An Examination of the Goodness of Fit Model: How is the Relationship Between Child Temperament and Behavior Expressed in Different Types of Classroom Environments?

by

Sasha Collins Blackwell

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2010

© 2010

Sasha Collins Blackwell

All Rights Reserved

This manuscript has been read and accepted for the
Graduate Faculty in Educational Psychology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Georgiana Shick Tryon, Ph.D.

Date — Chair of Examining Committee

Mary Kopala, Ph.D.

Date — Executive Officer

Dr. Ida Jeltova

Dr. Jay Verkuilen
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

THE CITY UNIVERSITY OF NEW YORK

Abstract

An Examination of the Goodness of Fit Model: How is the Relationship Between Child Temperament and Behavior Expressed in Different Types of Classroom Environments?

by

Sasha Collins Blackwell

Advisor: Georgiana Shick Tryon, Ph.D.

The present study examined how the relationship between child temperament and behavior is expressed in different types of classroom environments in prekindergarten settings. Other goals of the study were to further operationalize the goodness of fit model in school settings and to evaluate possible interactions of process variables indicative of classroom quality with child temperament to see if these interactions predicted child behavior and social skills. Participants included 130 students and their teachers (N = 11) in three prekindergarten settings. Child temperament was measured using the Total Temperament score from the Teacher and Caregiver Temperament Inventory for Children (TACTIC; Billman & McDevitt, 1998). Classroom quality and environment characteristics were measured using the Program Structure scale of the Early Childhood Environment Rating Scales-Revised (ECERS-R; Harms et al., 2005) and the Sensitivity subscale score from the Caregiver Interaction Scale (CIS; Arnett, 1989). Outcomes in behavioral and social domains were measured using the Externalizing Behavior Problems

and Social Skills subscales on the Preschool and Kindergarten Behavior Scales- Second Edition (PKBS-2; Merrell, 2002). Hierarchical linear modeling indicated that child temperament alone was the sole predictor of child externalizing behavior, while child temperament, disability status, and school program structure predicted child social skills. Overall, the study indicated that the goodness of fit model when operationalized in terms of the transactional relationship between temperament and environmental demand factors of characteristics of the classroom setting (as informed by the classroom quality literature) has predictive value and describes child behavioral and social outcomes in prekindergarten settings.

Acknowledgements

The task of acknowledging all those that helped to make this possible is not an easy one. Dr. Georgiana Shick Tryon, my advisor and mentor has diligently guided me through the process of selecting research interests, designing empirically sound methodology, analyzing and interpreting data, and expressing my thoughts in a clear and concise manner. I am so thankful for her patience, support, insight, and quick wit. Many thanks also are extended to my committee members, Dr. Jay Verkuilen and Dr. Ida Jeltova, who graciously offered their input and assistance.

A multitude of thanks is owed and offered to my family and friends who helped me trudge through the process of pursuing a doctorate degree and writing a dissertation. My parents, Allen and Stephanie Collins and siblings, Matthew and Alyssa, encouraged me throughout my years as a student, providing me with unconditional love, help, and guidance. My grandmother, Bertha Collins serves as my example as a stellar and fervent teacher and learner- I wish she could see this day. My mother- and sister-in-law, Barbara and Sarah Blackwell have helped me throughout this process on a moment's notice, always with smiles on their faces. My husband, Morgan Blackwell, my support center and sounding board, has provided me with love, laughter, and the utmost wisdom. Lastly, Haley Beth Blackwell, my daughter, has shown me how wonderful life can be. Thank you all, so very much.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

Introduction

Temperament consists of biologically based individual differences in behavior that are present early in life and are relatively stable across situations and time. While there are several theoretical models of temperament, this study considered that of Thomas, Chess, and Birch (1968), derived from the New York Longitudinal Study (NYLS) that began in the late 1950's. In this model, temperament is the stylistic component of behavior, "the how of behavior as differentiated from motivation, the why of behavior, and abilities, and the what of behavior" (Goldsmith et al., 1987, p. 508).

In 1956, the New York Longitudinal Study (NYLS) followed 141 infants for nearly 10 years and reported nine traits of temperament: activity level, rhythmicity, approach-withdrawal, adaptability, threshold of responsiveness, intensity of reaction, quality of mood, distractibility, and attention span and persistence. These nine temperamental traits clustered to form three temperament constellations: easy, slow-to-warm-up, and difficult (Chess & Thomas, 1996). This work "constituted part of the paradigm shift from a predominantly environmentalistic unidirectional perspective on child development, to one that acknowledged the child's own active part in the developmental process" (Sanson, Hemphill, & Smart, 2004, p. 143).

Chess and Thomas (1977) proposed the concept of goodness of fit that results when the opportunities, expectations, and demands of the environment are in accord with the child's temperament and other characteristics. When goodness of fit is present, they posited that optimal development in a positive direction is possible. However, a poorness of fit is thought to involve discrepancies and dissonances between environmental

opportunities, expectations and demands, and the temperament and other characteristics of the child, so that distorted development and maladaptive functioning may occur.

The majority of the existing research on goodness of fit focuses on the family, but goodness of fit between temperament and classroom environment also plays a powerful role in schools (Keogh & Speece, 1996). The extant research on goodness of fit in schools, however, measures more indirect aspects of the classroom environment (e.g., teacher expectations, teacher temperament) and focuses on cognitive and academic achievement and competencies (Churchill, 2003). However, goodness of fit is posited to predict a number of functional outcomes across settings, including personal and social adjustment (e.g., Carey; 1998; De Schipper, Taveecchio, Van IJzendoorn, & Jantine, 2004; Keogh, 1986). A good fit between a child's temperament and his or her environment may manifest itself in the child demonstrating higher levels of social competence, while a poor fit between a child's temperament and his or her environment may evolve into externalizing behavior problems that bring the child into conflict with others (e.g., Carey, 1998).

The goodness of fit concept was formulated retrospectively based on clinical records and through a process of inductive analysis (Chess & Thomas, 1977). As a result, the NYLS researchers never fully defined or operationalized it as a construct for purposes of experimental replication and extension. Consequently, methodological difficulties have plagued this area of research, with the measurement and conceptualization of fit proving to be most difficult (Windle & Lerner, 1986). As a result, investigators have operationalized goodness of fit in several ways: in terms of environmental demand factors of caregiver characteristics (i.e., caregiver temperament) (e.g., Sprunger, Boyce,

& Gaines, 1985; Wallander et al.,1988), environmental demand factors of caregiver expectations of children's behavior (i.e., the discrepancy between caregivers' ratings of children's actual and ideal temperaments) (e.g., Klein & Ballentine, 2001; Lerner, 1983, Palermo, 1982, Winde & Lerner, 1984), and environmental demand factors of characteristics of the setting (i.e., program structure, stability of setting, physical characteristics of setting) (De Schipper et al., 2004).

Studies that have operationalized goodness of fit in terms of environmental demand factors of the setting have measured classroom environment characteristics like program quality, stability, emotional climate, physical arrangements, stimulation, and discipline methods. This type of methodology is in accord with the ideas of Keogh (2003; Keogh & Speece, 1996) who proposed that several aspects of classrooms are important in the fit between students and classroom environments: the content and nature of the curriculum; the organization of space, time, and resources; and the nature of interactions between students, peers, and teachers. The results of the small number of studies that have assessed these characteristics typically indicate that high quality educational programs are of greater benefit to temperamentally easy children than to difficult children (e.g., De Schipper et al., 2004; Klein, 1982). Thus, this operationalization of goodness of fit appears to be promising because it offers a framework for understanding how child characteristics interact with the classroom environment to predict behavioral and affective outcomes.

Given that only a handful of studies have operationalized goodness of fit in terms of demand features of the setting, the investigator has looked to the early childhood classroom quality literature to determine the best way to measure classroom environment

characteristics that may interact with child temperament. This literature examines components of early childhood environments (i.e., preschool, prekindergarten, daycare) that relate to positive child outcomes in academic and social domains, such as social behavior, cognitive development, language development, compliance, and self-regulation (Hestenes, Kontos, & Bryan, 1993; Pianta et al., 2005). Accordingly, both the classroom quality and goodness of fit literatures seek to identify elements of school and classroom environments that are associated with positive developmental outcomes.

The classroom quality literature identifies two sets of variables that have been shown to be related to behavioral and social domains: structural variables and process variables. Structural variables are those like teacher credentials and education levels, and teacher-child ratios. While structural variables can provide some sense of the classroom characteristics that affect child outcomes (e.g., Burchinal et al., 2002; Howes et al., 1992), process variables, such as teacher-child interactions, type of instruction, classroom materials, program structure, and teacher sensitivity have consistently been demonstrated to relate to children's positive developmental outcomes (e.g., Bryant et al., 1994; Hestenes et al., 1993; Howes et al., 1994; Kontos & Wilcox-Herzog, 1997; Lambert et al., 2002; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001; Pianta, 1999; Pianta et al., 2002; Vandell, 2004).

The purpose of this study was to determine how the relationship between child temperament and behavior is expressed in different types of classroom environments in prekindergarten settings. Other goals of the study were to further operationalize the goodness of fit model in school settings and to evaluate possible interactions of process

variables indicative of classroom quality with child temperament to see if these interactions predict child behavior and social skills.

Child temperament was measured using the Total Temperament score from the Teacher and Caregiver Temperament Inventory for Children (TACTIC; Billman & McDevitt, 1998). To determine the relationship between temperament and demand characteristics of the school setting, classroom environment characteristics were measured. This study used the Program Structure subscale of the Early Childhood Environment Rating Scales-Revised (ECERS-R; Harms et al., 2005), and the Sensitivity subscale of the Caregiver Interaction Scale (CIS; Arnett, 1989). Direct observation (e.g., teacher-child ratios) and teacher responses to standard demographic questionnaires (e.g., years of education, completed degrees, number of years teaching) provided additional information regarding classroom structural variables. The Externalizing Behavior and Social Skills subscales on the Preschool and Kindergarten Behavior Scales- Second Edition (PKBS-2; Merrell, 2002), measured child outcomes in behavioral and social domains.

The results of this study indicated that child temperament is a highly important correlate and predictor of both student social skills and externalizing behavior. In addition, the study indicated that classroom quality, as described by the manner in which an educational program is structured and in terms of the warmth and sensitivity with which caregivers interact with their students is related to child outcomes in behavioral and social domains. Specifically, hierarchical linear modeling indicated child temperament alone to be the sole predictor of child externalizing behavior, while child temperament, disability status, and school program structure predicted child social skills.

Finally, the study indicated that the goodness of fit model when described in terms of the transactional relationship between temperament and environmental demand factors of characteristics of the classroom setting describes child behavioral and social outcomes in prekindergarten environments.

CHAPTER II

Literature Review

This chapter will provide a review of the literature pertaining to the three constructs that constituted the main focus of this dissertation: temperament, goodness of fit, and classroom quality. It will begin with a discussion of temperament in terms of its definition and as a theoretical construct that is important to study. After providing the formulation of temperament derived from the New York Longitudinal Study (NYLS) (Thomas et al., 1963), there will be brief descriptions of other models of temperament and an explanation of why the NYLS model of temperament is most relevant. A discussion of measurement issues that pertain to temperament and the relationship between temperament and behavior problems in children will follow.

The next large section of the chapter will provide an examination and definition of the goodness of fit construct through a review of the goodness of fit literature, including research, as it relates to school settings. Additionally, I will discuss the three different ways researchers have operationalized goodness of fit in school settings: in terms of environmental demand factors of caregiver characteristics, environmental demand factors of caregiver expectations of children's behavior, and environmental demand factors of characteristics of the setting.

The subsequent section of the chapter will focus on classroom quality and classroom environment characteristics as they relate to behavioral and social outcomes in children in early childhood environments. First, it will explain how the concept of classroom quality is relevant to the larger discussion of temperament and goodness of fit. Second, it will describe the two ways of describing and measuring classroom quality:

structural markers of quality and process quality. The chapter will end with a presentation of the rationale for this study and the hypotheses to be tested.

Temperament

The term temperament originates from the Latin word *tempare*, which one can define as to mix, meaning that various influences interact with one another to create behavior (Kristal, 2005). Temperament also has longstanding intuitive and folk literature definitions and is widely accepted as a fundamental aspect of the psychological mechanisms of behavioral functioning (Chess & Thomas, 1996). As a concept, temperament serves to "tie together a variety of primary behavioral dispositions commonly used to distinguish one individual from another" (Bates, 1989, p. 4). While many researchers have studied and sought to define this concept, the most general definition of temperament across theories is that it consists of biologically-based individual differences in behavior that are present early in life and are relatively stable across situations and time (Goldsmith et al., 1987). In addition, there is general agreement that temperament is manifested in the context of social interaction (Goldsmith et al., 1987). Notably, most temperament models justify their use of the term temperament by referring to Allport's definition:

> Temperament refers to the characteristic phenomena of an individual's emotional nature, including his susceptibility to emotional stimulation, his customary strength and speed of response, the quality of his prevailing mood, and all the peculiarities of fluctuation and intensity of mood, these phenomena being regarded as dependent upon constitutional make-up and therefore largely hereditary in origin. (Allport, 1961 as cited in Buss & Plomin, 1984, p. 3)

The concept of temperament is important to study because it offers advantages in discussing the origins of children's differences in behavioral style; as such, "it is a shorthand way of postulating that there are endogenous forces shaping the child's individual, social characteristics" (Bates, 1985, p. 11). In addition, it is a useful way to summarize individual differences in socially relevant behavior and it guides the selection of variables for measurement in the study of children's individual differences (Bates, 1989; Goldsmith et al., 1987). Moreover, it guides ways to formulate and understand children's behavior problems, because having a biological attribution for child's behavior can help caregivers to cope effectively with the child's characteristics, and understanding child's temperament can allow for the prevention of negative effects that may arise from a discordance between the child's temperament and characteristics of his or her environments (Bates, 1989; Keogh, 2003).

NYLS Group Formulation of Temperament

While there are several theoretical models of temperament, the present study considers that of Thomas, Chess, and Birch (1963) formulated as a result of the New York Longitudinal Study (NYLS) that began in the late 1950's. In this model, temperament is the stylistic component of behavior, "the how of behavior as differentiated from motivation, the why of behavior, and abilities, and the what of behavior," and can be equated to the term "behavioral style" (Goldsmith et al., 1987, p. 508). In this definition, style refers to the "how of behavior," the "what of behavior" refers to the content of behavior, and the "why of behavior" refers to the motivational component of behavior (Martin, 1983). The formulation of temperament in the NYLS model includes the following considerations: (a) temperament is an independent

psychological attribute; (b) one must differentiate temperament from motivations, abilities, and personality; (c) temperament is always expressed as a response to an external stimulus, opportunity, expectation, or demand; (d) temperament is a dynamic factor that shapes the influence of the environment on the individual's psychological structure (Bates, 1985; Goldsmith et al., 1987). In addition, the impact of temperament is bidirectional - the effect of a particular environmental influence will be influenced by the child's temperament; and the child's temperament will affect the judgments, attitudes, and behavior of the significant individuals in his/her environment (Goldsmith et al., 1987).

The work of the NYLS group "constituted part of the paradigm shift from a predominantly environmentalistic unidirectional perspective on child development, to one which acknowledged the child's own active part in the developmental process" (Sanson, Hemphill, & Smart, 2004, p. 143). Thus, this view of temperament is in accord with interactionist models that propose that development is neither biologically, maturationally, nor environmentally determined (Keogh, 2003). Accordingly, in this view, children are influenced by and also can influence their environments, and development and behavior are the result of ongoing, continuing interactions and transactions between the individual and the environment where temperament plays a contributing role to the nature of experience and development (Keogh, 2003).

Chess, Thomas, and colleagues developed their model of temperament as a result of the NYLS, when beginning in 1956, they followed 131 infants and children of 85 homogeneous sociocultural families for nearly 10 years, examining the role of temperament in child development. They collected data through parental and teacher

interviews and direct observations (Rice & Gaines, 1992). An empiric inductive analysis of these detailed behavioral data yielded nine dimensions of temperament on which individuals can vary from low to high: (a) activity level (the motor component present in a child's functioning and the daily proportion of active and inactive periods); (b) rhythmicity (the predictability and unpredictability in time of any biologic function); (c) approach-withdrawal (the nature of the initial response of a new stimulus, which can be positive or negative); (d) adaptability (the responses to new or altered situations); (e) threshold of responsiveness (the intensity level of stimulation necessary to evoke a discernable response); (f) quality of mood (the amount of pleasant behavior in contrast with unpleasant behavior, rated as the preponderance of positive versus negative mood expression); (g) intensity of reaction or mood expression (the energy level of response); (h) distractibility (the effectiveness of extraneous external stimuli in interfering with or in changing ongoing behavior); and (i) a double category of attention span (the amount of time an activity is pursued) and persistence (the continuation of an activity in the face of obstacles to the maintenance of the activity) (Chess & Thomas, 1991; 1996).

Chess, Thomas, and colleagues identified three temperament constellations through factor analysis as well as clinically (Chess & Thomas, 1991). Regularity, positive approach responses to new stimuli, high and quick adaptability, mild or moderate intense positive mood, good rhythmicity, and low frustration characterize the "easy temperament." Notably, children with easy temperaments formed 40% of the NLYS sample. A combination of negative response approach with slow adaptability, mild intensity of reactions characterize the "slow-warm-up" temperament. Comprising 15% of the NLYS sample, children with this temperament pattern also show quiet interest upon

experiencing new stimuli over time. Finally, irregularity in activity level and rhythmicity in biologic functions, negative withdrawal responses to new situations, nonadaptability to change, intense negative mood expressions, high frustration, and loud laughter typify the "difficult temperament" (representing 10% of the NYLS sample) (Chess & Thomas, 1991, 1996).

Not all individuals fall into one of the noted three temperament constellations (Chess & Thomas, 1991). In fact, 35% of the NYLS sample did not fall into one of these three temperament groups but fell into one of several intermediate categories of temperament (Chess & Thomas, 1996). Notably, the researchers did not describe the characteristics of these "intermediate" groups. In addition, it is quite possible for different combinations of the nine temperamental traits to occur, as well as for a person who fits within one of the three patterns of temperament to show a wide range in the degree of temperamental manifestation across different contexts. Notably, a large number of studies across the United States with heterogeneous samples as well as studies in a number of European and Asian countries (Thomas & Chess, 1977; Ciba Foundation Symposium 89, 1982) have identified the nine dimensions of temperament and three temperamental constellations elucidated in the NLYS.

Others have suggested additional clusters of the nine temperamental characteristics proposed by Chess and Thomas. For example, Keogh (1989) derived clusters that she posited to be useful to teachers and school psychologists for the evaluation of performance of school-age children: task orientation (composed of persistence, distractibility, and activity), personal-social flexibility (composed of adaptability, approach-withdrawal, and positive mood), and reactivity (a negative factor,

composed of intensity, threshold, and negative mood). Keogh's factors are described in short versions of Thomas and Chess (1977) Parent and Teacher Temperament Questionnaires and correspond to factors found by Windle and Lerner (1985) and are similar to those defined by Paget, Nagel, and Martin (1984).

Other Models of Temperament

While investigators and clinicians have used the nine temperamental traits originally proposed by Chess and Thomas widely in research and in clinical practice, some have voiced concerns of conceptual overlap that have led to the development of different approaches to conceptualization and definition. Accordingly, temperament theorists proposed different ideas about the origins, expressions, stability, and boundaries of temperament (Goldsmith et al., 1987). Gray (e.g., 1987) proposed a theory of temperament based on a neuropsychological model of brain functioning. He described internalizing and externalizing problems in terms of the inhibition and facilitation of behavior. In this model,

> behavior is seen as resulting from two different brain systems: a behavioral activation system (BAS) that generates appetitive and aggressive behavior and is sensitive to reward, and a behavioral inhibition system (BIS) that mediates the inhibition of behavior in novel situations and in the presence of cues signaling impending punishment. (Oldehinkel et al., 2004, p. 422)

Rothbart and colleagues formed an alternative theory of temperament based on the work of Chess and Thomas and colleagues as well as the work of Gray, founded on the premise that temperament is a multidimensional representation of constitutionally based individual differences in reactivity and self-regulation (Rothbart & Putnam, 2002).

Rothbart's model distinguishes eight temperamental dimensions; namely, high-intensity pleasure, shyness, fear, frustration, activation control, attention control, inhibitory control, and affiliation that reflect four broad dimensions of temperament (Rothbart & Putnam, 2002). These four broad temperament factors include: surgency, which is similar to approach-withdrawal and concerns an individual's orientation to and exploration of novelty; negative affectivity, which is similar to reactivity and encompasses both fear and frustration in reaction to stimuli; effortful control, which is similar to attention span and persistence and denotes the ability to regulate attention and behavior; and affiliation, which refers to the desire for closeness with others, independent of introversion or extroversion (Oldehinkel et al., 2004; Rothbart & Putnam, 2002).

Buss and Plomin also formulated the EAS (emotionality, activity, sociability) model of temperament based on the work of the NYLS group. They defined temperament as "a set of inherited personality traits that appear early in life" (Goldsmith et al., 1987, p. 508). These authors viewed dimensions of temperament as enduring personality traits that have genetic origins that appear early in infancy (specifically during the first year of life). In addition, the model identifies three general traits of temperament: emotionality, defined as primordial distress, which is assumed to differentiate into fear and anger during the first six months of life; activity, defined as sheer expenditure of physical energy; and sociability, defined as preferences for being with others rather than begin alone (Buss, 1989).

Yet another theory proposed by Goldsmith (1987) defines temperament as individual differences in the probability of experiencing and expressing primary emotions and arousal and includes dimensions such as activity level, joy/pleasure, social

fearfulness, anger proneness, and interest/persistence (Goldsmith et al., 1987). Thus, Goldsmith and colleagues viewed temperament as emotionality, because emotions regulate internal psychological processes and social behaviors (Kristal, 2005). In addition, while Goldsmith contended that genetics and physiology determine temperament, she focused on the behavioral nature of temperament because it is most meaningful in social contents and facilitates empirical investigation.

Even with these differences in the conceptualization and description of temperament, there is consensus across theories that temperament refers to individual differences rather than species-general characteristics and that temperament dimensions reflect behavioral tendencies rather than discrete behavioral acts, as well as an emphasis on the biological underpinnings and continuity of temperament (Goldsmith et al., 1987). Despite the availability of these alternative models of temperament, many have found that the NYLS group's formulation of temperament has the most clinical relevance, specifically as it applies to school settings (e.g., Carey, 1998; Kristal, 1986; McDevitt, 1994). This may be because these dimensions are "the only set of traits derived from clinical observations and the only ones used extensively and successfully in a variety of clinical settings" (Carey, 1998, p. 523). In addition, clinicians in pediatrics, psychiatry, psychology, and education have readily applied and developed the NYLS formulation of temperament, while others have developed alternative views of temperament based purely on empirical approaches (Kristal, 2005; McDevitt, 1994). Moreover, the NYLS conceptualization of temperament has made it possible to understand individual differences in children's responses to similar environmental situations (e.g., Dunn & Kendrick, 1982; Graham, Rutter, & George, 1973; Maziade et al., 1985).

Maziade and colleagues' (1985) study on the predictive value of difficult temperament is an example of research that demonstrates the utility of the NYLS model in understanding individual differences in children's responses to analogous environments. The study sample included 980 12-year-old children from the general population of Quebec City whose temperaments had been assessed and determined to be difficult or easy at age 7 using a French translation of the Thomas and Chess (1977) Parent Temperament Questionnaire (described on p. 17). The authors balanced the difficult and easy groups for age, sex, and socioeconomic status. They administered semi-structured interviews to parents to assess the children's clinical status. In addition, parents completed several ratings scales to assess child behavior and responded to a semi-standardized interview to assess family functioning in terms of behavior control according to the McMaster model of family functioning (Epstein, Bishop, & Levin, 1978). This study's findings indicated that temperamentally difficult children had more clinical disorders at age 12. In addition, results suggested that there was an increased risk for temperamentally difficult children in homes where "parents show[ed] little consensus and if family rules and demands [were] not clear or a lack firmness and consistency" (Maziade et al., 1985, p. 402). However, this was not found to be the case in families of difficult children found to have "superior functioning." These results lead the authors to hypothesize that a positive change in behavior control in dysfunctional families of difficult children might improve the psychosocial prognoses of the children.

In summary, the concept of temperament is one that has been regarded as an important contributor to human behavior for many years. Thomas and Chess and colleagues (1968) conducted the seminal research on child temperament of young

children in the New York Longitudinal Study. A plethora of research followed the NYLS, each describing temperament in several different manners. The NYLS formulation of temperament, however, is the most widely accepted and researched model, as its clinical applications appear to be the most useful.

Measurement of Temperament

A wide variety of instruments exist to assess temperament that operationalize the theoretical framework proposed by the NLYS group (Hubert et al., 1982). These include the Parent and Teacher Temperament Questionnaires (PTQ,TTQ; Thomas & Chess, 1977), designed to provide a systemic manner of describing children on each of the nine temperament dimensions as described by the NYLS; the Parent and Teacher Temperament Questionnaires- Short Form, formulated by Keogh, Pullis, and Cadwell (PTTQ-S; 1982), which are abbreviated forms of the Thomas and Chess (1977) questionnaires, used for research purposes, which contain the 3 factors (described above); the Carey Temperament Scales (CTS; Carey, 2000), which include the Early Infancy, Revised Infant, Toddler, Behavioral Style, and Middle Childhood Temperament Questionnaires, often used in clinical and pediatric practice; the Dimensions of Temperament Survey (DOTS) and DOTS-Revised, developed to assess age-continuous aspects of temperament and used for longitudinal research (DOTS, Lerner, Palermo, Spiro, & Nesselroade, 1982; DOTS-R, Windle & Lerner, 1986); the Temperament Assessment Battery for Children (TABC; Martin, 1988), which assesses six out of the nine NLYS temperament dimensions; and a newer instrument, the Teacher and Caregiver Temperament Inventory for Children, (TACTIC; Billman & McDevitt, 1998), a

comprehensive measure of behavior that differentiates between normal temperamental variations and clinical problems.

The three most common approaches to measuring temperament are interviews, direct observations, and rating scales or questionnaires (Rice & Gaines, 1992). Respondents, mainly parents and teachers, make a judgment about the child's behavioral style. This study utilized a teacher-completed rating scale to assess child temperament. Rating scales or questionnaires are the most frequently used assessment approach (Teglasi, 1998). This is partly due to the benefits that rating scales and questionnaires offer, such as ease of administration and cost and time effectiveness. However, parent- and teacher-completed questionnaires may be influenced by observer bias and inaccuracy in recollections, observations, or interpretations (Teglasi, 1998) (discussed on pp. 20-21).

Major differences in assessment instruments (e.g., rating scales and questionnaires) occur in several areas (McDevitt, 1994). First, instruments vary in the concept and dimensions of temperament being measured. Some researchers (e.g., Buss & Plomin, 1975) have challenged the NLYS temperament constructs and argued that their conceptual overlap calls for them to be reduced to a smaller number of purer dimensions. Yet, the temperament dimensions derived from the NYLS have demonstrated clinical utility that allows for the provision of prevention and intervention services to children and their caregivers (McDevitt, 1994).

For example, The Parent Temperament Program in LaGrande, Oregon was designed in consultation with Chess and Thomas to teach parents positive parenting techniques, especially those targeted to children classified as having the difficult temperament profile (Smith, 1994). This program has several stages, the first of which is

an assessment phase where child temperament is assessed according to the NYLS model in a semi-structured parent interview and child behavior is measured using the 36-item Eyberg Child Behavior Inventory (ECBI; Eyberg & Ross, 1978). The second phase is one that focuses on parent-child relationship strengthening, where parent-child stress is reduced which contributes to achieving a "good fit." The third phase includes specific parenting advice and support, where a temperament specialist works with the child's parents to develop a plan for resolving some or all of the concerns that prompted their entry into the program. This may include providing referrals to other agencies, consultations with teachers, and the provision of long-term emotional support to families. The final phase is an ongoing one of continued availability, where parents are encouraged to return to the program if a new issue should arise or old issues are to resurface. Program outcomes, assessed through a Parent Satisfaction Survey, indicate that 79% of parents have been generally satisfied with the program with 79% also reporting "some "or "much" improvement in their child's behavior (Koroloff, 1989, 1990, 1991, 1992).

A second area in which assessment instruments vary is on how they consider the effect of various contexts on temperament. Some researchers believe (e.g., Carey & McDevitt, 1978; Thomas & Chess, 1977) that there is unconvincing evidence of the cross-situational consistency of temperament. This has resulted in the development of an interactional viewpoint, where temperamental styles are seen as having a "range of specific expressions, with the 'real' temperament being at some midpoint or most frequently expressed value around which the majority of typical situations tended to cluster" (McDevitt, 1994, p. 196), which is accordingly reflected in assessment instruments (i.e., CTS; Carey, 2000; TACTIC, Billman & McDevitt, 1998). The present

study used a rating scale that views the relationship between temperament and context in this manner. Others (e.g., Buss & Plomin, 1975) believe that temperament is genetically determined and therefore independent of context. Thus, analogous instruments (e.g., EAS Survey; Buss & Plomin, 1984) are aimed at reducing cross-situational consistency by having raters respond to items with no context included at all (McDevitt, 1994).

A third concern about assessment information is that there are questions about caregivers' (both parents and teachers) ability to observe and rate children's behavioral styles objectively, therein calling the validity of caregiver report into question (Keogh, 2003; McDevitt, 1994). These concerns are the result of studies that indicated poor correspondence between caregiver report and direct observation (correlations in the .20-.40 range) (e.g., Vaughn et al., 1987). However, it may be that parents' and teachers' views of children's temperaments are selective, reflect individual expectations for behavior and conduct, and/or change over time (Keogh, 2003). Also, teachers, in particular, have a normative basis of comparison, which enables them to recognize consistencies in students' behaviors and temperaments, as well as to be aware of slight temperamental differences between students (Keogh, 1982; 2003).

Teachers may use their ability to recognize differences in students' temperaments to make decisions in the classroom. Pullis and Cadwell (1982) studied the influence of children's temperament characteristics on teachers' decision strategies in a sample of 321 elementary school children and 13 primary-level teachers from a small suburban school district. The Short Version of the Teacher Temperament Questionnaire (TTQ-Short; Keogh et al., 1982) measured child temperament and student aptitude, teacher classroom management, and placement decisions were assessed using a teacher-completed

questionnaire rated on a 6-point Likert scale. The most significant finding of this study was the "teachers' reliance on temperament information when making classroom management decisions. Even when controlling for ratings of ability, motivation, and social interaction skills, temperament consistently entered the regression equation (Pullis & Cadwell, 1982).

In another study, Pullis and Cadwell (1985) investigated the influence and utility of information about children's temperament in the process of assessment of educational risk and the implementation of intervention approaches in a group of 54 students who were identified as being "at risk" educationally from a sample of 247 first and second grade students. Ten primary level teachers provided ratings of children's temperaments using the TTQ-Short (Keogh et al., 1982), and rated children's general ability, academic performance, teacher classroom management, and placement decisions on a questionnaire that had a 6-point Likert scale. In addition, parents consented to the provision of information on students' achievement and aptitude tests to the researchers. Results from a series of multiple regression analyses indicated significant relationships between temperament and teachers' classroom management decisions. Specifically, temperament was most useful in terms of guiding teachers' efforts at adaptive instruction at the students' instructional level.

A fourth and final area of instrument contention is the fundamental differences between those who believe that temperament as a concept "includes so much rater bias that it is partially (or largely) defined by a measurement artifact…. and those that believe that scores on temperament measures may contain a certain amount of error attributable to inaccuracies by the individual rating the infant of child" (McDevitt, 1994, p. 197). In

this manner, a perception-orientated approach either deals with raters by labeling temperament ratings as perception because this source of bias cannot be eliminated (e.g., Wolk et al., 1992), or by viewing the perception rating as being the more accurate predictor of later status since both sides of the interaction are being measured by the assessment (e.g., Keogh, 1982). A third approach measures both specific ratings (i.e., dimensions) and general impressions (i.e., global impressions/perceptions) of temperament (i.e., Carey, 1989). This approach discriminates between the two and uses them in different ways.

In sum, while rating scales and questionnaires are the most widely used measurement tools of temperament, there are several major differences in assessment techniques. Given the plethora of these issues, there remains to be an available objective method for determining temperamental status that investigators accept as a benchmark for a comparison of validity with other measures. Until someone develops such a measure, it is important that one judges temperament assessments on a number of characteristics, including their factorial integrity, validity, and reliability; breadth and specificity of item and scale content; age appropriateness; potential biasing effects; psychometric qualities; and clinical utility (Keogh, 2003; McDevitt, 1994).

Temperament and its Relationship to Behavior Problems

Certain temperamental characteristics may predispose a child to behavior and adjustment problems. The NYLS indicated five major temperament constellations related to behavior problems: (a) negative mood combined with irregularity, nonadaptability, and negative withdrawal responses; (b) excessive persistence; (c) withdrawal and mild negative reactions to new situations combined with slow adaptability; (d) extreme

distractibility; and (e) very high or low activity level (Thomas, Chess, & Birch, 1968). In addition, while Thomas and colleagues found youngsters in the NLYS sample who were classified as having easy temperaments to adapt relatively smoothly to the demands of schooling (because they were able to behave in ways that were consonant with those of their peers), this was not the case for who were classified as having difficult temperaments (Thomas et al., 1968). The NYLS also found that individual differences in negative mood, nonadaptability, lack of rhythmicity, and intensity during early childhood were associated with externalizing behavior problems in late childhood.

A large body of research has followed the NLYS, indicating that these temperament characteristics continue to be relevant. Using data from the NLYS (discussed on pp. 8-11), Terestman (1980) found that prekindergarten teachers' ratings of mood and intensity discriminated between children with and without behavior problems. Teglasi and MacMahon (1990) (discussed on pp. 25-26) surveyed parents of 362 children, ages 8 to 12 years-old, to determine sources of friction within the family, and found strong associations between externalizing behavior problems and several dimensions of temperament, including negative mood, low adaptability, low persistence, nonapproach responses, and high activity levels.

In another study, using the NYLS nine-dimension model of temperament, Wertlieb and colleagues (1987) found significant relationships between temperament, stress, hassles with parents, and behavior problems in 158 6- to 9-year-old children. Behavior problems, as measured by the Parent Report Form of the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1983), related most to a temperament profile of negative mood, withdrawal responses, low adaptability, low persistence, high

distractibility, and unpredictability, as measured by the parent-completed Middle Childhood Temperament Questionnaire (MCTQ; Hegvik et al., 1982), that is one of the CTS (Carey, 2000; discussed on p. 17). In addition, Hagekull (1994) found that negative emotionality, impulsivity, and high activity level in toddlerhood, assessed using the Toddler Behavior Questionnaires (TTQ; Hagekull & Bohlin, 1981), which corresponds to the NYLS model of temperament, and the Emotionality, Activity, Sociability, Impulsivity Questionnaires (EASI; Buss & Plomin; 1975, 1984) , which corresponds to Buss & Plomin's EAS model of temperament (described on p. 14), predicted externalizing behavior problems, measured using the Preschool Behavior Questionnaire (PBQ; Behar & Stringfield, 1994) in a sample of 376 4-year-old children. Overall, these results indicate the importance of the effect of temperament on children's behavior and adjustment (Sanson, Hemphill, & Smart, 2004).

Global measures of "difficult" temperament, as described in the difficult temperament profile also predict the development of externalizing behavior problems in children. This study used such a measure of temperamental difficulty. The NLYS found that when measured at 3, 4, and 5 years of age, difficult temperament characteristics related to numerous adjustment problems at school, at home, and during early adulthood (Caspi & Silva, 1995). Tschann and colleagues (1996) proposed that "children with more difficult temperaments may be especially vulnerable to stress and therefore likely to develop behavior problems, under stress" (p. 187). A longitudinal study conducted by Maziade and colleagues (1990) followed 38 children from 7 to 16 years of age to investigate the influence of temperament on the development of externalizing behavior problems. A parent-completed semi-structured interview according to DSM-III

diagnostic criteria established clinical status, and a French translation of the Thomas and Chess (1977) Parent Temperament Questionnaire (described on p. 17) assessed temperament. Results indicated that children with clinically significant externalizing behavior problems scored higher on measures of difficult temperament.

Guerin and colleagues (1997) conducted a 10-year longitudinal study that followed 100 children ages 1.5 through 12 that examined the relationships between temperament and behavior problems. A parent-completed rating scale, the Infant Characteristics Questionnaire (ICQ; Bates et al., 1979), measured temperament and parent report when children were 3.25 through 12 years old and teacher report annually from 6 through 11 years using several instruments, including the Eyberg Child Behavior Inventory (ECBI; Robinson, Eyberg, & Ross, 1980), the Preschool Behavior Questionnaire (PBQ; Behar & Springfield, 1974), and the Child Behavior Checklist for ages 4-18 (CBCL; Achenbach, 1991) assessed behavior problems. These authors found that the temperament difficultness factor correlated consistently with parent reports of behavior problems across all ages, more strongly for externalizing behavior problems than for internalizing behavior problems, and highly with teacher report of externalizing behavior problems at 6 to 8 years of age.

Others have studied the relationship between temperament and behavior problems in schools. In a study of 362 children between the ages of 8 and 12, Teglasi and MacMahon (1990) found that teachers conceptualized good citizenship (assessed by a teacher rating form created for the study that included 13 items, grouped into two categories: citizenship and work habits) in terms of how well children collaborate with peers, their respect for property and for others, and their responsibility for personal

behavior. Teachers' ratings of students' behavior in the classroom (as measured by the Problem Behavior Survey, an instrument developed for the study) related significantly and positively to children's temperaments, as measured by parent's ratings on the Middle Childhood Temperament Questionnaire (Hegvik et al., 1982), which is one of the CTS (Carey, 2000) that assesses the 9 NYLS temperament traits in 8- to 12-year-old children, and specifically to temperamental dimensions of predictability (a form of rhythmicity) and persistence. These findings dovetail with those of Guerin, Gottfried, Oliver, and Thomas (1994) who also found significant associations between teachers' views of children's temperaments and how well the children function in the classroom. This study included data drawn from the Fullerton Longitudinal Study that followed 130 1-year-olds and their families for 10 years. Mothers completed the Infant Characteristics Questionnaire (ICQ; Bates et al., 1979) to assess temperament at 1.5 years old, and teacher reports (as well as parent report) annually from 6 through 11 years using several rating scales assessed behavior problems. Notably, this study found temperament to be the sole predictor of classroom behavior.

Overall, results of these studies provide support for the notion of difficult temperament as an important contributor to behavior and adjustment problems. However, it must be noted that difficult temperament alone does not lead to maladjustment. Carey (1998) noted that "temperament is a matter of style; behavioral maladjustment means substantial disturbance of social relationships, autonomy, or task performance. A volatile temperament is behavioral style; social alienation due to temper is behavioral maladjustment" (p. 132). In addition, Carey emphasized that even the extremes of

temperament are neither equivalent to behavior problems nor do they inevitably lead to them.

Instead, behavior problems may result from one or more situations (Keogh, 2003). First, it is possible that temperament and behavior problems are expressions of the same underlying condition, but that temperament is a factor that is evident earlier in the child's development (Keogh, 2003). For example, evidence of difficult temperament early in infancy may be an indication of a behavior disorder that will emerge as the child gets older (e.g., Guerin et al., 1997). Thus, this explanation suggests a lack of independence between temperament and behavior problems. However, this is not likely the case, because not all children with difficult temperament go on to develop behavior disorders (Keogh, 2003). For example, Maziade and colleagues (1990) (discussed on p. 24) found that extreme temperament at age 7 predicted psychiatric status in preadolescence and adolescence only when family functioning was also taken into account.

A second explanation that Carey (1998) offered is that behavior problems may result when children's temperaments are reinforced and become incorporated into generalized maladaptive coping strategies. The literature has not formally documented this explanation. In a study that examined the relationship between temperament, coping abilities, and responses to stress, Carson and Bittner (1994) administered the Stress Impact Scale (SIS; Hutton & Roberts, 1990) to 60 9- to 12-year-old children. In addition, trained observers completed an observational rating scale, the Coping Inventory (CI; Zeitlin, 1985), and mothers completed The Stress Response Scale (SRS; Chandler, 1985, 1986) and the Middle Childhood Temperament Questionnaire (Hegvik, McDevitt, & Carey, 1980). Results indicated that temperamentally difficult children may be more

vulnerable to both major and minor stresses in life, and therefore more likely to go on to develop a behavior disorder later in life. Thus, this second explanation is not likely, because while coping style has been shown to be partially influenced by temperament (Carson & Bittner, 1994), there is a difference between the two concepts.

Finally, behavior problems may result from poor fit situations in which the learning, independence, social, and/or environmental demands and characteristics clash with the child's temperament or exceed the child's skill level (Pullis, 1989). For example, Thomas et al. (1968) found that children with difficult temperaments in their sample went on to develop behavior problems mainly when parents had not been able to manage them in ways that diminished the stress in their relationships. In addition, Maziade (1989) discovered that even extremely difficult temperaments resulted in behavior problems only in children whose families had dysfunctional relationships. For additional examples, see examples provided in the discussion of the concept of goodness of fit (pp. 29-52). This explanation is the most plausible one that best conceptualizes and understands temperament and behavior problems within an interactional model. "This view means getting away from thinking of problems as 'in the child' and rather of thinking of problems in terms of the 'child in context'" (Keogh, 2003, p. 106).

Thus, a plethora of research indicates that certain temperament characteristics (i.e., negative mood, low adaptability, low persistence, nonapproach responses, withdrawal responses, high distractibility, high intensity, and high or low activity level) (e.g., Hagekull, 1994; Teglasi & MacMahon, 1990; Thomas et al., 1968; Wertlieb et al., 1987) as well as the difficult temperament profile (e.g., Caspi & Silva, 1995; Guerin et al., 1997; Maziade et al., 1990; Teglasi & MacMahon, 1996; Thomas et al., 1968)

predispose a child to behavior and adjustment problems. However, temperament is not posited to be the sole cause of behavior problems. Instead, behavior problems are most likely the result of poor fit situations in which discrepancies and dissonances between environmental opportunities, expectations and demands, and the temperament and other characteristics of the child occur (Chess & Thomas, 1977).

Goodness of Fit

Defining Goodness of Fit

The NYLS model of temperament is an interactionist (or transactional) one, because a central tenet of this approach is the "goodness of fit" construct (Chess and Thomas, 1977; Thomas et al., 1968). Chess and Thomas (1977) proposed this concept and defined it in terms of the nature of the interactions of the individual and the environment. The following quote from Thomas et al. (1968) explains the idea that temperament is situationally embedded and that normal and pathogenic processes do not depend on temperament alone:

> In themselves, temperamental attributes are neither good nor bad. Whether a given temperamental trait of a child meets with approval or disapproval, results in praise or criticism, or proves convenient or inconvenient to adults and peers can depend upon its appropriateness to the situations in which it is expressed and upon the degree to which its manifestations corresponds to the value judgments of others. (p. 100)

Thomas and Chess' (1977) conceptualization of goodness of fit is that it results when the opportunities, expectations, and demands of the environment are in accord with the child's temperament and other characteristics. When goodness of fit is present,

optimal development in a positive direction is thought to be possible (Thomas and Chess, 1977). However, a poorness of fit is posited to involve discrepancies and dissonances between environmental opportunities, expectations and demands, and the temperament and other characteristics of the child, so that distorted development and maladaptive functioning may occur.

As noted, goodness of fit places great emphasis on the meaning and relevance of temperament as it is situationally embedded (Keogh, 2003). Thomas and Chess (1977) suggested that the impact of temperament for psychosocial adaptation depends on whether a person's characteristics (i.e., temperament) provide a "goodness of fit" with the demands present in the physical and social environment. "A person's individuality, in meeting these demands, will provide a basis of the feedback he receives; thus, people whose temperamental characteristics fit their setting should show evidence of more adaptive behavioral functioning than should mismatched people" (Lerner, 1983, p. 150). However, these "demands" are conceptualized and defined in different manners by different researchers (see pp. 37-52).

Chess and Thomas (1977) formulated the goodness of fit concept as the result of a review of all subjects in the NYLS clinical case records. In their (1977) book, Chess and Thomas explained this process.

> We pulled out at random from our NYLS records, one group with well-functioning subjects, then one group with behavioral disturbance cases. The three of us sat down and reviewed each record, analyzing the nature of the interactional process in the healthy children versus the clinical cases. The analysis of goodness of fit with the healthy children, and by contrast, a poorness of fit with the

> disturbed children confirmed the usefulness of the concept for every subject. (p. 14)

After they developed this organizing principle, Chess and Thomas (1977) looked for evidence of goodness of fit in some of their other studies. These included a sample of 95 children of working class Puerto Rican parents who were followed from early infancy to adolescence (Thomas, Chess, Sillen, & Mendez, 1974), 56 children with mild mental retardation who were followed from 5 to 11 years of age (Chess & Korn, 1970), and a population of 243 children with congenital rubella who were followed from toddlerhood through adolescence (Chess, Korn, & Fernandez, 1971). Notably, data gathered on children and parents of these three groups used the same methodological procedures as the NYLS (see pp. 9-12 for a description) (Chess & Thomas, 1977).

As an example of one of these studies, Thomas, Chess, Sillen, and Mendez (1974) examined the relationship between temperament and behavior problems in a Puerto Rican working-class (PRWC) sample and the NYLS middle- and upper-middle-class sample. As noted above, both studies employed the same methodology. The clinical cases sampled in this study include 31 children drawn from PRWC families and 42 children from middle- and upper-middle-class families from the NYLS. Notably, a greater percentage of the middle class (NYLS) children were diagnosed as having behavior problems by the time the children were 9 years of age as compared with those in PRWC sample (31% versus 10%). When discussing potential explanations for this difference, Thomas et al. (1974) offered the following:

> It is possible that the earlier presentation of symptoms in the middle class group also stemmed from the greater demands for task performance made upon them

> both a home and in school. Their parents emphasized the early accomplishment of self-care activities, particularly feeding and dressing…. In addition, many adaptations to new situations and new people were required of the preschool middle-class children because their involvement in nursery school and kindergarten brought them into contact with unfamiliar experiences…. In contrast, the demands of the PRWC children in the first 5 years of life and their extrafamiliar experience with new people and new situations were much more limited. (p. 57)

In summary, the goodness of fit construct is central to Thomas and Chess' (1968) formulation of temperament and its role in child development. Put simply, goodness of fit results when the demands and characteristics of the environment are in accord with the child's temperament and other characteristics and its presence posited to lead to positive developmental outcomes. Notably, this concept was formulated retrospectively based on clinical records and a process of inductive analysis (Chess & Thomas, 1977). As a result, the researchers never fully defined or operationalized it as a construct for purposes of experimental replication and extension. Accordingly, methodological difficulties and inconsistencies have plagued this area of research, with the measurement and conceptualizing of fit proving to be most challenging (Windle & Lerner, 1986). As a result, investigators have operationalized goodness of fit in several different manners. Specifically, goodness of fit has been operationalized and studied in terms of environmental demand factors of caregiver characteristics (i.e., caregiver temperament) (e.g., Sprunger, Boyce, & Gaines, 1985; Wallander et al.,1988), environmental demand factors of caregiver expectations of children's behavior (i.e., the discrepancy between

caregiver's ratings of children's actual and ideal temperaments) (e.g., Klein & Ballentine, 2001; Lerner, 1983; Palermo, 1982, Windle & Lerner, 1984), and environmental demand factors of characteristics of the setting (i.e., structure of the setting, stability of setting, physical characteristics of setting) (e.g., De Schipper et al., 2004; Klein, 1982). Given this study's focus, I will discuss manners of operationalizing goodness of fit in the classroom environment. However, prior to this discussion, it is necessary to provide an introduction to this topic of interest.

Goodness of Fit and Schooling

While there is a good deal of research that has demonstrated the ability of temperament to predict children's adjustment and social behavior in school settings (e.g., Garrison, Earls, & Kindlon, 1984; Jewsuwan et al., 1993; Martin et al., 1983; Paget et al., 1984), relatively little research has been directed at understanding the interaction between child temperament and the school environment. The studies that considered temperament alone in relation to child adjustment typically indicate that high adaptability, positive mood, and low emotional intensity are related to children's social adjustment, based on both classrooms observations and teacher reports (Martin et al., 1983; Slee, 1986). While the goodness of fit concept is an intuitively appealing way of thinking about the relationship between temperament and school performance and adjustment, this construct has not received the necessary empirical scrutiny (Keogh, 1986).

"Children's functioning at school is influenced by the congruence between temperamental characteristics and characteristics and contextual requirements, rather than temperament style by itself" (Sanson, Hemphill, & Smart, 2004, p. 157). Accordingly, characteristics of the school and classroom environment must be considered as powerful

contexts in which children are situated. In this respect, the classroom can be thought of as a series of related activities, each containing explicit or implicit expectations for learning, increasing maturity or independence, and appropriate social interactions with teacher and classmates (Keogh, 2003). Thus, this conceptualization requires the specification of the types of tasks and interpersonal demands and expectations within classrooms that interact with children's individual temperament characteristics. A subsequent section, which focuses on the different manners that investigators have operationalized goodness of fit in school settings, will discuss these interpersonal and task demands.

In her book, *Temperament in the Classroom*, Keogh (2003) noted that that several characteristics of classrooms are important in the fit between students and classroom environments. This section of the chapter will elucidate these elements of classrooms. However, readers should note that these described classroom characteristics are ideational propositions rather than research-based conclusions. Nevertheless, these will be discussed and will later be elaborated upon.

Keogh (2003; Keogh & Speece, 1996) first noted that the content and nature of the curriculum may relate to the goodness of fit between a child's temperament and the classroom environment. She explained that teachers organize their classrooms according to their views about how instruction should be delivered (e.g., through exploration, cooperative learning, or direct instruction). It is here that Keogh (2003) believed that the match between students' cognitive abilities and subject matter skills and the level and content of the instructional program affects learning and behavior. In a review of research on school effects, Speece (1992) argued that children in the same classroom experience different educational programs, and are exposed to different content as well as different

instruction. In fact, Cooper and Speece (1990) found associations between first-grade children's personal characteristics and school placement decisions.

Second, Keogh (2003; Keogh & Speece, 1996) noted that the organization of space, time, and resources may also interact with children's temperaments. Every classroom has many ongoing activities taking place, and individual children will respond to various conditions in the classrooms differently, depending in part on their temperament (Keogh, 2003). While neither Keogh and Speece's (1996) nor Keogh's (2003) discussion of goodness of fit offered this example, a study conducted by Klein (1982) provided some support for this notion by comparing parent-reported temperaments of 52 low socioeconomic Head Start preschools and 23 middle socioeconomic kindergartens on teacher judgments of school adjustment. The Thomas and Chess (1977) Parent Temperament Questionnaire measured temperament. The Behar Preschool Behavior Questionnaire as well as a teacher-rated global adjustment score assessed child adjustment characteristics. Educational settings differed with respect to program structure and goals, with the Head Start setting being less structured and more permissive (the author inferred these characteristics, they were not directly measured). Correlations and linear regression indicated that different temperament characteristics were associated with school adjustment depending on the characteristics of the setting. For the Head Start preschoolers, high persistence and intensity related to poor adjustment, whereas, kindergarteners with high thresholds of responsiveness, low persistence, and high levels of withdrawal responses had like adjustment scores.

Third, Keogh (2003; Keogh & Speece, 1996) asserted that the nature of interactions between students, peers, and teachers may affect goodness of fit. Classrooms

are public places where students must take risks, and over time, students become exceedingly aware of one another's successes and failures. Children's temperaments interact with the public nature of the classroom as well as with the temperaments of their peers. Moreover, a poor fit may result when a child's temperament is discrepant from the teacher's view of what makes a child teachable and teachers may also differ in how they interpret a child's temperament (Keogh, 2003).

The role of temperament in teachers' views has in fact been demonstrated in research that links students' temperaments to teachers' ratings of students' teachability (Keogh, 1994). In a study of 360 children in different school placements (i.e., general elementary school, general preschool, special education elementary school, special education preschool), students whom teachers rated high on teachability (on the Teachable Pupil Survey; (Kornblau, 1982)) had positive temperament profiles (as rated by teachers on the Teacher Temperament Questionnaire-Short Form). In contrast, teachers rated those with negative temperament profiles (i.e., low task orientation, low personal-social flexibility) to be less teachable.

Clearly, the goodness of fit between a child's temperament and the classroom environment will affect child outcomes in school settings. The majority of the extant research on goodness of fit in schools, however, focuses on students' academic achievement. This research has shown that when there is a good fit, achievement is enhanced (e.g., Chess, Thomas, & Cameron, 1976; Keogh, 1983; Lerner, 1983; Lerner, Lerner, & Zabski, 1985; Lerner et al., 1986; Pullis & Cadwell, 1982). As one example of this research, Lerner (1983) found that children with high correspondence between self-ratings of temperament and teacher ratings of desirable temperament characteristics

achieved at higher levels and were rated by teachers to be of higher ability than those children with a poorer fit (see p. 40).

Despite this focus on academic achievement, others have posited goodness of fit to predict a number of functional outcomes across settings, including personal and social adjustment (e.g., Carey; 1998; De Schipper et al., 2004). A good fit between a child's temperament and his or her environment may manifest itself in the child demonstrating higher levels of social competence, while a poor fit between a child's temperament and his or her environment may evolve into externalizing behavior problems that bring the child into conflict with others. A subsequent section will demonstrate this relationship. However, it is first necessary to review the different manners in which researchers have operationalized goodness of fit in school settings.

Operationalizing Goodness of Fit in School Settings

Despite the importance of the goodness of fit model for their conceptualization of temperament, Thomas and Chess (1977; 1995) never fully defined or operationalized this construct (Seifer, 2000), as the concept was defined and through clinical examples and broad generalizations. Accordingly, methodological difficulties and inconsistencies have beleaguered this area of research, with the measurement and conceptualizing of fit proving to be most challenging (Windle & Lerner, 1986). As a result, investigators have operationalized goodness of fit in several different manners. Specifically, researchers have operationalized and studied goodness of fit in terms of environmental demand factors of caregiver characteristics (i.e., caregiver temperament) (e.g., Sprunger et al., 1985; Wallander et al.,1988), environmental demand factors of caregiver expectations of children's behavior (i.e., the discrepancy between caregiver's ratings of children's actual

and ideal temperaments) (e.g., Klein & Ballentine, 2001; Lerner, 1983, Palermo, 1982, Windle & Lerner, 1984), and environmental demand factors of characteristics of the setting (i.e., structure of the setting, stability of setting, physical characteristics of setting) (e.g., De Schipper et al., 2004; Klein,1982). This section will present these manners of operationalizing goodness of fit, specifically as they relate to school settings.

Environmental demand factors of caregiver characteristics. Research that considers various characteristics of caregivers as demand factors in the environment operationalizes goodness of fit in terms of behavior matching. Investigators compare caregiver and child characteristics (e.g., child and caregiver temperament) for level of fit. These matching paradigms are generally discussed in terms of statistical interactions of child temperament by some contextual factor used in the prediction of child outcomes (e.g., behavior) (Bates, 1989). Notably, studies that conceptualized goodness of fit in this manner are confined to goodness of fit in the child-parent-family context (e.g., Sprunger et al., 1985; Wallander et al., 1988). In addition, this research that compared the fit between child and parent temperaments did not find that goodness of fit so defined was predictive of behavior problems (e.g., Wallander et al., 1988).

For example, in a study of temperament characteristics and their goodness of fit as potentially important moderators of adjustment in 50 children with congenital physical disabilities, Wallander et al. (1988) did not find support for the goodness of fit model defined in such a manner. They measured maternal and child temperament using two parallel forms of the DOTS (Lerner et al., 1982) and assessed child adjustment using the parent-completed CBCL (Achenbach & Edelbrock, 1983). They tested the goodness of fit model in two manners. An examination of the increment in explained variance in child

adjustment resulting from the interaction of child and mother temperament characteristics (a behavior matching technique) resulted in no significant findings. However, combining mother and child temperament characteristics to predict adjustment (an additive contribution model of child and maternal temperament to adjustment) produced some significant multiple correlations, in terms of their maximized unique contributions to adjustment.

Others using similar paradigms also did not find support for the goodness of fit model conceptualized in this respect (e.g., Sprunger, et al., 1985). It may be that this approach for conceptualizing goodness of fit yields insignificant findings due to the fact that it "generally does not fully capture the theoretical sprit of the Thomas and Chess construct, being more concerned with exact comparability rather than the developmental implications of degree of match" (Seifer, 2000, p. 270).

Environmental demand factors of caregiver expectations. A number of studies investigating the relationship between child temperament and behavior and adjustment operationalized goodness of fit as the discrepancy between teacher ratings of children's actual and ideal temperaments. Thus, these studies consider the interaction between children's temperaments and teacher's expectations for children's temperaments. In other words, these studies inferred goodness of fit from a difference or discrepancy score between the caregiver's attitudes about desirable temperament and the child's actual temperament characteristic.

Lerner and colleagues (e.g., Lerner, 1983, Palermo, 1982, Windle & Lerner, 1984) measured the additudinal/expectational demands held by the significant persons in a child's life by obtaining a demand score for each attribute measured by the DOTS

(Lerner et al., 1982). They did this by using a version of the DOTS that words each item in the form of an attitudinal demand. In this respect, peer demands on the DOTS are phrased, "I think children should usually….;" where teacher demands are phrased, "I expect my students to be;" parent demands are worded, "I expect my child to be;" and perceived demands are phrased, "My parents/teachers/friends want me to usually" (Lerner, 1983). Lerner and colleagues employed stepwise multiple regression analyses to explicate the effects of difference scores (between actual and ideal temperament) on various outcome measures including academic achievement, social competence, and adjustment in elementary, middle school, and high school students.

For example, in a study of the role of temperament in psychosocial adaptation in early adolescents, Lerner (1983) operationalized fit in the above manner, and assessed both temperament as well as demands for behavioral style in the classroom maintained by each student's teacher and peer group using the DOTS (Lerner et al., 1982) for 110 students and 4 teachers located in a large suburb of New York City. To assess goodness of fit, all participants received four sets of discrepancy scores, one for each context: perceived peer, perceived teacher, actual peer, and actual teacher. Within each context there were nine different discrepancy scores, one for each of the NYLS temperamental characteristics. Lerner used the absolute value of the difference between a participant's *z*-score for each attribution and the deviation score of the demand that pertained to that attributed to make up the set of discrepancy score for all participants across contexts. Outcome measures included 25 items from Coppersmith's Self-Esteem Inventory (SEI; Coppersmith, 1967), an index of positive and negative peer relations developed by Lerner and Lerner (1977), teacher ratings of students' academic and social competence on a 7-

point Likert scale, and students' current grades. Results indicated that students whose temperaments best matched demands had favorable teacher ratings of adjustment and ability, better grades, more positive peer relations, fewer negative peer relations, and more positive self esteem that did students who temperaments were less well matched with either teacher and/or peer demands. In addition, fit in one context predicted fit in the other context.

Others have used similar paradigms to measure the relationship between goodness of fit and school adjustment. In a study examining teacher-child and teacher-parent relationships and their impact on child outcomes within Head Start settings, Churchill (2003) found significant positive correlations between goodness of fit and child cognitive and social outcomes in a sample of 45 parents and three teachers. The author measured goodness of fit by an intraclass correlation between the teacher's ratings of their expectations of children's temperament and the teacher's rating of the child's actual temperament. Specifically, Churchill used a version of the Temperament Assessment Battery (TAB; Martin, 1996) to assess both teacher and parent's expectations for children's behavior. The directions for the questionnaire asked informants to rate each item as to how "okay" it was for a child to be that way. For example, a question might state, "How ok is it if a child runs rather than walks everywhere?" The directions further instructed informants to rate what they would tolerate, not what they though was ideal. Teachers also rated the children's actual temperaments using the TAB. Child outcomes were measured by observing children in Head Start classrooms using the MAPS Developmental Observational Scales (Bergen et al., 1992), an activity based developmentally appropriate approach to the assessment of young children's abilities to

assess child outcomes where this study used three subscales: math, preliteracy, and social skills.

Results (Churchill, 2003) indicated that both child cognitive and social outcomes were significantly correlated with teacher-child fit, where fit (measured through intraclass correlation of teachers ratings of children actual temperament sand teacher's expectations for children's temperaments) was positively correlated with the MAPS subscales of math (r = .36, p < .01) and social development (r = .42, p < .01) (the author did not provide the respective degrees of freedom). However, given that this study simply employed correlation analysis to measure goodness of fit, questions remain about both the construct validity of difference scores reflecting temperament within a goodness of fit framework and their minimal variability. In addition, this study measured teacher's tolerance for temperamental deviations, rather than expectations for temperament, calling the operationalization of goodness of fit in this respect into question.

In another study testing the influence of cultural settings on the relationship between temperament and adjustment in children, Klein and Ballentine (2001) found several relationships in the data from a sample of 180 elementary-school children and 30 teachers. The authors measured individual student temperament as well as one ideal temperament profile per classroom using the Teacher Temperament Questionnaire (described on p. 17) (TTQ; Thomas et al., 1968). To measure teachers' expectations of children's temperament, the investigators asked teachers to think of the temperament characteristic of a hypothetical child whom they believed would make the best possible adjustment to their classrooms and then asked the teachers to keep this hypothetical child in mind as they completed one TTQ for Ideal Temperament Judgment. In addition,

teachers completed an Adjustment Ranking scale, where they ranked children on three dimensions (forming healthy relationships with peers; fitting in appropriately with curriculum, routines, and schedules; and formatting age appropriate relationships with teachers and staff). Results indicated that teachers viewed children as having better adjusted to school if the children's actual temperament profiles were more similar to ideal profiles (i.e., higher in persistence, adaptability, and approach and lower in activity). However, a multiple regression analysis that considered temperament predictors of adjustment only examined actual temperament rankings, therein calling both the generalizability of study's results and operationalization of goodness of fit into question.

In summary, studies that operationalized goodness of fit as matching expectations and behavior typically involved measuring the degree of behavioral match with teacher or peer expectations as they related to some measure of competence or adjustment (e.g., Churchill, 2003; Klein & Ballentine, 2001; Lerner, 1983, Palermo, 1982, Windle & Lerner, 1984). However, this view has been criticized in several respects. First, this view is thought to become problematic because the ideal score becomes close to a constant, and the fit, or discrepancy score, can be nearly identical to the actual temperament score (Sanson, Hemphill, & Smart, 2004). In addition, the ideal temperament scores from which the difference score was derived generally is limited in variability, thus limiting the difference scores' reliability (Plomin & Daniels, 1984). It may be that this limited variability suggests the likelihood of a priori attitudes that teachers hold about preferred temperament characteristics. Furthermore, "in general, the ability to predict outcomes for these goodness of fit measures has been modest, with little unique predictive variance that is independent of ordinary temperament ratings" (Seifer, 2000, p. 270). While

operationalizing goodness of fit in terms of the discrepancy between ratings of actual and ideal temperaments is an interesting theoretical concept, as teacher's expectations of children's behavior certainly may be an aspect of the degree of fit between students and their school environment, the construct validity and overall clinical applicability of this approach is lacking.

Environmental demand factors of characteristics of the setting. A handful of investigators have examined the goodness of fit construct by assessing classroom environment characteristics as demand factors in the school setting. Within this class of research, some considered the relationship between temperament, classroom environment characteristics, and child behavior and adjustment to be of primary concern (e.g., De Schipper et al., 2004; Klein,1982) and others measured temperament as a potential moderating variable that must be accounted for (e.g., Hagekull & Bohlin, 1995; Lamb et al., 1988). This second class of studies examined various child and environmental characteristics in relation to behavioral and social outcomes measured temperament (as one of these child characteristics), but viewed it as a possible moderating variable or confound.

Operationalizing goodness of fit in terms of demand characteristics of the classroom environment is consistent with the ideas of Keogh (2003; Keogh & Speece, 1996) who proposed that several aspects of classrooms are important in the fit between students and classroom environments: the content and nature of the curriculum; the organization of space, time, and resources; and the nature of interactions between students, peers, and teachers (discussed on pp. 34-36). This section will first present research that examined the relationship between temperament, classroom environment

characteristics, and behavior, and then will do the same for studies that examined various child and environmental characteristics in relation to behavioral and social outcomes that consider temperament as a moderating variable.

Only a few studies have examined the relationship between child temperament and classroom environment characteristics as it relates to child adjustment and behavior. A study conducted by Klein (1982) that examined relationship between children's temperament and adjustment to kindergarten and Head Start settings, indicated that classroom environment characteristics may play a role in goodness of fit. In a sample of 52 kindergarten students and 23 Head Start students, Klein assessed temperament using the Thomas and Chess Parent Temperament Questionnaire (TTQ Thomas and Chess (1977). This questionnaire has 72 items that are equally divided among the nine temperamental traits. She presented these items verbally to parents who rated each item on a 7-point Likert scale from 1 (hardly ever) to 7 (almost always). Child adjustment was measured using the Behar Preschool Behavior Questionnaire (Behar & Stringfield, 1974), which contains 30 behavioral descriptions scored on three behavioral dimensions: Hostile-Aggressive, Anxious, and Hyperactive-Distractible. In addition, the investigator derived an overall Behar Adjustment Score, with a high score indicating high levels of overall problems. Teachers also provided a global rating of the child's adjustment, by answering the question, "How does this child's adjustment to kindergarten (or Head Start) compare to other children of his or her age and experience?" on a 5-point scale from 1 (very well adjusted) to 5 (poorly adjusted). Results of study indicated that situational characteristics are likely to affect the relationship between temperament and behavior. In the kindergarten setting, children who were rated by teachers as having high

sensory thresholds were viewed as having the greatest adjustment difficulties. In addition, children in this setting who were rated by teachers as having low levels of persistence and approach responses were also rated as having a higher frequency of behavior problems. However, in the Head Start setting, children perceived as being high in persistence or high in intensity of reactions were considered to be more poorly adjusted.

While Klein (1982) did not use a measure of the classroom environment, in her analysis of the findings, she suggested that kindergarten settings may be more academically focused and include more teacher-controlled activities than Head Start settings, which she noted may be more socialization focused and grant the children more freedom to self-select activities. Therefore, while this study provided some indication that school and classroom environment characteristics may influence the expression of the relationship between temperament and behavior, Klein did not directly measure these characteristics, indicating the need for additional validation of this relationship.

In a study that assessed classroom environment characteristics directly, De Schipper et al. (2004) applied the concept of goodness of fit within day care settings in a sample of 186 children, ages 6-30 months to address possible relations of children's temperament with their social-emotional wellbeing and problem behavior. The authors measured goodness of fit in terms of the characteristics of the daycare program using the Leiden Inventory for Child's Wellbeing in Day Care (LICW-D; Van IJzendoorn et al., 1998a), the Leiden Inventory for Daily Stability in Center Care (LIDS; De Schipper, Van IJzendoorn, & Tavecchio, 2004), and child-professional caregiver ratios. The LICW-D is a 28 item scale, adapted from the Wellbeing Scale used in a previous study (Van Ijzendoorn et al., 1998b) where respondents rate four aspects of wellbeing in daycare

(general wellbeing, wellbeing in the presence of caregivers, wellbeing with group members, and wellbeing within the physical care setting) on a 6-point Likert scale. The LIDS focuses on children's daily experiences of stability and includes two scales: daily caregiver availability and daily arrangement stability. The daily caregiver availability scale includes seven items regarding the availability of trusted caregivers, and the daily arrangement stability scale includes three items that indicate the number of childcare arrangements in addition to daycare center case.

De Schipper et al. (2004) assessed temperament using the Infants Characteristics Questionnaire (ICQ; Bates, Freeland, & Lounsbury, 1979), a parent-completed measure that consists of 33 7-point Likert type items and four subscales based on several theoretical models of temperament: Fussy/Difficult, Unadaptable, Dull, and Unpredictable. They assessed child behavior using the Child Behavior Checklist Teacher Report Form (ASEBA TRF; Achenbach, 1997), using the children's internalizing, externalizing, and total problem behavior scores as indicators.

De Schipper et al. (2004) used multivariate regression analysis to explore child adjustment to daycare and children's wellbeing in daycare. They explored hypotheses using a multivariate approach in which variables were entered in three or four blocks for each child adjustment variable: gender, child temperament, child care characteristics, and interaction terms (temperament x caregiver availability, temperament x arrangement stability). Results indicated partial support for the goodness of fit hypothesis in several respects. First, the stability of day care effected children of easy and difficult temperamental styles in different manners, where more unstable care arrangements negatively effected temperamentally difficult children more than it did temperamentally

easy children (r (72) = 0.10, $p < 0.41$ *vs.* r (110) = -0.37, $p < 0.001$). Second, when children had access to trusted caregivers, there were stronger associations between temperament and social-emotional wellbeing in temperamentally easy children than in difficult children (r (38) = -0.10, $p < 0.54$ *vs.* r (51) = -0.53, $p < 0.001$). Overall, these results indicated that classroom environment characteristics interact with children's temperaments to predict behavioral outcomes. In addition, these results suggested that children with the difficult temperament profile may be especially vulnerable to experience stress within daycare settings.

Notably, the De Schipper et al. (2004) study is unique in several respects. First, its examination of the goodness of fit model directly considered classroom environment characteristics as they interact with children's temperaments. Second, the authors employed multivariate regression analysis to test this relationship, where the interaction terms of temperament by classroom environment characteristics were primary measures of goodness of fit. Finally, the study integrated the classroom quality literature to identify which classroom environment characteristics may interact with child temperament. The present study also integrated some of these unique methodological components.

As noted, another class of studies assessed how various child characteristics interact with early childhood environments to predict behavioral and social outcomes. These studies will be discussed as part of the goodness of fit literature because they offer insights into the relationship between demand features of the classroom/school setting and child temperament. However, this research is first and foremost concerned with how the classroom quality of early childhood environments (i.e., preschool, prekindergarten, daycare) contributes to children's positive developmental outcomes. A presentation of

this body of literature can be found on pages 53-65. Temperament is sometimes also measured in the classroom quality research to examine its role as a potential moderating variable (e.g., NICHD Early Child Care Research Network, 1998; NICHD Early Child Care Research Network, 2003; Zajdeman & Minnes, 1991). Specifically, these studies typically view temperament as a background factor that is assumed to be one of several potential moderating factors of care quality effects with either buffering of potentiating effects on social and affective outcomes (Hagekull & Bohlin, 1995). This section will provide examples of studies that consider the moderating role of temperament in this area of research.

Hagekull and Bohlin (1995) conducted a longitudinal study that considered the effects of day care quality in interaction with child and family characteristics. The authors also measured goodness of fit in terms of the characteristics of the daycare program. In addition, they assessed two child characteristics, gender and temperament (where child temperament was conceptualized as child manageability).

Hagekull and Bohlin (1995) assessed temperament using The Toddler Behavior Questionnaire (TBQ; Hagekull & Bohlin, 1981), at 29-months and at 4-years of age, and they measured quality of care in day care settings using an instrument designed specifically for the study. This measure examined several aspects of quality including levels of stimulation, discipline methods, physical arrangements, and emotional climate between children as well as between children and caregivers. Investigators assessed child socioemotional development at 29-months using an instrument developed by the authors that assessed externalizing behavior, internalizing behavior, and positive socioemotional behavior, and at 4 years using another instrument developed by the authors that measured

behavior problems, social competence, and ego strength/effectance. While the relationship between temperament and behavior was not a particular focus, one of several findings was that temperamentally easy toddlers showed reduced aggressiveness at 4 years-old when they experienced high quality day care, while difficult children did not demonstrate reduced aggressiveness in high quality settings. In their discussion of these findings, the authors explained that this might be an effect of what Thomas and Chess (1977) described as a possible developmental course for an easy child and on the other hand, nonmalleability in temperamentally difficult children (Hagekull & Bohlin, 1995).

Lamb and colleagues (1988) examined the effects of out-of-home care on the development of social competence in Swedish preschoolers. These authors considered one aspect of temperament (a temperamental difficulty factor) as a child background factor. They used Rothbart's Infant Behavior Questionnaire (IBQ; Rothbart, 1981) from which the researchers only used one composite measure (a perceived difficulty score). Specifically, *t*-tests determined if group differences existed in this background factor across three settings (home settings, family daycare settings, and daycare centers). Given that *t*-tests did not indicate significant differences in this factor across these groups, Lamb et al. did not explore the role of temperament in the relationship between the daycare setting and social and affective outcomes further.

In another study that examined the relationship between daycare quality and children's emotional expression, Hestenes, Kontos, and Bryan (1993) viewed temperament as a variable that must be taken into account because children's emotional responses are a reflection of both internal (i.e., temperament) and external (i.e., child care environment) influences. Thus, they measured temperament as a potential moderating

variable to ensure that intrapersonal influences (temperament) were not mistakenly attributed for the cause of children's emotional expressions. In addition, these investigators looked at temperament because the impact of the environment on children's emotional expression may be moderated by temperament (Goldsmith et al., 1987).

Sixty children, ages 3 to 5, their mothers, and their day care teachers participated in the study (Hestenes et al., 1993). The authors measured child temperament using an abbreviated form of the mother-completed Behavior Style Questionnaire (McDevitt & Carey, 1978). They assessed the quality of the daycare environment using the Early Childhood Environment Scale (ECERS; Harms & Clifford, 1980), as well as an observational measure of teacher engagement developed by Howes and Stewart (1987). Children's emotional expression was measured using an observational affect code developed by Caruso and colleagues (1991), where facial expressions were coded in terms of quality, intensity, and duration. Results indicated that temperament did not moderate the relation between daycare environment and children's display of affect. However, the study used only one temperamental variable (intensity) throughout analyses as a moderating variable because this was the only temperamental variable that significantly correlated with affect. Moreover, the authors noted that, "it is possible that this lack of significant findings was due to methodology…. [and that] in the future it may be beneficial to have the out-of-home caregiver's view of the child's temperament" (Hestenes et al., 1993, p. 305). In addition, while the authors employed multiple regression analyses, given temperament's role as a designated moderating variable, temperament by environment interaction terms were not considered.

Studies that examine temperament as a moderating variable in the classroom quality literature offer limited insights into the relationship between child temperament, demand factors of the classroom environment, and child behavior. If anything, they indicate the need to explore this relationship further and suggest that temperamentally easy children may benefit more from high quality educational programs.

Overall and in sum, studies that measure classroom environment characteristics as an aspect of goodness of fit measure characteristics like program quality, stability, emotional climate, physical arrangements, stimulation, and discipline methods. This type of methodology is in accord with the ideas of Keogh (2003; Keogh & Speece, 1996) who proposed (albeit without providing the necessary empirical support) that several aspects of classrooms are important in the fit between students and classroom environments: the content and nature of the curriculum; the organization of space, time, and resources; and the nature of interactions between students, peers, and teachers. The results of the small number of studies that have assessed these characteristics typically indicate that high quality educational programs are of greater benefit to temperamentally easy children than to difficult children. Thus, this operationalization of goodness of fit appears to be promising because it offers a framework for understanding how child characteristics interact with the classroom environment to predict behavioral and affective outcomes. However, it must be said that only a handful of studies implemented this sort of operationalization. Therefore, I looked to the early childhood classroom quality literature to determine the best way to measure classroom environment variables that may relate to child temperament to predict behavioral and social outcomes.

Classroom Quality as it Relates to Child Behavioral and Social Outcomes

This body of literature examines components of early childhood environments (i.e., preschool, prekindergarten, daycare) that relate to positive child outcomes in academic and social domains, such as social behavior, cognitive development, language development, compliance, and self-regulation (Hestenes, Kontos, & Bryan, 1993; Pianta et al., 2005). Accordingly, both the classroom quality and goodness of fit literatures seek to identify elements of school and classroom environments that are associated with such positive developmental outcomes. As previously noted, the current study is particularly interested in children's externalizing behavior problems and social competence.

Social competence as a general developmental construct is multifaceted and has many definitions (Howes, 1987). Within the classroom quality literature, investigators generally consider social competence to extend to successful social functioning with peers and include prosocial behavior, complex play, and gregarious behaviors (Howes, 1987; Howes & Hamilton, 1993). Thus, social competence is generally measured in terms of the actual social skills a child displays (e.g., Howes, 1987; Howes & Hamilton, 1993). On the other hand, these same authors consider the lack of social competence to include aggressive and withdrawn behaviors (Howes & Hamilton, 1993). Thus, social competence is manifested in both child behavior and affect. The development of social competence occurs within a social system composed of the child's caregivers and peers in the primary settings in which they interact (Howes, 1987). Clearly, the prekindergarten environment is one of these primary settings.

Studies that measure classroom quality of prekindergarten environments typically assess two aspects of classrooms: structural markers of quality and process quality (e.g.,

Early et al., 2007; Hagekull & Bohlin, 1995; Head Start Research Consortium, 2003; Hestenes, Kontos, & Bryan, 1993; Lamb et al., 1988; Lambert, Abbott-Shim, & McCarty, 2002; LoCasale-Crouch et al., 2007; Pianta et al., 2005; NAEYC, 1997). The following section will provide a brief review of these manners of assessing of classroom quality and focus on how they relate to behavioral and social outcomes in children. Notably, this review must be considered as part of the broader discussion of classroom environment characteristics of early childhood environments that may relate to child temperament to predict child behavior and social skills.

Structural Markers of Classroom Quality

A number of structural markers of early childhood environments are posited to be associated with positive social and behavioral outcomes in children (e.g., Lambert et al., 2002; LoCasale-Crouch et al., 2007; Pianta et al., 2005; Rice, 2003). Structural markers of quality include static variables like teacher credentials and education levels, teacher-child ratios, and program location. Investigators obtain structural information through surveys and questionnaires where teachers and program administrators serve as respondents. The literature includes two broad categories of structural markers of classroom quality: program variables (under which classroom variables are subsumed) and teacher variables (Lo-Castle-Crouch et al., 2007).

Program variables. Program variables connected with classroom quality include the number of students enrolled, adult-child ratios, program location, program day length, and adequate classroom space (e.g., Lambert et al., 2002; LoCasale-Crouch et al., 2007; Pianta et al., 2005). The National Institute of Early Education Research and The National Association of Education of Young Children identify both number of students enrolled

and teacher-child ratios as contributors to classroom quality that relate to positive child outcomes (LoCasale-Crouch et al., 2007). In addition, research also indicates that higher teacher-child ratios increase teacher detachment and harsh behavior (Whitebook et al., 1990) as well as teachers' controlling behavior (Kontos & Fiene, 1987).

Despite this evidence for the contribution of program variables to quality and children's social competence, a recent study of classroom quality and program characteristics in a sample of 238 classrooms representing six states' prekindergarten programs found small, but not clear cut, relationships between observed quality and location of the program, child-staff ratios, and length of day (Pianta et al., 2005). Investigators used three assessment systems of classroom quality: the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms et al., 1998) (discussed below); the CLASS (Pianta, La Paro, & Hamre, 2007), which assess the emotional climate of schools; and the Emerging Academic Snapshot, a 27-item instrument that observers code within a 20-second period. To determine if structural variables were associated to observed classroom quality, the authors entered structural variables as prediction variables in a hierarchical regression equation, where composite scores on the ECERS-R and CLASS were used as outcome variables. As noted, program variables were significantly, albeit very modest, predictors of observed quality. Specifically, program location, child-to-staff ratio, and length of day had no relation to quality. The results suggested that classroom quality was most related to proximal teacher attributes (i.e., demographics, attitudes/beliefs, depressive feelings) and child characteristics (i.e., socioeconomic status). Given these conflicting findings, it appears necessary for

additional research to determine the extent to which structural markers of program quality relate to classroom quality and higher levels of social competence in children.

Teacher variables. A number of studies have also connected staff characteristics such as training and education levels of teachers to the quality of the prekindergarten classroom environment. Rice (2006) identified five broad categories of measurable and policy-relevant indicators of teacher quality that relate to enhanced child outcomes in academic and social domains: teacher experience, teacher preparation programs and degrees, teacher certification, teacher coursework, and teachers' own test scores. In the classroom quality literature, most research indicates a link between higher levels of teachers' education and higher classroom quality (e.g., Burchinal et al., 2002; Howes et al., 1992), yet other studies have found largely null (e.g., Early et al., 2006) or inconclusive (e.g., Pianta et al., 2005) associations of teacher education with classroom quality. On the other hand, research has consistently demonstrated that adults without education in child development tend to engage in behaviors that may impede child wellbeing (Doherty-Derkowski, 1995). This relationship has been found both internationally and in the United States, and has been replicated across different types of classroom settings (Arnett, 1989; Pence & Goelman, 1991; Whitebook et al., 1990).

Based on these conflicting findings, Early and colleagues (2007) suggested that the lack of common definitions of teacher education and training, coupled with the use of different controls in different studies severely limit the ability to draw clear conclusions from the existing literature. Therefore, they conducted a study that allowed for the direct testing of the effects of teacher education on classroom quality using value-added specification. They analyzed several large data sets (Early Head Start Follow-Up, Head

Start Family and Child Experiences Survey, Georgia Early Care Study, More at for Evaluation, National Center for Early Development and Learning, Study of Early Child Care and Youth Development, and Preschool Curriculum Evaluation Research Program) using similarly defined variables and equivalent models specification to answer a common set of questions. Results indicated inconsistent and largely null findings, and thus did not provide compelling evidence of an association between structural markers and improved child outcomes (Early et al., 2007).

Given the largely inconsistent findings for an association between structural markers of classroom quality and improved outcomes in both child academic and social domains, it appears that while some structural indicators (i.e., teacher credentials and education levels, teacher-child ratios) provide some sense of the classroom resources available that effect child outcomes, they only can be considered as proxy quality measures because they do not provide a complete description of the classroom environment. In addition, it is important to note that measures of process quality (discussed on pp. 58-64), like the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Cryer, & Clifford, 1998; discussed on pp. 54-55) are often used as an indicator when this type of research attempts to determine if structural variables are an aspect of classroom quality (Burchinal et al., 2000; Scarr et al., 1994). Thus, it follows that investigators need to identify and measure additional indicators of classroom quality, such as the quality of the interactions between students and teachers as well as the interactions among peers, when behavioral and social outcome variables are of interest.

Process Quality

The ecological theory of development suggests that development occurs through interactions between the individual and the environment, where ongoing interactions with both people and objects shape a child's development over time (Bronfenbrenner & Morris, 1997). Thus, in addition to structural aspects of the classroom environment, one must also consider the daily interactions found in the classroom setting as powerful process variables that relate to children's social development. Thus, process variables encompass many of the most important elements of a high quality program and can be regarded as one of the most important components of classroom quality (Essa & Burnham, 2001).

Process quality in prekindergarten classroom settings involves social, emotional, physical, and instructional elements of interactions with young children (Pianta et al., 2005). Classroom process studies repeatedly identify two general factors that predict children's development: social/emotional climate and instructional support (LoCasale-Crouch et al., 2007). Social and emotional aspects of interactions in the classroom include teacher sensitivity to children's needs and a positive classroom climate. In addition, aspects related to instructional support include highly skilled teachers who monitor student performance and provide additional explanations and ideas to their students (LoCasale-Crouch et al., 2007).

Process quality is typically measured by observing the experiences in the early child program and classroom and rating the multiple dimensions of the program, such as teacher-child interactions, type of instruction, room environment, materials, relationships with parents, and safety routines (Espinosa, 2002). The Early Childhood Environment

Rating Scale-Revised (ECERS-R; Harms, Cryer, & Clifford, 1998), measures many of these dimensions, and is the most commonly used comprehensive measure of process quality in prekindergarten environments that has served as the standard measure in the field of education for over 25 years (Mashburn et al., 2008). The ECERS-R describes the classroom physical environment and materials and the warmth and responsivity of child-teacher interactions, aspects that predict both current and future child competencies (Pianta et al., 2003).

Harms and colleagues (1998) developed the ECERS-R from quality indicators for classroom environments elaborated by The National Association for the Education of Young Children (NAEYC), including small group size, low teacher-child ratios, developmentally appropriate curricula, adequate teacher training, parent-teacher communication, and a safe, clean, and stimulating classroom environment (Schwartz, Garfinkle, & Davis, 2002). Item development also took into account a number of practices recommended by the Division of Early Childhood (DEC) of the Council for Exceptional Children, and thus included an assessment of scheduling, transitions, natural environments, peer interactions, and use of effective instructional strategies (Schwartz et al., 2002). The revised edition includes 43 items organized into 7 areas of center-based care for children aged 2.5 through 5 years, which include personal care routines, space and furnishing, language-reasoning, interaction, activities, program structure, and parents and staff.

A large body of literature indicates that process quality ratings of early childhood programs such as those obtained through the use of the Early Childhood Environment Rating Scale- Revised (ECERS-R) and measures like it are associated with the level of

social functioning in children (e.g., Bryant et al., 1994; Hestenes et al., 1993; Lambert et al., 2002; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001; Vandell, 2004). Using the ECERS-R, Hestenes et al. (1993) examined the relationship between day care classroom quality, children's emotional expression, and temperament among 60 children ages 3 to 5 (discussed on pp. 50-51). Multiple regression analysis revealed that higher classroom quality (described with the ECERS-R total scale score) settings were associated with significantly more positive affect. In fact, children in high quality settings displayed more smiling and laughing, showed a greater intensity of this positive affect, and displayed less intense negative affect than children in lower quality settings. In addition to indicating that classroom quality relates to more positive displays of emotion and affect, the study also found that teacher engagement (defined above) predicated children's affect, but that child temperament did not moderate the relationship between day care classroom quality and children's emotional expression.

Hagekull and Bohlin (1995) investigated the effects of day care quality (as measured by the ECERS) in interaction with child and family characteristics on socioemotional development concurrently at 29 months and longitudinally at 4 years (discussed on pp. 49-50). In a group of 52 children, the authors studied main effects of day care quality and effects due to interactions between such quality and the background characteristics of home environment quality, gender, and infant manageability. The results indicated a main effect of day quality on children's expression of positive emotions, and interactive effects for several of the other indicators of socioemotional functioning (including temperament).

In a study that considered the relation of preschool child-care quality to children's cognitive and social developmental trajectories, Peisner-Feinberg et al. (2001) followed 733 children from ages 4 to 8 years of age. This study was part of the Cost, Quality, and Child Outcomes (CQO) in Child Care Centers Study, a study of center-based community childcare and children's longitudinal outcomes in four states (California, Connecticut, North Carolina, and Colorado) in the United States. The data included measurement of preschool childcare quality and longitudinal assessment of children's language, cognitive, and socioemotional functioning over a 4-year period. Peisner-Feinberg and colleagues measured childcare quality using several instruments, including the ECERS-R. They measured child outcomes in behavioral, cognitive, and social domains using the Classroom Behavior Inventory (CBI; Shaefer, Edgerton, & Aaronson, 1978) at each of the 4 years. The results of this study indicated longitudinal effects for several developmental domains, including cognitive and attention skills, problem behaviors, and sociability, indicating that children who had better quality preschool experiences were more advanced in their development over a 5 year period.

Caregiver sensitivity. As noted, important features of process quality are the social and emotional aspects of interactions in the classroom. Investigators have identified teacher sensitivity as a key component that contributes to the quality of children's early educational experiences and has been linked to positive developmental outcomes in social domains (Gerber, Whitebook, & Weinstein, 2007; Helburn, 1995; NICHD ECCRN, 2005). "Attention to caregiver sensitivity emerged when child development research gradually began to shift from a focus on children's inborn qualities to the role of children's earliest environments in shaping their developmental trajectories"

(Gerber et al., 2007, p. 328). In particular, the longitudinal research of Werner (1996) indicated that children born into poverty were more resilient at age 3 when more protective factors were present, including caregiver sensitivity, defined in terms of higher levels of warmth and positive interactions.

State quality initiatives, such as those in Colorado, Iowa, Kentucky, Maryland, Montana, North Carolina, Oklahoma, Pennsylvania, Tennessee, and Vermont, often use the Arnett Caregiver Interaction Scale (CIS; Arnett, 1989), an observational assessment of caregiver interaction and sensitivity with children, as a process quality indicator (NAEYC, 2005). Different from the ECERS-R, a measure of broad process quality (i.e., classroom structure, materials, and tone), the CIS rating scale is used to assess teacher's caregiving behaviors. The CIS has 26 items measuring four subscales: sensitivity/positive interaction (e.g., speaks warmly to children), punitiveness (e.g., seems critical of the children), detachment (e.g., doesn't seem interested in children's activities), and permissiveness (e.g., doesn't reprimand children when they misbehave).

The CIS was formulated based on a study conducted by Arnett (1989), who was interested in the level of training caregivers in childcare centers and its effect on their interactions with children in those childcare centers. Arnett studied 59 caregivers, whose education levels ranged from having no training to a four-year early childhood education program or degree in a related field. The CIS was used to assess the caregivers. Results indicated that caregivers with more training were rated higher on the CIS in terms of the quality of interactions with the students. Specifically, caregivers with more training showed more warmth towards the students, were more enthusiastic, and encouraged students to cooperate with their peers. Arnett's (1989) findings, that caregivers with more

training were more likely to engage in sensitive and positive interactions with children in their care have been replicated in several studies (e.g., Burchinal et al., 2002; Ghazvini & Mullis, 2002; Howes, 1997).

The CIS is typically used as a supplemental measure of teacher's sensitivity, one aspect of process quality that has been shown to be related to higher levels of academic and social competency and lower levels of behavior problems (NAEYC, 2005). For example, the Cost, Quality, and Outcomes Study (Peisner-Feinberg et al., 2001; discussed on p. 56) used four summary measures of classroom quality, one of which was the CIS, which was used as a measure of teacher sensitivity. This study indicated that higher levels of classroom quality were related to positive child outcomes in developmental domains, including cognitive and attention skills, problem behaviors, and sociability, indicating that children who had better quality preschool experiences were more advanced in their development over a 5 year period. This finding, that higher levels of teacher sensitivity relates to positive child outcomes in behavioral and social domains has been found across age groups and educational settings (e.g., Howes, Matheson, & Hamilton, 1994; Kontos & Wilcox-Herzog, 1997; Pianta, 1999; Pianta et al., 2002).

Studies that investigated if caregiver training and early childhood center accreditation is associated with higher levels of classroom quality also used the CIS as such an indicator (Cassidy et al., 1998; Kontos, Howes, & Galinsky, 1996; NAEYC, 2005) For example, a study conducted by the National Association for the Education of Young Children (NAEYC, 2005) to determine if their accreditation is linked to higher levels of quality in early childhood environments used the CIS (along with the ECERS-R) as an indicator of process quality. In addition, a study to document the impact of

training and ensuing experience on the knowledge, beliefs, and practices of AmeriCorps national service volunteers in child care classrooms in the state of North Carolina used the CIS as one of two measures of classroom interaction (Cassidy et al., 1998). Also, several studies have used the CIS as a classroom process quality validation measure (e.g., De Kruif et al., 2000; Maxwell et al., 2001; Zellman et al., 2008) to explore the validity of proposed classroom quality assessments systems.

In summary, the process quality of early childhood environments is typically conceptualized in terms of the multiple dimensions of the educational program, such as teacher-child interactions, type of instruction, room environment, materials, and program structure as well as in terms of specific caregiver behaviors, such as teacher sensitivity. No matter how they are conceptualized, process variables have been demonstrated to be related to positive child outcomes in behavioral and social domains (i.e., Bryant et al., 1994; Hestenes et al., 1993; Howes et al., 1994; Kontos & Wilcox-Herzog, 1997; Lambert et al., 2002; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001; Pianta, 1999; Pianta et al., 2002; Vandell, 2004). It is for this reason that studies should measure classroom environment characteristics related to process quality at the classroom level, for it is quite probable that along with child temperament at the student level, they predict social and behavioral outcomes in children.

As noted, the early childhood classroom quality literature was reviewed to supplement the goodness of fit literature's failings to identify specific and direct aspects of classroom environment characteristics that relate to child temperament and predict positive developmental outcomes in behavioral and social domains. This review indicated that both structural and process variables have been measured in early childhood

environments. However, while some structural variables (e.g., teacher credentials and education levels, teacher-child ratios) have been shown to provide some sense of the classroom resources available that affect child outcomes (e.g., Burchinal et al., 2002; Howes et al., 1992), structural variables most likely have a more indirect influence. Process variables, on the other hand, such as teacher-child interactions, type of instruction, room environment, materials, and program structure as well as in terms of specific caregiver behaviors, such as teacher sensitivity have consistently and unequivocally demonstrated to be related to children's positive developmental outcomes (e.g., Bryant et al., 1994; Hestenes et al., 1993; Howes et al., 1994; Kontos & Wilcox-Herzog, 1997; Lambert et al., 2002; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001; Pianta, 1999; Pianta et al., 2002; Vandell, 2004).

A Pilot Study for the Present Dissertation

Blackwell (2008) conducted a pilot study for the present dissertation. The goal of this study was to determine if the relationship between temperament and behavior was expressed differently in dissimilar classroom environments in prekindergarten settings. Two prekindergarten classrooms in a Head Start center consisting 14 and 15 students and their head teachers participated in this study. These classes will be referred to in this discussion as Class 1 and Class 2. Teachers rated child temperament using the Teacher and Caregiver Temperament Inventory for Children (TACTIC; Billman & McDevitt, 1998), a 95-item rating scale that assesses temperament in children two through six years of age. The study examined four temperamental dimensions on the TACTIC (adaptability, activity level, mood, and persistence). The investigator selected these dimensions based on the plethora of research that indicated that certain temperament

characteristics (i.e., negative mood, low adaptability, low persistence, and high or low activity level) predispose a child to behavior and adjustment problems (see pp. 22-24 for further discussion) (e.g., Hagekull, 1994; Teglasi & MacMahon, 1990; Thomas et al., 1968; Wertlieb et al., 1987).

The study measured several classroom environment characteristics based on the recommendations of Keogh and Speece (1996) (discussed on pp. 34-36). Specifically, the investigation used three subscales of the Early Childhood Environment Rating Scales-Revised (ECERS-R; Harms, Clifford, & Cryer, 2005), an observational tool designed to assess the process quality of group programs for children: the Activities, Interaction, and Program Structure scales. In addition, the study also used the Total Scale score and Sensitivity subscale score from the Caregiver Interaction Scale (CIS; Arnett, 1989), an observational assessment of caregiver sensitivity. Direct observations (e.g., teacher-child ratios) and teacher responses to standard demographic questionnaires (e.g., years of education, received degrees, number of years teaching) provided additional information regarding classroom structural variables. A classroom's scores on the three ECERS-R subscales, CIS Total Scale, and Sensitivity subscale scores indicated overall classroom quality along with classroom structural variables (i.e., teacher-child ratios, teacher years of education, teacher received degrees, teacher number of years teaching). The Externalizing Behavior and Social Skills subscales on the Preschool and Kindergarten Behavior Scales- Second Edition (PKBS-2; Merrell, 2002, a 76-item teacher-completed rating scale designed to measure both problem behaviors and social skills of children ages 3-6) measured child outcomes in behavioral and social domains.

The results of this study indicated that the two classes varied on several levels, and specifically that Class 2 was of higher overall quality. In terms of structural markers of quality, Class 2 had a slightly lower child to teacher ratio, and the teacher in Class 2 had several more years of education (20 *vs.* 16 years) and teaching experience (21 *vs.* 18 years) as well as a higher academic degree than the teacher in Class 1 (Masters *vs.* Bachelors degree) (see Appendix A). On measures of process quality (i.e., ECERS-R; CIS), classroom environment ratings were consistent across raters and time (kappa range, .95 – 1.0, $p < 0.001$; r range .92 – 1.0, $p < 0.001$) (see Appendix A). In addition, Class 2 received higher levels of ratings on the CIS, and on two out of three subscales on the ECERS-R (see Appendix A). Thus, on a purely descriptive level, the two classrooms differed from one another, where Class 2 appeared to be of higher quality based on structural and process quality indicators.

Independent sample *t*-tests assessed possible differences between measures of externalizing behaviors and social skills between classrooms. Despite the apparent differences between the two classrooms levels of quality, *t*-test results indicated no significant differences between classes in relation to mean scores of either Externalizing Behaviors (t (27) = -.001, p = .999) or Social Skills (t(27) = 1.23, p = .23) (see Appendix A).

Correlation analyses within classrooms assessed the relationships among the four temperament variables and both externalizing behaviors and social skills. (See Appendix A.) Results revealed positive relationships between all measures of temperament and externalizing behaviors in both classrooms; however the magnitude of these relationships differed greatly. In Class 1, all measures of temperament significantly and positively

related to externalizing behaviors. In Class 2, however, the only relationship to reach significance was that between activity level and externalizing behavior. Results assessing the relationships between the temperament variables and social skills within classrooms revealed significant negative relationships between all measures of temperament and Social Skills in both classes.

To assess if there were significant differences between the strength of correlations between classrooms, all correlations within each class were transformed using the Fisher's *r* to *z* transformation method (Fisher, 1958). This method allows the ability to test whether the correlations computed for each of the classrooms are equal, or if there are significant differences between the strength of the correlations between classrooms. Results revealed significant differences in the strength of correlations between classrooms for the relationships of activity, adaptability, and persistence with externalizing problems (see Appendix A).

In the last stage of analyses, a series of individual regressions were conducted for each classroom using externalizing behaviors and social skills as dependent variables, and the four temperament variables as independent variables. As the sample size for this study was relatively small, individual regression analyses were conducted for each of the temperament variables on the corresponding dependent variables instead of including all four predictors in one model. In Class 1, all four measures of temperament were significant predictors of both externalizing behavior and social skills. In Class 2, results also revealed that all four temperament variables were significant predictors of social skills, however, the only significant predictor of externalizing problems was activity level (see Appendix A).

The results of this pilot study (Blackwell, 2008) indicated that stronger relationships between temperamental variables and child behavior and social skills emerged in classrooms lower overall quality. This finding is in contrast to research that demonstrates stronger relationships between measures of children's temperament and well-being in settings where children have greater availability of trusted caregivers and a lack of such an association in settings where children have less access to trusted caregivers (De Schipper et al., 2004). Nevertheless, the results of the present study are in line with the common argument that schools play in important role in reducing disparities in students' skills and characteristics (Mann, 1965). In addition, the classroom quality literature has indicated that when children at risk are exposed to high levels of observed instructional and emotional support from teachers, rates of behavior, social, and academic problems decrease (Pianta et al., 2005). It may be that classrooms of higher quality serve to equalize disparities among children, where high levels of classroom quality weaken the documented relationship between temperament and child behavior and adjustment as described in the goodness of fit model.

The results of this study were limited in several respects. First, study's sample was extremely small sample and thus limited the statistical analysis that could be performed. Second, the study design did not permit a discriminant analysis of the classroom environments to determine the magnitude of their differences. Third, because the sample included a normal population, differences in the degree of children's externalizing behavior problems and social skills were unlikely between the two classrooms. Despite these limitations, this pilot's findings (Blackwell, 2008) indicated that stronger relationships between temperamental variables and child behavior and social

skills are likely to emerge in classrooms lower overall quality. Additional research is needed to determine the specific nature of the relationship between classroom environment characteristics, child temperament, and child behavior and social skills.

Rationale and Hypotheses

Based on the reviewed literature, it is clear that the interaction of child temperament and classroom environment characteristics relates to various behavioral and social functional outcomes throughout childhood within the school setting, indicating the salience of the goodness of fit model (e.g., Keogh, 1986; Thomas & Chess, 1977; Thomas et al., 1968). In addition, the early childhood classroom quality literature indicates that children in higher quality prekindergarten settings, defined in terms of process variables, demonstrate more social competence than those in lower quality settings (e.g., LoCasale-Crouch et al., 2007; Pianta, 1999). However, no study has fully explored how the relationship between temperament and behavior is expressed in classroom environments that have different characteristics. In addition, while several researchers (e.g., De Schipper et al., 2004; Hagekull & Bohlin, 1995; Klein, 1982) have used classroom environment characteristics as an aspect of goodness of fit, they have provided inconsistent operationalization and measurement of these characteristics.

Purpose

The purpose of this study was threefold. First, it sought to determine how the relationship between temperament and behavior is expressed in different types of classroom environments in prekindergarten settings. Second, it sought to further operationalize the goodness of fit model as it relates to child outcomes in behavioral and social domains in school settings. Third, it sought to determine if process variables

indicative of classroom quality interact with child temperament to predict child behavior and social skills in school settings.

I recruited classrooms in several prekindergarten settings. Child temperament was measured using the Teacher and Caregiver Temperament Inventory for Children, a 95-item rating scale that assesses temperament in children 2 through 6 years of age (TACTIC; Billman & McDevitt, 1998). This dissertation used the Total Temperament score, derived by summing the scale's 95-items, which is a composite measure of temperamental difficulty. I selected the Total Temperament score based on its clinical utility (S. McDevitt, personal communication, October 9, 2007) and based on research indicating that the difficult temperament profile predisposes a child to behavior and adjustment problems (see pp. 24-26 for further discussion) (e.g., Caspi & Silva, 1995; Guerin et al., 1997; Maziade et al., 1990; Teglasi & MacMahon, 1996; Thomas et al., 1968).

To determine the relationship between temperament, classroom environment characteristics and behavior, I measured classroom environment characteristics based on the early childhood classroom quality literature and on the recommendations of Keogh (2003; Keogh & Speece, 1996). This dissertation used the Program Structure scale of the Early Childhood Environment Rating Scales-Revised (ECERS-R; Harms et al., 2005), an observational tool designed to assess the process quality of group programs for children of preschool through kindergarten age. In addition, the study also used the Sensitivity subscale score from the Caregiver Interaction Scale (CIS; Arnett, 1989), an observational measure of caregiver sensitivity. Direct observation (e.g., teacher-child ratios) and teacher responses to standard demographic questionnaires (e.g., years of education,

completed degrees, number of years teaching) provided additional information regarding classroom structural variables. The Externalizing Behavior and Social Skills subscales on the Preschool and Kindergarten Behavior Scales- Second Edition (PKBS-2; Merrell, 2002), a 76-item teacher-completed rating scale designed to measure both problem behaviors and social skills of children ages 3-6, measured child outcomes in behavioral and social domains (externalizing behavior and social skills).

Hypotheses

Contained within the temperament literature, investigators have operationalized goodness of fit in several manners, of which the most promising is when goodness of fit is conceptualized in terms of environmental demand factors of characteristics of the setting. This view is in accord with the ideas of Keogh (2003; Keogh & Speece, 1996), who proposed that several aspects of classrooms are important in the fit between students and classroom environments: the content and nature of the curriculum; the organization of space, time, and resources; and the nature of interactions between students, peers, and teachers. The few studies that have considered classroom environment characteristics have measured setting demand factors such as program structure, emotional climate, physical arrangements, stimulation, and discipline methods (De Schipper et al., 2004).

These classroom environment characteristics originated from the early childhood classroom quality literature, which has identified several salient process variables, and like the goodness of fit literature, is interested in characteristics of the environment which relate to positive child outcomes in developmental domains (e.g., Hestenes et al., 1993; Howes et al., 1994; Kontos & Wilcox-Herzog, 1997; Lambert et al., 2002; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001; Pianta, 1999; Pianta et al.,

2002). While the classroom quality literature has integrated child temperament in its investigation, investigators have only assumed it to be one of several potential moderating factors, which has limited the scope of its analysis (e.g. Hagekull & Bohlin, 1995; Hestenes et al., 1993; Lamb et al., 1998). Thus, additional research was necessary to further operationalize the goodness of fit model and determine if child temperament and classroom environment characteristics of early childhood environments, as well as their interaction predict child outcomes in behavioral and social domains. Therefore, I hypothesized:

H01: It is expected that there will be a significant negative relationship between child temperament (TACTIC Total Temperament score) and child social skills (PKBS-2 Social Skills subscale score).

H02: It is expected that there will be a significant positive relationship between child temperament and child externalizing behavior (PKBS-2 Externalizing Behavior Problems subscale score).

H03: It is expected that there will be a significant negative relationship between child social skills and child externalizing behavior.

H04: It is expected that child temperament will be a significant positive predictor of child externalizing behavior.

H05: It is expected that child temperament will be a significant negative predictor of child social skills.

H06: It is expected that classroom process quality (ECERS-R Program Structure and CIS Sensitivity scores) will be a significant negative predictor of child externalizing behavior.

H07: It is expected that classroom process quality will be a significant positive predictor of child social skills.

H08: It is expected that the interaction between child temperament and classroom environment characteristics (ECERS-R Program Structure and CIS Sensitivity scores) will be significant negative predictors of child externalizing behavior.

H09: It is expected that the interaction between child temperament and classroom environment characteristics will be significant positive predictors of child social skills.

CHAPTER III

Method

This chapter presents the methodology that this study used to address the research questions and hypotheses concerning the operationalization of the goodness of fit model in school settings and the manner in which children's temperaments evidence expression in prekindergarten classrooms that have different classroom environment characteristics. As such, the chapter describes the study's participants as well as the instruments and procedures that the study used to assess participants' temperament and behavior and the characteristics and quality of their prekindergarten classrooms. In addition, the investigator presents the study's design and methods for data analysis.

Participants

The investigator solicited participation from school administrators in several diverse prekindergarten settings. The study used prekindergarten settings because they represent the major out-of-home socialization experience for many young children, and as such, the nature of the opportunities for social development within this setting can be very important for children's development (Lambert, Abbott-Shim, & McCarty, 2002). In addition, there has been extensive study of the relationship between classroom quality and children's behavioral and social outcomes, indicating that children in higher quality prekindergarten settings demonstrate more social competence and fewer externalizing behavior problems than those in lower quality settings (e.g., Bryant et al., 1994; Hestenes et al., 1993; Howes et al., 1994; Kontos & Wilcox-Herzog, 1997; Lambert et al., 2002; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001; Pianta, 1999; Pianta et al., 2002; Vandell, 2004).

It was also important that this study included settings that were as diverse as possible, as it sought to determine the manner in which children's temperaments evidenced expression in prekindergarten classrooms that have different classroom environment characteristics. As hierarchical linear modeling (HLM) was used and there is no power analysis for this procedure, the investigator attempted to solicit participation from as many classrooms as possible.

The investigator contacted administrators at ten potential school sites by email, phone, and verbally to inform them about the nature and content of the study (see Appendix B). Of these ten sites that were contacted, three (30%) agreed to participate in the study. A total of 11 teachers were contacted (one in each participating classroom) and all 11 teachers gave consent to participate (100%). As a result, this study included 11 prekindergarten classrooms in two Head Start settings and one private school setting. The following section will describe each of the three school sites that participated in this study; as well as the study's participants.

Two of the 11 classrooms that participated in this study were in a Head Start setting located on the Lower East Side of Manhattan. This head start offers a free, bilingual, preschool education to 127 students in two sites and seven classrooms. The program offers breakfast, lunch, and snack to all students as well as integrated special education and related services to 10% of their students. The school's students are equally distributed by gender, where 90% are Hispanic, 5% are African American, 2% are Asian, and 3% are Other. The educational program implements Creative Curriculum, a project-based early childhood curriculum designed to foster the development through teacher-led small and large group activities. One of the participating classrooms in this site had a

total of 19 students, of which 15 parents gave consent for their child to participate in this study (79%). The second classroom that participated had a total of 18 students, of which 14 parents gave their consent (78%).

Seven of the classrooms that participated in this study were from another Head Start setting located in Brooklyn, New York. It offers free and bilingual preschool education, integrated special education and related services as well as social serves to 130 children and their families in three sites and eight classrooms. The program also offers breakfast, lunch, and snack to all students. The educational program is based on the Reggio Emilia approach, a child-centered educational philosophy. Each classroom follows the inclusion model, where 45% of the students are classified as having a disability. The school's students are equally distributed by gender, and 68% are Hispanic, 29% are African American, and 3% are Caucasian. The seven classrooms that participated in this study in this site had the following total number of students/students for whom parental consent was received: classroom 1: 13/20 (65%); classroom 2: 18/20 (90%); classroom 3: 11/18 (61%); classroom 4: 14/20 (70%); classroom 5: 9/17 (53%); classroom 6: 11/12 (92%); classroom 7: 8/12 (67%).

Two of the classrooms that participated in this study were from a graduate school-based child development and learning center located in midtown Manhattan. The center offers preschool education to a total 27 students in two classrooms. It follows a play-based curriculum and emphasizes the process of interaction with materials and integration of knowledge versus acquisition of specific academic skills. The center also serves as a lab school for faculty and students doing research pertinent to preschool children. The school's students are equally distributed by gender, where 75% are White,

11% are Other, 7% are Asian, and 7 % are Black. While the preschool does not provide special education or related services, 4% of their student population is classified as having a disability. As data at this site was collected during the summer, there were fewer students in each classroom than during a typical academic year. Classroom 1 had a total of 10 students, of which nine parents provided their consent for their child to participate (90%). Classroom 2 in this site had a total of 12 students, of which eight parents provided consent for their child to participate (67%).

Table 1 provides descriptive statistics for the demographic information collected for teachers and students who participated in this study. The majority of the students who participated in this study was Hispanic, spoke English as a primary language, received free lunch at school, and did not have a documented disability. Students' ages ranged from 35 to 66.3 months or 2.9 to 5.5 years (M = 52.5, *SD* = 7.6). The majority of the teachers who participated in this study were female, black or Hispanic, and had completed a master's degree. The average number of students in the 11 classrooms that participated in this study was 16.2 (see Appendix C for student demographic information in each participating classroom).

Table 1

Student and Teacher Demographics

Student Demographics	Variable	*N*	%
Gender	Male	62	47.69
	Female	68	52.31
Ethnicity	Black	34	26.15
	Asian	2	1.54
	Hispanic	71	54.62
	White	13	10.00
	Other	10	7.69
Disability Status	No	108	83.08
	Yes	22	16.92
Free Lunch	No	18	13.80
	Yes	112	86.20
English Primary Language	No	49	37.69
	Yes	81	62.31

Age in Months	Range	M	SD
	35 - 66	52.49	7.63

Teacher Demographics	Variable	*N*	%
Gender	Male	1	9.00
	Female	10	82.00
Teacher Ethnicity	Black	4	36.36
	Hispanic	4	36.36
	White	3	27.27
Highest Degree	Associate's Degree	1	9.09
	Bachelor's Degree	3	27.27
	Master's Degree	6	54.55
	Post Master's Degree	1	9.09

Instruments

Student Level Variables

Teacher and Caregiver Temperament Inventory for Children (TACTIC; Billman & McDevitt, 1998). The TACTIC is a comprehensive inventory of behavior in 2.5 to 6 year-old children measuring temperament and attention with screening for problem areas of conduct and emotions. This rating scale was selected for use in this study because of the paucity of reliable and valid instruments reflective of the NYLS model that assess child temperament in early childhood settings where teachers serve as respondents.

The TACTIC includes 95 items divided into 13 scales. Nine of the scales reflect NYLS temperament dimensions and include seven items each (i.e., activity level, rhythmicity, adaptability, approach-withdrawal, intensity, mood, persistence, distractibility, and sensory threshold), and four of the scales reflect problem areas in attention, conduct, and emotion (i.e., ADHD Inattentive (nine items), ADHD Hyper-Impulsive (nine items), Conduct Problems (seven items), and Emotional Issues (seven items)). Examples of items assessing temperamental dimensions include: Is ok with changes in schedule and follows directions without any problem; Loses his/her temper when frustrated (teased, displeased); Ignores strong smells, such as cooking odors, paint, glue, etc.; Shows strong emotion. Both teachers and other child caregivers may complete the TACTIC with ratings on a 6-point scale in which the anchor points are the following: 1 = almost never; 2 = rarely; 3 = variable, usually does not; 4 = variable, usually does; 5 = frequently; 6 = almost always. A Total Temperament score can be derived by summing 65 items among the nine scales that reflect temperament dimensions. This total score is

indicative of temperamental difficulty. As it is an average of the scale's 65 items which are scored on a 6-point scale, it can range from 1 to 6 (M = 2.47, *SD* = .99).

The TACTIC also includes nine items that ask the rater to provide her general impressions of the child's temperament. Each item reflects a temperament dimension and is rated on a 6-point scale. For example, the Activity Level item defines the dimension "Activity Level- the amount of physical motion of movement shown during the day." The 6-point scale is the following: 1 = very inactive; 2 = inactive; 3 = somewhat inactive; 4 = somewhat active; 5 = active; 6 = very active.

In this dissertation, the 11 head teachers completed a TACTIC for each participating child in their class. This study used the Total Temperament (raw) score, a measure of temperamental difficulty, derived by summing the nine temperament scale's 65 items. I selected the Total Temperament score based on its clinical utility (S. McDevitt, personal communication, October 9, 2007) and on the plethora of research that indicates that the difficult temperament profile predisposes a child to behavior and adjustment problems (e.g., Caspi & Silva, 1995; Guerin et al., 1997; Maziade et al., 1990; Teglasi & MacMahon, 1996; Thomas et al., 1968).

Temperament items for TACTIC assess child temperament according to the NYLS model (Thomas et al., 1968), and Billman and McDevitt (1998) selected temperament items for the TACTIC from the Preschool Temperament Inventory (PTI; Billman, 1981) that was developed in response to problems in the reliability and validity of teacher-rated early childhood questionnaires. The PTI uses teacher ratings on eight of the NYLS temperament categories, including activity level, approach/withdrawal, adaptability, intensity, mood, persistence, distractibility, and threshold. To establish

reliability, teachers rated 54 preschool children on 72 items of a pilot instrument. A chi-square test on each item differentiated the 15 high scorers and the 15 low scorers in each temperament category. Items not showing a difference at the $p < .05$ level of significance were eliminated from the revised version, which resulted in a questionnaire with a total of 54 items. Questionnaires were then rescored only including reliable items, and the PTI was administered to an additional sample of teachers of 73 preschool children. To validate the instrument, a subsample of 31 children was rated on the Behavioral Style Questionnaire (BSQ; McDevitt & Carey, 1978), a well established parent-rated temperament rating scale. Here, moderate correlation coefficients were found for the majority of eight scales with its subscales. Overall, Billman (1981) concluded that the PTI reliably measures temperament traits of children 3 to 6 years of age.

Billman and McDevitt (1998) developed additional items for the TACTIC that measure attentional issues and conduct and emotional problems through a process of inductive-empiric analysis. They theoretically derived items measuring attentional issues from symptoms described in the DSM-IV. The authors rationally derived items measuring conduct and emotional problems. They then subjected these items addressing clinical problems to a series of repeated item analyses, where the only items that they retained were those with moderate to high intercorrelations (S. McDevitt, personal communication, March 11, 2009).

Billman and McDevitt (1998) computed coefficient alphas to determine internal consistency on a sample of 99 children, ages 2.5 to 6 years old. The coefficients ranged from .70 to .93 for the temperament category scores, and the coefficient for the Total scale score was .85. In the present study, the coefficient alpha obtained for the Total scale

score study was .96. The temperament items included in the TACTIC are based on a well-known temperament theory (Thomas et al., 1968) and as such, users can regard the scale as having high construct validity. In addition, Billman and McDevitt (1998) used quantitative methods to study and control item validity. The resultant questionnaire therefore has high content validity and the behaviors described compare favorably with those identified in the literature as specific to children ages 2.5 to 6 years old (Billman & McDevitt, 1998).

This present study as well as a pilot study for this dissertation (Blackwell, 2008) established convergent and discriminant validity for the TACTIC, as it found significant positive relationships between TACTIC measures of temperament (and externalizing behavior (as measured by the Preschool and Kindergarten Behavior Scales- Second Edition (PKBS-2; Merrell, 2002)), and significant negative relationships between measures of temperament and social skills (as measured by the PKBS-2). The pilot study also found significant positive relationships between TACTIC scales of Hyper-impulsivity, Inattention, Conduct Issues, and Emotional Issues with the Externalizing Behavior Problems scale on the PKBS-2, and significant negative relationships between these scales on the TACTIC and the Social Skills scale on the PKBS-2 (Blackwell, 2008).

The Preschool and Kindergarten Behavior Scales- Second Edition (PKBS-2; Merrell, 2002). In this study, head teachers completed a PKBS-2 for each student in the classroom. The PKBS-2 is a 76 item norm-referenced, standardized rating scale designed for use in evaluating social skills and problem behaviors of preschool- and kindergarten-age children, ages 3 through 6 years. The PKBS-2 was used in this study because it yields scores indicative of children's social skills and externalizing behavior problems, the two

child outcome variables in this study. In addition, the instrument measures social skills and problem behaviors specifically manifested in preschool settings, a central focus of this study.

The PKBS-2 is a norm-referenced and standardized instrument developed for use in a number of settings and by a variety of informants. Either home-based (e.g., parents) or school-based (e.g., teachers, school administrators) raters can complete the PKBS-2. The PKBS-2 contains two composite scales: Social Skills (34 items) and Problem Behavior (42 items). Both of these scales include empirically derived subscales that are posited to be useful in identifying specific clusters of social skills and problem behaviors (Merrell, 2002). Head teachers completed a PKBS-2 for each participating student in their classroom.

Raters respond to both the Problem Behavior and Social Skills items on a 4-point scale in which the anchor points are the following: 0 = never; 1 = rarely; 2 = sometimes; 3 = often. In addition, the PKBS-2 yields Standard Scores (SS), with a mean of 100 and a standard deviation of 15. The scale requires approximately eight to 12 minutes to be completed by a caregiver who has been familiar with the child for at least 3 months prior to the rating period.

The Social Skills scale includes 34 items that describe adaptive behaviors that are likely to lead to positive personal and social outcomes that are characteristics of well-adjusted children, ages 3 through 6 years. The manual notes that the "PKBS-2 is designed to focus on social skills (specific behaviors which may lead to adequate social competence) and peer relations (the social outcome of social skills) rather than the entire dimension of social competence" (Merrell, 2002, p. 8). The Social Skills scale includes

several empirically derived subscales: Social Cooperation (12 items; reflecting adult-related social adjustment), Social Interaction (11 items; reflecting peer-related social behaviors, like making friends) and Social Independence (11 items; reflecting peer-related social behaviors that allow one to achieve independence within the peer group). Overall, the scale yields three subscale scores, which can range from a SS of 21 to 120, as well as one total Social Skills scale score, which can range from a SS of 42 to 152. This study used the Social Skills scale score as one of two child outcome variables. Examples of items on this scale are: Works independently; Is invited by other children to play; Shows self-control; Follows Instructions from adults. High scores on this scale indicate well developed social skills (Merrell, 2002). The investigator selected this scale because it focuses on typical and general social competencies displayed by children, 3 through 6 years of age.

The 42-item Problem Behavior scale includes items that describe various problem behaviors and difficulties in adjustment typically seen among children in this age group. The items are divided into two empirically derived subscales, Internalizing Problems (15 items) and Externalizing (Behavior) Problems (28 items). The internalizing dimension includes symptoms of depression, social withdrawal, anxious and inhibited behaviors, and the development of somatic problems. The externalizing dimension consists of aggressive, defiant, disruptive, oppositional, overactive, and acting-out behaviors. Because this study focused on behavior problems, the Externalizing Behavior Problems subscale score was used as the other child outcome variable (in addition to Social Skills). High scores on this scale are indicative of self-centeredness, explosiveness, attention problems, over-activity, and antisocial/aggressive tendencies (Merrell, 2002). Examples

of items on this subscale are: Disobeys rules; Acts impulsively without thinking; Destroys things that belong to others; Bothers and annoys other children. The Externalizing and Internalizing subscale scores can range from a Standard Score (SS) of 76 to 168, and Problem Behavior scale score from an SS of 42 to 152.

The PKBS-2 was used in this study because it yields scores indicative of children's social skills and externalizing behavior problems, the two child outcome variables in this study. In addition, the instrument measures social skills and problem behaviors specifically manifested in preschool settings, a central focus of this study. Raters responded to both the Problem Behavior and Social Skills items on a 4-point scale in which the anchor points are the following: 0 = never; 1 = rarely; 2 = sometimes; 3 = often. The PKBS-2 yields SSs, with a mean of 100 and a standard deviation of 15. The scale requires approximately eight to 12 minutes to be completed by a caregiver who has been familiar with the child for at least three months prior to the rating period.

Merrell (2002) used three methods to determining reliability of the PKBS-2. Coefficient alpha and split-half reliability determined internal consistency reliability. The coefficients ranged from .81 to .97 for the subscale scores, and .94 to .97 for the overall scores. Coefficient Alpha coefficients found by the test's author for the Externalizing Problems subscale and Social Skills scale were $\alpha = .97$ and $\alpha = .96$, and split-half coefficients were $r = .95$ and $r = .94$, respectively (Merrell, 2002). Coefficient alphas obtained for these subscales in the present study were $r = .95$ and $r = .97$, respectively.

Merrell (2002) established test-retest reliability using ratings by preschool teachers of 82 children at three different time periods: a baseline rating, a retest three weeks after baseline, and a retest 3 months after the baseline measure. At three weeks the

resulting coefficients of stability, ranging from .58 to .87, were in the moderate to high range. At three months, most coefficients were also in the moderate to high range, ranging from .66 to .78, with the exception of the Anxiety/Somative Problems subscale (within the Internalizing dimension), which had a coefficient of .36. Coefficients for the Externalizing Problems subscale and Social Skills scale were .87 and .58 at 3 weeks, and .78 and .69 at three months, respectively (Merrell, 2002).

To establish inter-rater reliability, teachers and classroom aides rated 82 children during the same one-week period. Coefficients ranged from .36 to .63, all significant at an alpha level of $p < .001$. Coefficients for the Externalizing Behavior Problems subscale and Social Skills scale were .63 and .48, respectively. The author noted that the modest level of agreement may have been the result of different amounts of experience among the raters (Merrell, 2002).

Merrell (2002) also provided four descriptions of validity evidence: content, construct, convergent/discriminate, and criterion. To ensure content validity, the author explained that the test was initially constructed using a rational-theoretical approach, where using existing literature, he created a pool of items that a panel of 16 experts then modified and reviewed. Using factor analysis, Merrell dropped several items that did not align with the theoretical structure of the scale. Item analysis then further eliminated several items that did not correlate highly with the total score.

Correlations among the various subscales were moderate to high, suggesting that they are somewhat related, but that they measure specific types of the central construct of interest (i.e., social skills or problem behaviors). In addition, Merrell (2002) used a variety of factor analytic methods, where item loading also supported the scales'

construct validity. Specifically, three factors emerged in the Social Skills scale, labeled as: social cooperation, social interactions, and social independence. The Problem Behavior scale yielded a more complex factor structure with five major factors: self-centered/explosive, antisocial aggressive, social withdrawal, and anxiety/somatic problems. The author also noted that the data in the Problem Behavior scale also fit a two-factor structure: internalizing problems and externalizing problems (Merrell, 2002).

Merrell (2002) investigated the PKBS-2's convergent and discriminant validity by correlating its scores with those from other instruments. He found strong associations of the Social Skills scores on the PKBS-2 with the established measures of social skills, such as the social skills scores of the Social Skills Rating System (SSRS; Gresham & Elliot, 1990), Walker-McConnell Scale of Social Competence and School Adjustment (SSCSA; Walker & McConnell, 1995) and the School Social Behavior Scales (SSBS; Merrell, 1993). The PKBS-2 Problem Behavior scores also were strongly associated, particularly within the domains of externalizing and internalizing problems, with established measures of problem behavior, such as the Conners' Teacher Rating Scale-39 (CTRS-30; Conners, 1990), Teacher's Report Form (TRF; Achenbach, 1991), Adjustment Scale for Children and Adolescents (ASCA, McDermott, Marston, & Stott, 1993), and Matson Evaluation of Social Skills with Youngsters (MESSY, Matson, Rotari, & Helsel, 1985). In addition, the Social Skills score on the PKBS-2 was inversely associated with the measures of problem behavior, and PKBS-2 Problem Behavior score was inversely associated with the measures of social skills. PKBS-2 scores also had very weak relationship with scores from other measures of unrelated constructs (Merrell, 2002).

Finally, when investigating criterion validity, Merrell (2002) found significant group differences on the PKBS-2 between males and females, between typically developing children and children considered to be at risk for social-behavior problems (i.e., Child Find, Internalizing, Externalizing, and ADHD status), and between younger and older children.

Classroom Level Variables

Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 2005). The ECERS-R was used in this study to measure classroom environment characteristics. The ECERS-R is a criterion-referenced observational tool designed to assess the process quality, which is the quality of the social, emotional, physical, and instructional elements of interactions with young children in educational settings (Pianta et al., 2005) of group programs for children, ages 2.5 through 5 years of age. The ECERS-R assumes an expansive characterization of early child environments, including "those spatial, programmatic and interpersonal features that directly affect the children and adults in early childhood settings" (Harms et al., 2005, p. 1). This measure was selected for use in this study because it is aligns with Developmentally Appropriate Practices (Bredekamp & Copple, 1997) that are described in NAEYC's definition of programmatic quality in early childhood environments (Harms et al., 2005).

The ECERS-R has seven subscales, consisting of a total of 43-items and 470 quality indicators. Each item is rated on a 7-point scale, with the odd number descriptors of *inadequate* (1), *minimal* (3), *good*, (5), and *excellent* (7), by an observer who spends of a minimum of 3hours in the classroom being rated (Harms et al., 2005). Scores for each subscale, ranging from 1 to 7, are derived by averaging the subscale score by the total

number of points possible for a given domain. A total score for the entire classroom can also be derived from the summation of all ratings, which is then averaged with the total numbers of points possible, leaving this overall score to fall between 1 and 7, with higher scores for the total and each subscale representing better classroom quality.

Based on the recommendations of Keogh (2003; Keogh & Speece, 1996), trained and independent observers completed three ECERS-R subscales: the Activities, Interaction, and Program Structure subscales. The Activities subscale is comprised of 10 items that focus on the environmental areas of (a) fine motor skills; (b) art; (c) music and movement; (d) blocks; (e) sand and water; (f) dramatic play; (g) nature and science; (h) math and number; (i) use of television, video and/or computers; and (j) promoting acceptance of diversity (Harms et al., 2005). The Interaction subscale includes five items that focus on (a) supervision of gross motor activities, (b) general supervision of children, (c) discipline, (d) staff-child interactions, and (e) interactions among children. The Program Structure Scale is composed of four items that focus on (a) schedule, (b) free play, (c) group time, and (e) provisions for children with disabilities (Harms et al., 2005). Items in these subscales are rated on a score sheet that also asks raters to provide information on the number of staff and children present (teacher:child ratio). This information (i.e., teacher:child ratio) was used as one of the classroom structural variables.

The ECERS-R is a revision of the ECERS (Harms & Clifford, 1980). Harms et al. (2005) used three main sources of information in the revision: a content analysis of the relationship of the ECERS to other global program quality assessment instruments and documents examining early childhood programmatic issues, data from studies using the

ECERS in early childhood settings, and feedback from ECERS users. Changes to the revised scale included deletion of infant-toddler alternative items, creation of separate descriptions of quality indicators, expansion of Notes for Clarification, separation of several items to deepen content, addition of several items, addition of examples to make items more inclusive and culturally sensitive, and creation of a more consistent scoring system.

There are three types of reliability data available for the ECERS-R: inter-rater, internal consistency, and test-retest. Harms et al. (2005) conducted an extensive set of field tests of the ECERS-R in 45 classrooms. In response to unsatisfactory reliability of the measure, the authors made several revisions to the instrument. After these revisions, inter-rater reliability was again assessed and it was found that the ECERS-R was a reliable instrument at both the indicator and item level. The percentage of agreement across the full 470 indicators in the scale was 86.1% with no item having an indicator agreement level below 70%. At the item level, the proportion of agreement was 48% for exact agreement and 71% for agreement within one point (Harms et al., 2005). A pilot study for this dissertation (Blackwell, 2008) found item level exact percent agreement of 96%. In the present study, at the (sampled) item level, the proportion of agreement was 70% for exact agreement and 88% for agreement within one point.

Harms et al. (2005) also examined internal consistency at the subscale and total score levels. Subscale intraclass correlations ranged from .71 to .88, with the Activities, Interaction, and Program Structure having intra-class correlations of .88, .86, and .77, respectively. In addition, internal consistency among scale items as measured by Cronbach's alpha on the authors' sample was excellent at .92. (Harms et al., 2005).The

present study found Cronbach's alphas of .54, .81, and .79 for the Activities, Interaction, and Program Structure Scales, respectively. The authors did not provide test-retest reliability. The pilot study for this dissertation (2008) found test-retest reliability coefficients of .96, .92, and 1.0 for the Activities, Interaction, and Program Structure subscales, respectively. The present study did not assess test-retest reliability.

The authors noted that because the ECERS-R maintains the same conceptual framework, basic scoring approach, and administration of the original version, which has a long history of demonstrating that the classroom quality that it measures has good predictive validity (i.e., Peisner-Feinberg & Burchinal, 1997; Whitebook, Howes, & Phillips, 1990) (discussed in Chapter 2, pp. 58-59). The revised edition "would be expected to maintain that form of validity" (Harms et al., 2005, p. 1). Thus, Harms et al. (2005) did not provide information specific for the validity of the measure. Bredekamp and Copple (1997) indicated that the ECERS-R has face and content validity because 37 of the total 43 ECERS-R items are directly linked to statements of developmentally appropriate practice (embodied in NAEYC's definition of programmatic quality (Harms et al., 2005).

Several studies (e.g., Helburn, 1995; Howes et al., 1992; Phillipsen, Burchinal, Howes, & Cryer, 1997) have examined the factor structure of the ECERS with item loadings generally supporting the scale's construct validity. For example, Cassidy and colleagues (1995) explored the factor structure of the ECERS-R using a sample of 1,313 classrooms. Exploratory and confirmatory factor analyses resulted in a two-factor solution that accounted for 69% of total item variance. The first factor, named Activities/Materials, consisted of nine ECERS-R items, and the second factor,

Language/Interactions, consisted of seven items. In addition, several authors (i.e., Essa & Burnham, 2001; Perlman, Zellman, & Le, 2004; Sylva et al., 2006) have suggested that the ECERS-R statistically yields a single construct related to the general quality of an early childhood developmental environment.

The Arnett Caregiver Interaction Scale (CIS; Arnett, 1989). The CIS, a 26-item criterion referenced observational tool that assesses the content of teacher's interactions with children (based on the recommendations of Keogh (2003; Keogh & Speece, 1996) was also used to measure classroom environment characteristics. The CIS was formulated based on a study conducted by Arnett (1989), who was interested in the level of training caregivers in childcare centers and its effect on their interactions with children in those childcare centers. The investigator selected this scale for use because it is different from the ECERS-R, which is a measure of broad process quality (i.e., classroom structure, materials and tone). The CIS is used to assess teacher's caregiving behaviors and overall sensitivity in the classroom.

CIS items are usually organized into the following four subscales: Sensitivity (10 items), Harshness (8 items), Detachment (4 items), and Permissiveness (4 items) (de Kruif et al., 2000). When combined, these four scales yield an overall caregiver quality score. Observers rate caregiver-child interactions on a 4-point Likert-type scale, with the following descriptors: (1) not at all true, (2) somewhat true, (3) quite a bit true, and (4) very much true. Here, scores can range from 1 to 4, with higher scores reflecting high quality caregiver behaviors. The Sensitivity subscale measures the degree of caregiver enthusiasm, warmth of the caregiver's interaction with children, and the developmental appropriateness of his/her communication with children. An example of an item on this

scale is: Speaks warmly to the children. The Detachment subscale measures the extent to which the caregiver is uninvolved with and uninterested in the children. An example of an item of this scale is: Seems distant or detached from the children. The Harshness subscale measures the extent to which the caregiver engages in hostile, threatening, and harshly critical behavior toward the children. For example: Places high value on obedience. Finally, the Permissive subscale contains items that refer to the extent to which teachers permit inappropriate behavior to occur. For example: Doesn't try to exercise much control over the children (Arnett, 1989). This study intended to use the Total Scale score and Sensitivity subscale score, along with the ECERS-R scales to assess aspects of the classroom environment (i.e., process quality). The investigator selected these scores because of their typical use in research studies that examined the relationship between classroom quality and child outcomes in behavioral and social domains (e.g., Burchinal et al., 2002; Ghazvini & Mullis, 2002; Howes, 1997).

There are three types of reliability information available for the CIS: internal consistency, inter-rater reliability, and test-retest reliability. Resnick and Zill (1997) obtained a Cronbach alpha of .98, and Maxwell and colleagues (2001) found an alpha level of .93 for the Total Scale score. In addition, Layzer and colleagues (1993) obtained a Cronbach alpha of .91 for the Sensitivity subscale. The present study found Cronbach alphas of .81 for the Total Scale and .81 for the Sensitivity subscale. Peisner-Feinberg and colleagues (1997, 2001) reported inter-rater reliability coefficients ranging from .89 to .98 for each of the four subscales, with median subscale scores from .92 to .95. Maxwell and colleagues (2001) reported that inter-rater agreement within 1 point for the CIS was 100% across classrooms; exact agreement ranged from 27% to 85%, with a

mean of 60%. The pilot study for this dissertation (Blackwell, 2008) found exact agreement of 96%. In the present study, exact inter-rater agreement was 76% and agreement within 1 point was 96%. While this study did not assess test-retest reliability, a pilot found test-retest coefficients of .96 for the Sensitivity subscale and .98 for the Total Scale score (Blackwell, 2008).

There are two types of validity evidence available for the CIS: construct validity and concurrent validity. Several studies (e.g., Burchinal et al., 2002; Ghazvini & Mullis, 2002; Howes, 1997) have replicated Arnett's (1989) findings that caregivers with more training are more likely to engage in sensitive and positive interactions with children in their care. In addition, other studies (e.g., Arnett, 1989; Ghazvini & Mullis, 2002, de Kruif et al., 2000) have generally replicated the factor structure of the CIS. Layzer et al. (1993) reported correlation coefficients ranging between.43 and.67 between the CIS and other instruments measuring more global aspects of early childhood classroom environments. The authors did not expect the coefficients to be large because the Arnett scale focuses more narrowly on aspects of teacher behavior (i.e., teacher sensitivity) that other observational instruments do not measure directly. However, Phillipsen et al. (1995) reported a correlation of .76 between the Arnett and the ECERS (Harms & Clifford, 1980).

Demographic Questionnaires

The investigator developed the demographic questionnaires used in this study. There were two questionnaire forms, a parent form (see Appendix D) and a teacher form (see Appendix E). The teacher form allowed the investigator to gain information on teacher demographic variables, such as ethnicity, as well as various classroom structural

variables that may relate to classroom quality. For example, the teacher provided information on the number of years of completed education, received degrees, and the number of years she has worked as a teacher. The parent form included such information as relationship to the child, language(s) spoken in the home, ethnicity, disability status if any, and the number of years the child has been in out of home care.

Procedure

When the school administrators at each of the three sites agreed to participate in this study (see Appendix B), they communicated with prekindergarten classroom teachers to ensure that they were also willing to partake in the study. When the investigator received this approval from the administrators, she gave teachers an introductory letter describing the study (see Appendix F), along with a consent form (see Appendix G), which teachers completed and returned either by giving them to the investigator, a research assistant, or via mail in a provided stamped self-addressed envelope. Introductory letters describing the nature of the study (see Appendix H) and consent forms (see Appendix I) were also distributed to the parents of students in the participating classrooms. Parents received the introductory letter and consent form by hand along with a self-addressed, stamped envelope in which to return the signed form providing permission for participation. The parent or guardian was also given the option to return the signed consent forms directly to the school.

The parent consent form indicated that I was the principal investigator for the study, and that the child's name would not appear on any of the information gathered (as children would be assigned an identification number). It also highlighted that all results would be kept strictly confidential and that all data would be kept in a locked file cabinet.

In addition, the letter informed parents that they could stop their child's participation at any time without any penalties. Teachers' consent letters also informed them that their names would not appear on any of the information gathered and that all classrooms would be referred to by a randomized number created by the principal investigator. In addition, teachers' letters also informed them that they could withdraw from the study at any time without penalties of any kind. If parents did not consent to their children's participation in the study, the teacher did not complete any rating scales for those children.

When teachers and parents returned the signed consent forms, participating parents and teachers completed their respective demographic questionnaires. Appendix J provides a second letter that the investigator provided to parents that describes the purpose of the demographic questionnaire.

The head teachers in each classroom also received in the packet with the demographic questionnaire the rating scales (i.e., PKBS-2 and TACTIC) to complete for each participating child as well as a letter provided instructions on how to complete the rating scales (see Appendix K) along with a copy of the consent form that the teacher and investigator had signed. In addition, in their packet with the demographic questionnaire, the parents/guardians received a copy of the consent form that they and the investigator had signed.

Two trained and independent observers rated each classroom on the several scales of the ECERS-R (the Activities, Interaction, and Program Structure scales) and the entire CIS to obtain measures of classroom environment characteristics and classroom process quality. These observers were the investigator's research assistants, and were second-year

graduate students in a school psychology certificate program. These research assistants completed the ECERS-R and CIS simultaneously once. The investigator chose simultaneous versus repeated observations because studies that examined how classroom process quality relates to positive child outcomes in behavioral and social domains typically only considered inter-rater (versus test-retest) reliability (i.e., Bryant et al., 1994; Hestenes et al., 1993; Howes et al., 1994; Kontos & Wilcox-Herzog, 1997; Lambert et al., 2002; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001). In addition, a pilot study for this dissertation that used both repeated and simultaneous ratings from the ECERS-R and CIS found test-retest coefficients of .96, .92, and 1.0 for the Activities, Interaction, and Program Structure scales on the ECERS-R, respectively, and coefficients of .98 and .96 for the CIS Total Scale and Sensitivity subscale, respectively (Blackwell, 2008). These observational ratings of the classroom environment took about 1.5 hours in entirety per classroom. The investigator trained these observers to use both instruments. Appendix L outlines these training procedures. As the description specifies, during training procedures, the proportion of inter-rater agreement was calculated that indicated sufficient levels of agreement (see p. 91 for the ECERS-R and p. 95 for the CIS).

Data Analysis

This study used an observational design. The investigator used several procedures to analyze the data in order to answer this study's research questions. In the first stages of analyses, all of the variables used in this study, including teacher and child demographic information, were analyzed via descriptive statistics. In addition to basic descriptive

statistics, the investigator conducted correlational analyses among study variables to assess the direction and strength of variable relationships.

In the second stage of analysis, the investigator conducted hierarchical linear modeling (HLM). The basic concept of HLM is similar to regular regression. At the individual or student level (Level 1), an outcome variable is predicted from a set of predictors. At the classroom level (Level 2), Level 1 slopes and intercepts become dependent variables being predicted from Level 2 variables. Following this, one can model the effects of Level 1 variables on the outcome and the effects of Level 2 on the outcome. Table 2 presents the student- and classroom-level variables used in the analyses in this study. Student-level variables included demographic characteristics and socioeconomic-related information. Classroom-level variables included teacher demographic characteristics, classroom structural characteristics, and classroom process characteristics. In addition, Table 3 highlights corresponding Level 1 and Level 2 equations for a two level HLM model.

Table 2
HLM Variables Used in this Study by Level

Student Level Variables	Classroom Level Variables
Gender	Teacher Yrs of Education
Ethnicity	Teacher Yrs of Teaching
Disability Status	Teacher Degree Status
Eligibility status for free/reduced lunch	CIS Sensitivity
	ECERS-R Program Structure
TACTIC	

Table 3

Equations for a Two Level HLM Model

Symbol	Meaning
Level 1 Equation	$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$ where $r_{ij} \sim N(0, \sigma^2)$
Y_{ij}	Value of the DV at Level 1 for student i in classroom j
X_{ij}	Level 1 Predictor
β_{0j}	Intercept for DV in classroom j
β_{1j}	Slope for the relationship in classroom j between the DV and Level 1 predictor; change in DV with 1 unit increase in Level 1 predictor in group j
r_{ij}	Random error of prediction for student i in group j
σ^2	Variance of Level 1 random errors
Level 2 Equations	$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$
	$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}$
W_j	Level- 2 Predictor
γ_{00}	Overall intercept; Grand mean of the DV scores across all classrooms when all predictors are equal to zero
γ_{01}	Overall regression coefficient for the relationship between a Level 2 predictor and the DV
u_{0j}	Random error for the deviation of the intercept of a classroom from the overall intercept; the unique effect of classroom j on the intercept
γ_{10}	Overall regression coefficient for the relationship between Level 1 predictor and the DV when the Level 2 predictor(s) are equal to zero
γ_{11}	Change in slope for the relationship between the Level 1 predictor and DV with a one unit increase in the Level 2 predictor
u_{1j}	Random error for the slope for relationship between the Level 1 predictor and DV; unique effect of group j on slope for relationship between Level 1 predictor and the DV, controlling for the Level 2 predictor(s)
τ_{00}	Variance of u_{0j}
τ_{01}, τ_{10}	Covariance between u_{0j} and u_{1j}
Combined Equation	$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \gamma_{11} W_j X_{ij} + u_{0j} + u_{1j} X_{ij} + r_{ij}$

Raudenbush and Bryk (2002) recommended a build-up strategy for HLM analyses. This strategy was followed in this study. The first analysis was conducted through an unconditional model, or a one-way ANOVA model with random effects for each classroom. This model provided useful preliminary information about how much variation in the outcome lay within and between classrooms and about the reliability of each classroom's sample mean as an estimate of its true population mean. As such, this model was used as a baseline for comparison with subsequent, more complex models. These more complex models included TACTIC scores at Level 1 as a student covariate, classroom level predictors that included CIS Sensitivity and ECERS-R Program Structure total scale scores, and interaction effects between CIS Sensitivity and ECERS-R Program Structure scores. Table 4 highlights the sequential modeling analyses that were conducted in this study for both outcome variables (i.e., externalizing behavior and social skills), emphasizing both student level, as well as classroom level predictors entered into each model.

Table 4

Structure of Fitted Models for Student Outcomes

Model	Covariates included in Level 1	Covariates included in Level 2
Model a	Unconditional Model	Unconditional Model
Model b	Student Demographics	Teacher Demographics
Model c	Student Demographics TACTIC	Teacher Demographics
Model d	Student Demographics TACTIC	Teacher demographics CIS Sensitivity ECERS Program Structure
Model e	Student Demographics TACTIC	Teacher demographics CIS Sensitivity ECERS Program Structure CIS Sensitivity x ECERS Program Structure

To assess how much variation in student outcomes can be attributed to classroom level characteristics, the intraclass correlation was used to assess the baseline model. The intraclass correlation was computed by:

$$\rho = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}^2} \qquad (1)$$

where $\hat{\tau}_{00}$ is the total explainable variation at the Level 2 (classroom level), and $\hat{\sigma}^2$ is the total variance in outcomes within classrooms that can be explained by a Level 1, or student level model. The percentage of variance that can be attributed to the student level (Level 1) can be easily found according to $\rho - 1.$ Generally, if the amount of variation in outcome variables that can be attributed to Level 2 is beyond 10%, HLM is warranted.

In order to assess the amount of variation accounted for by the two level HLM models in this study, as suggested by Snijders and Bosker (1994), separate examinations of variance at both Level 1 and Level 2 were evaluated using pseudo-R^2 values. The pseudo-R^2 is a way to derive separate measures of the proportional reduction in mean squared prediction error for each level, or more specifically, the percent of variation accounted for at each level explained by the predictors included at that level. The process of calculating this value begins with the null model, or the unconditional model, which serves as a benchmark for determining the R^2 at each level of the hierarchy. By using variance estimates from the null model and variance estimates from the final model (with all predictors), the percent reduction in Level 2 variance is calculated by (Raudenbush and Bryk, 2002, p. 74):

$$R_2^2 = \frac{\hat{\tau}_{00}(null) - \hat{\tau}_{00}(final)}{\hat{\tau}_{00}(null)}, \tag{2}$$

and for the percent reduction in Level 1 variance:

$$R_1^2 = \frac{\hat{\sigma}^2(null) - \hat{\sigma}^2(final)}{\hat{\sigma}^2(null)}, \tag{3}$$

where $\hat{\tau}_{00}$ and $\hat{\sigma}^2$ were explained above. Although this statistic is not analogous to the R^2 used in standard multiple linear regression that represents the amount of variance accounted for by the model, it does give some indication of the proportion reduction in mean squared error at each level.

Another important issue in HLM is centering predictor variables, or using the deviation scores from a mean instead of the raw scores. By doing so, model intercepts will become more meaningful. The meaning of intercept terms, however, depends on

which mean is subtracted from the original scores. Kreft, Leeuw, and Aiken (1995) discussed different centering methods in hierarchical linear models and concluded that the decision about which centering method should be used depends on the theory and research questions asked. Grand mean centering is more often used in studies on large-scale achievement assessment (e.g., Braun, Jenkins, & Grigg, 2006), and often improves the interpretability of coefficients because zero takes on a meaning. In this study, Level 1 and Level 2 predictors were grand mean centered.

CHAPTER IV

Results

The primary aim of this study was to determine how the relationship between child temperament and behavior is expressed in different types of classroom environments in prekindergarten settings. Another goal of the study was to evaluate possible interactions of process variables indicative of classroom quality with child temperament to see if these interactions predicted child behavior and social skills. This chapter provides results for these hypotheses that operationalized the goodness of fit model as it relates to child outcomes in behavioral and social domains in prekindergarten settings.

Descriptive Statistics and Correlations of Study Measures

Table 5 presents study descriptive statistics for student, classroom, and structural variables, respectively. Results indicate that mean student Externalizing Behavior and Social Skills scores on the PKBS-2 (Merrell, 2002) as well as the TACTIC Total Scale difficult temperament scores (Billman & McDevitt, 1998) within this sample were within the average ranges of the normative samples for these instruments. Thus, the scores indicate that on average, participants demonstrated typical levels of externalizing behavior, social skills, and temperamental characteristics. This finding is notable, as about 17% of this study's sample was classified as a having a disability. However, it was communicated by school administrators that the majority of these students had speech/language impairments versus emotional disturbances that would be associated with extreme behavior problems and marked impairments in social skills.

While the authors of the CIS (Arnett, 1989) and ECERS-R (Harms, Clifford, & Cryer, 2005) did not provide expected ranges for their measures, the ranges found in this study appear to be slightly higher than those found in other studies (e.g., Burchinal et al., 2002; Gerber et al., 2007; Harms et al., 2005). Classroom structural variables indicated that teachers who participated in this study had on average completed a master's degree, 17.45 years of education, and had been teaching for 11.55 years. In addition, the mean student to teacher ratio was 4.92. These values obtained for classroom structural variables for the classrooms in this study are also higher than those typically found (e.g., Early et al., 2007; Pianta et al., 2005).

Table 5

Study Descriptive Statistics for Student, Classroom, and Structural Variables

Variable	*N*	*M*	*SD*	Min	Max
Student Level Variables					
Externalizing Behavior	130	92.59	16.29	21.00	146.00
Social Skills	130	103.39	14.84	61.00	122.00
TACTIC	130	2.68	0.64	1.54	4.08
Classroom Level Variables					
CIS Sensitivity	11	3.04	0.47	2.00	3.65
CIS Total	11	3.39	0.28	2.74	3.77
ECERS Activity	11	5.14	0.44	4.50	6.00
ECERS Interaction	11	6.22	1.09	3.70	7.00
ECERS Program Structure	11	5.64	1.55	3.50	7.00
Classroom Structural Variables					
Highest Degree	11	2.64	0.81	1.00	4.00
Years of Education	11	17.45	1.75	15.00	20.00
Years of Teaching	11	11.55	7.23	2.00	25.00
Ratio of students per teacher	11	4.92	2.79	2.25	10.00

Table 6 presents correlations among students' Externalizing Behavior, Social Skills, and TACTIC Total Scale scores. As expected, in line with the study's hypotheses, student externalizing behavior was significantly negatively related to social skills. In addition, the TACTIC Total Scale score, a measure of temperamental difficulty, was significantly positively related to externalizing behavior and significantly negatively related to social skills. These findings indicate that H01, H02, and H03 (i.e., there will be a significant negative relationship between temperament (TACTIC Total Temperament

score) and social skills (PKBS-2 Social Skills subscale score); there will be a significant positive relationship between temperament and externalizing behavior (PKBS-2 Externalizing Behavior Problems subscale score); there will be a significant negative relationship between social skills and externalizing behavior, respectively) were supported. This indicates that difficult temperament and externalizing behavior are positively associated with one another, and that difficult temperament and externalizing behavior are negatively associated with social skills. Appendix M presents correlations among students' Externalizing Behavior, Social Skills, and TACTIC category scores.

Table 6

Correlations Among Student Variables (N = 130)

	Externalizing Behavior	Social Skills	TACTIC
Externalizing Behavior	1		
Social Skills	-0.44**	1	
TACTIC	0.53**	-0.43**	1

**Significant at the $p < .001$ level.

Table 7 presents the correlations among classroom variables assessed in this study: CIS Sensitivity, CIS Total, ECERS-R Activities, ECERS-R Interaction, and ECERS-R Program Structure. Significant positive relationships were found between CIS Total, CIS Sensitivity, and ECERS-R Interaction scores. In addition, there was a significant positive relationship between ECERS-R Activities and ECERS-R Program Structure scores. This indicates that ECERS-R variables were strongly correlated, and most likely reflect a single construct, which corroborates existing findings (e.g., Cassidy

et al., 2005; Perlman et al., 2004) and suggests that the ECERS-R measures composite classroom quality. In addition, the CIS Sensitivity and CIS Total Scale were highly correlated, also reflecting that they are most likely a single construct, that of the warmth and sensitivity which caregivers bring to their interactions with students. Notably, the correlations between ECERS-R Interaction, CIS Sensitivity, and CIS Total were extremely strong, indicating that the measures tap similar constructs, also that of the warmth and sensitivity that caregivers bring to their interactions with students.

Table 7

Correlations Among Classroom Variables (*N* = 11)

	CIS Sensitivity	CIS Total	ECERS Activities	ECERS Interaction	ECERS Program Structure
CIS Sensitivity	1				
CIS Total	0.97**	1			
ECERS Activity	0.37	0.43	1		
ECERS Interaction	0.88**	0.93	0.46	1	
ECERS Program Structure	0.30	0.36	0.72*	0.56	1

* Significant at the $p < .05$ Level. **Significant at the $p < .001$ level.

HLM Analyses

In order to assess how the relationship between temperament and behavior is expressed in different types of classroom environments, I conducted separate analyses for each dependent variable (e.g., externalizing behavior and social skills). Due to the nested structure of the data (students within classrooms) the analyses used hierarchical linear

regression (HLM). All variables included in the HLM analyses were standardized, which allowed effect sizes to be interpreted using the beta coefficients.

A series of separate HLM analyses or sequential modeling were conducted in order to evaluate the most important predictors and interactions for the dependent variables of interest (externalizing behaviors and social skills). Tables 8 and 9 depict the structure, or sequence, of the fitted models (Model A through E) that will be expanded upon below, for both externalizing behavior and social skills, respectively. Notably, due to limitations in sample size, only two measured variables, CIS Sensitivity and ECERS-R Program Structure, were included at Level 2. These variables were selected based on correlational data (i.e., they demonstrated the lowest correlation among variables in Table 7) suggesting these two variables reflected disparate constructs related to classroom process quality. Although not presented here, separate HLM analyses were conducted to assess all other teacher level variables.

Table 8

Structure of Fitted Models for Externalizing Behavior

Model	Covariates included in Level 1	Covariates included in Level 2
Model a	None	None
Model b	Student Demographics	Teacher Demographics
Model c	Student Demographics TACTIC	Teacher Demographics
Model d	Student Demographics TACTIC	Teacher Demographics CIS Sensitivity ECERS Program Structure
Model e	Student Demographics TACTIC	Teacher Demographics CIS Sensitivity ECERS Program Structure TACTIC x CIS Sensitivity TACTIC x ECERS Program Structure

Table 9

Structure of Fitted Models for Social Skills

Model	Covariates included in Level 1	Covariates included in Level 2
Model a	None	None
Model b	Student Demographics	Teacher Demographics
Model c	Student Demographics TACTIC	Teacher Demographics
Model d	Student Demographics TACTIC	Teacher Demographics CIS Sensitivity ECERS Program Structure
Model e	Student Demographics TACTIC	Teacher Demographics CIS Sensitivity ECERS Program Structure TACTIC x CIS Sensitivity TACTIC x ECERS Program Structure

HLM Fitted Models

Model A. Model A reflected an unconditional model that did not include any predictors. One can think of this model as being equivalent to an ANOVA model. The unconditional model yields a decomposition of the total variance into within- and between-classroom components, which is useful in calculating the amount of variance accounted for when adding Level 1 (student) and Level 2 (classroom) variables.

Model B. Model B controlled for student demographics at Level 1 including gender, ethnicity, disability status, and eligibility for free lunch, as well as whether

English was the primary spoken language. It also controlled for teacher demographics at Level 2, which included the number of years of teaching, the highest degree held, and years of education. Both the number of years of teaching and number of years of education were grand mean centered.

Model C. Model C again included both student and teacher control variables at Levels 1 and 2 respectively, but in addition, included student TACTIC scores as a covariate in Level 1. The TACTIC was grand mean centered, or more specifically, centered around the overall mean of all of the students' TACTIC scores.

Model D. Model D included both student and teacher control variables at Levels 1 and 2, students TACTIC scores at Level 1 (covariate), and classroom CIS Sensitivity and ECERS-R Program Structure scores at Level 2.

Model E. In Model E, or the full model, student and teacher control variables were included at Level 1 and Level 2 respectively, TACTIC scores were included as a covariate in Level 1, classroom variables CIS Sensitivity and CIS Total were entered in Level 2, and interactions between CIS Sensitivity and TACTIC scores, as well as ECERS-R Program Structure and TACTIC were included at Level 2. TACTIC scores and classroom Level variables were again grand mean centered.

Final Models. For both outcome variables (externalizing behavior and social skills), final models were decided on by using model fit statistics. These included the deviance statistics (-2 residual log likelihood, AIC, AICC, and BIC). Normally, full maximum likelihood estimation (FML) is used in HLM to compare models that vary by both fixed and random effects. Using FML estimation provides that advantage of both fixed and random effect model comparisons, where goodness of fit tests are used based

on competing model deviance statistics and the number of model parameters estimated. Unfortunately, due to the small student sample size and minimal number of classrooms used at Level 2 in this study, problems with convergence required overall values of AIC, AICC, and BIC fit statistics based on restricted maximum likelihood (REML) to be used. Subsequently, final models were adjudged by lower overall values of these fit statistics.

Student Externalizing Behavior as Outcome Variable

In the first stage of HLM analyses, student externalizing behavior was used as the dependent variable, and models A through E were analyzed as outlined above. As was described in Chapter 3, to calculate the variance accounted for at the classroom level using externalizing behavior as the outcome, the intraclass correlation was used based on the unconditional model HLM analysis (see p. 102, Formula 1). The results from the unconditional model (Model A) revealed that 18% of the variance in externalizing behavior scores was accounted for at the classroom level, while 82% of the variance in externalizing behavior was accounted for at the student level. This amount of variance accounted for at Level 2 supported the use of HLM. The pseudo-R2, or percent reduction in variance calculated at Level 1 and Level 2 was 33% and 28%, respectively (see p. 103, Formulas 2 and 3). Appendix N presents the model summary.

Table 10 represents the results for Models A through E. None of the student demographics or teacher demographics included at Level 1 (Model B) had a significant effect on externalizing behavior. Free lunch status at the student level, however, as well as the number of years of teacher education at the classroom level, approached significance.

In Models C, D, and E, results indicated that TACTIC had a significant positive effect on externalizing behavior, however no other significant effects were found. This indicates that H04 (that temperament will be a significant positive predictor of externalizing behavior) was supported.

Results did indicate that overall, CIS Sensitivity, as well as the modeled interactions (i.e., CIS Sensitivity x TACTIC), both had negative effects on students' overall externalizing behavior. In addition, ECERS-R Program Structure had positive effects on externalizing behavior; however, these effects were not significant. This indicates that H06, that classroom process quality (ECERS-R Program Structure; CIS Sensitivity) will be a significant negative predictor of externalizing behavior, was partially supported with respect to the relationship between externalizing behavior and one process quality measure, CIS Sensitivity. While not significant, the directionality of the coefficient indicates that such a relationship may be significant with a larger sample size. Notably, given the limitations in the present study's sample size, it was not feasible to test possible interactions between Level 1 and Level 2 variables (i.e., Model E and HO8, that the interaction between child temperament and classroom environment characteristics (ECERS-R Program Structure; CIS Sensitivity) will be significant negative predictors of externalizing behavior). While not significant, the interactions between CIS Sensitivity and TACTIC did show a negative effect on externalizing behavior.

Based on overall fit statistics, Model C was chosen as the final model using student externalizing behavior as the outcome variable. This model included control variables at both the student and teacher level as well as student TACTIC scores as a

covariate in Level 1. In regard to variables at the student level, as can be seen from Table 6, the TACTIC covariate had a significant positive effect on students' externalizing behavior. More concisely, higher TACTIC scores led to higher student externalizing behavior, supporting H04 that predicted this result. No significant effects were found for any of the demographic variables entered at the student level, however, both ethnicity and disability status had positive effects on students externalizing behavior, and both student gender and free lunch status were found to have negative effects on externalizing behavior. More specifically, those students classified with a disability status had higher reported externalizing behavior, and females, as well as those students with reported free lunch status exhibited lower overall externalizing behavior scores. At the classroom level, all teacher control variables (years of teaching, years of education, and teacher ethnicity) were found to be nonsignificant yet positive.

Table 10

HLM Estimates with Standard Errors for Externalizing Behavior as Outcome

Fixed Effects	Model A	Model B	Model C	Model D	Model E
Intercept	.051 (.154)	.679 (.536)	.219 (.508)	.244 (.590)	.250 (.642)
Ethnicity		.039 (.498)	.045 (.047)	.046 (.048)	.043 (.049)
Gender		-.092	-.053 (.136)	-.047 (.137)	-.067 (.144)
Lunch		(.164)	-.700†(.358)	-.842 (.587)	-.834 (.623)
Disability Status		-	.090 (.195)	.088 (.196)	.108 (.200)
Years Teaching		.785†(.380)	.005 (.022)	.009 (.034)	.009 (.037)
Years of Education		.299 (.232)	.032 (.097)	.041 (.109)	.052 (.118)
Teacher Ethnicity		-.041 (.022)	.079 (.091)	.109 (.113)	.110 (.123)
TACTIC		.216† (.096)	.640** (.090)	.651** (.091)	644**(.094)
CIS Sensitivity		-.039		-.156 (.235)	-.181 (.256)
ECERS PS		(.091)		.081 (.820)	.093 (.118)
CIS SENS*TACTIC					-.051 (.664)
ECERS PS*TACTIC					-.005 (.134)
Random Effects					
Intercept	.188† (.118)	.120 (.110)	.135† (.102)	.180 (.149)	.214 (.183)
Residual	.832 (.108)	.814 (.107)	.561 (.074)	.562 (.074)	.568 (.076)
Model Fit Statistics	Model 1	Model 2	Model 3	Model 4	Model 5
Deviance	360.00	366.50	325.20	327.00	332.10
AIC	364.00	370.50	329.20	331.00	336.10
AICC	364.10	370.60	329.30	331.10	336.20
BIC	364.80	371.30	330.00	331.80	336.90

* $p < .05$. ** $p < .001$. †Approaches Significance.

Social Skills as Outcome Variable

Table 11 represents the results for Models A through H as well as a final model using student social skills as the outcome variable. Again, the intraclass correlation was used to calculate the variance accounted for at the classroom level based on the unconditional model HLM analysis. Results indicated that for the unconditional model (Model A), 6% of the variance in social skills scores was accounted for at the classroom level (see p. 95, Formula 1). The pseudo-R2, or percent reduction in variance calculated at Level 1 and Level 2 was 35% and 16%, respectively (see p. 96, Formulas 2 and 3). Appendix N presents the model summary.

In Model B, results indicated that disability status had a significant negative effect on students' social skills. In addition, modeling TACTIC scores as a student level covariate (Model C) also resulted in significant negative effects on social skills. This indicates that H05 was supported, that child temperament is in fact a significant negative predictor of social skills.

The inclusion of teacher demographics at Level 2, as well as teacher CIS Total and CIS Sensitivity scores revealed no significant effects (Model C and Model D). In Model E, ECERS-R Program Structure was included at Level 2, and results revealed that ECERS-R Program Structure had a positive effect on social skills that approached significance. Given the study's sample size, this coefficient would most likely reach significance in a larger sample. This indicates that H07 (i.e., that classroom process quality will be a significant positive predictor of social skills) was partially supported with respect to the relationship between social skills and ECERS-R Program Structure.

Subsequently, Model D was chosen as the final model using the lowest overall model fit statistics. This model included student disability status and TACTIC at Level 1, as well as the ECERS-R Program Structure at the classroom level, supporting H05 and partially supporting H07 (i.e., temperament will be a significant negative predictor of social skills; classroom process quality will be a significant positive predictor of social skills, respectively). As can be seen from Table 7, at the student level, the TACTIC covariate as well as student disability status had significant negative effects on students' social skills. More specifically, higher TACTIC scores were associated with lower social skills, as well as student classification with a disability status. No other significant effects were found for any other control variables entered at the student level. However, results suggest that being female was associated with slightly increased social skills scores, and students' classification as having free lunch status was related to slightly lower social skills scores. At the classroom level, similar to findings in the final externalizing behavior model, all teacher control variables (years of teaching, years of education, and teacher ethnicity) were nonsignificant.

Given the limitations in the present study's sample size, it was not possible to test possible interactions between Level 1 and Level 2 variables (HO9, that interaction between temperament and classroom environment characteristics will be significant positive predictors of social skills). However, while not significant, the interactions between CIS Sensitivity and TACTIC had a positive effect on social skills (see Table 11).

Table 11

HLM Estimates with Standard Errors for Social Skills as Outcome

Fixed Effects	Model A	Model B	Model C	Model D	Model E
Intercept	-.018 (.114)	.258 (.410)	.759 (.513)	1.028*(.460)	1.145† (.528)
Ethnicity		-.061 (.058)	-.061 (.050)	-.046 (.050)	-.037 (.050)
Gender		.115 (.170)	.092 (.520)	.086 (.143)	.128 (.149)
Lunch		.300 (.278)	.134 (.362)	-.374 (.456)	-.524 (.519)
Disability Status		-.700* (.237)	-.534* (.205)	-.493* (.204)	-.546*(.207)
Years Teaching		.030† (.015)	-.011 (.022)	.028 (.025)	.033 (.028)
Years of Education		-.033 (.062)	.149 (.097)	.133 (.076)	.112 (.089)
Teacher Ethnicity		-0.79 (.059)	-.204† (.090)	-.169† (.077)	-.182 (.110)
TACTIC			-.612** (.094)	-.665** (.166)	-.657** (.097)
CIS Sensitivity				-.048 (.166)	-.001 (.195)
ECERS PS				.511† (.253)	.536 (.286)
CIS SENS*TACTIC					.148 (.121)
ECERS PS*TACTIC					-.044 (.107)
Random Effects					
Intercept	.061 (.067)	.006 (.047)	.126 (.114)	.051 (.076)	.084 (.100)
Residual	.946 (.123)	.870 (.114)	.619 (.082)	.611 (.081)	.607 (.081)
Model Fit Statistics	Model 1	Model 2	Model 3	Model 4	Model 5
Deviance	369.30	369.10	336.3	333.20	337.00
AIC	373.30	373.10	340.3	337.20	341.00
AICC	373.40	373.20	340.4	337.30	341.10
BIC	374.10	373.90	341.1	338.00	341.80

* $p < .05$. ** $p < .001$. †Approaches Significance.

Summary of Findings Related to Study Hypotheses

Table 12 presents this study's hypotheses and indicates that each hypothesis tested (only the first seven of the nine hypotheses could be tested) was either supported or partially supported. H01 through H03 were tested through correlations between student variables. H01 asserted that there would be a significant negative relationship between temperament and social skills. A correlation of $r = -0.43$, significant at the $p < .001$ level was found between these constructs, supporting this hypothesis. H02 indicated that there would be a significant positive relationship between temperament and externalizing behavior. A correlation of $r = 0.53$, significant at the $p < .001$ level was found between these constructs, also supporting this hypothesis. These findings indicate that more difficult child temperament is associated with higher levels of externalizing behavior problems, and lower levels of social skills. H03 noted that there would be a significant negative relationship between social skills and externalizing behavior. A correlation of $r = -0.44$, significant at the $p < .001$ level was found between these constructs, supporting this hypothesis and indicating an inverse relationship between the externalizing behavior and social skills, the outcome variables in this study.

H04 through H07 were tested through a series of HLM analyses. H04 indicated that temperament would be a significant positive predictor of externalizing behavior. In the final model, as well as in all other tested models where TACTIC scores were included as a covariate in Level 1, child temperament was found to have a significant positive effect on externalizing behavior, supporting this hypothesis. H05 asserted that temperament would be a significant negative predictor of social

skills. In the final model, as well as in all other tested models where TACTIC scores were included as a covariate in Level 1, child temperament was found to have a significant negative effect on externalizing behavior, also supporting this hypothesis.

H06 indicated that classroom process quality, defined in terms of ECERS-R Program Structure and CIS Sensitivity scores, would be a significant negative predictor of externalizing behavior. While both variables were not found to be significant predictors in entered models, CIS Sensitivity had negative effects on students' overall externalizing behavior. Notably, the directionality of the coefficient is in line with the study's hypothesis, suggested that such a relationship may reach significance with a larger sample size, partially supporting H06. H07 noted that classroom process quality, defined by ECERS-R Program Structure and CIS Sensitivity scores, would be a significant positive predictor of social skills. ECERS-R Program Structure was found to have a positive effect on social skills which approached significance, and as such was included in the final model at the classroom level, partially supporting H07.

I did not test H08 and H09. H08 indicated that the interaction between temperament and classroom environment characteristics (defined in terms of ECERS-R Program Structure and CIS Sensitivity scores) would be significant negative predictors of externalizing behavior. While not significant, the interaction between CIS Sensitivity and temperament (TACTIC) did show a negative effect on externalizing behavior. H09 asserted that the interaction between temperament and classroom environment characteristics would be significant positive predictors of social skills. While not significant, the interaction between CIS Sensitivity and

temperament (TACTIC) was also found to have a positive effect on social skills. Given the limitations in sample size, finding significant interactions between Level 1 and Level 2 variables was unlikely. Therefore, I contend that it was not possible to test H08 and H09.

Table 12
Overview of Study Hypotheses

H0 Number	Study Hypothesis	Supported/ Not Supported
H01	There will be a significant negative relationship between temperament (TACTIC Total Temperament score) and social skills (PKBS-2 Social Skills subscale score).	Supported
H02	There will be a significant positive relationship between temperament and externalizing behavior (PKBS-2 Externalizing Behavior Problems subscale score).	Supported
H03	There will be a significant negative relationship between social skills and externalizing behavior.	Supported
H04	Temperament will be a significant positive predictor of externalizing behavior.	Supported
H05	Temperament will be a significant negative predictor of social skills.	Supported
H06	Classroom process quality (ECERS-R Program Structure; CIS Sensitivity) will be a significant negative predictor of externalizing behavior.	Partially Supported
H07	Classroom process quality will be a significant positive predictor of social skills.	Partially Supported
H08	The interaction between temperament and classroom environment characteristics (ECERS-R Program Structure; CIS Sensitivity) will be significant negative predictors of externalizing behavior.	Not Tested
H09	The interaction between temperament and classroom environment characteristics will be significant positive predictors of social skills.	Not Tested

CHAPTER 5

Discussion

This chapter describes the key findings obtained from the statistical analyses in the present study, as well as educational implications of the findings, limitations of this study, and directions for future research.

Key Findings

The present study sought to determine how the relationship between child temperament and behavior is expressed in different types of classroom environments in prekindergarten settings. Other goals of the study were to further operationalize the goodness of fit model in school settings and to evaluate possible interactions of process variables indicative of classroom quality with child temperament to see if these interactions predicted child behavior and social skills. Overall, the study revealed that child temperament is a highly important correlate and predictor of both student social skills and externalizing behavior. In addition, the study indicated that classroom quality, as described by the manner in which an educational program is structured and in terms of the warmth and sensitivity with which caregivers interact with their students, is related to child outcomes in behavioral and social domains. Finally, the study indicated that the goodness of fit model when operationalized in terms of the transactional relationship between temperament and environmental demand factors of characteristics of the classroom setting (as informed by the classroom quality literature) has predictive value and describes child behavioral and social outcomes in prekindergarten settings.

Generally, the hypotheses of this study were supported by the results. The discussion of the results of this study is organized in terms of findings related to the two

main constructs in this study: temperament and goodness of fit. The first part discusses the results related to child temperament. The second part discussed results related to the goodness of fit model, and is organized in terms of the two outcome variables used in the study: child externalizing behavior and child social skills.

Temperament

Following the model of Thomas, Chess, and Birch (1968), results indicated that children in this study's sample demonstrated typical levels of temperamental characteristics. In addition, as expected, results indicated that difficult temperament is positively correlated with child externalizing behavior and negatively correlated with child social skills (e.g., Caspi & Silva, 1995; Guerin et al., 1997; Thomas et al., 1968). The various temperamental dimensions that comprise the difficult temperament profile have been shown to predispose a child to behavior and adjustment problems (Keogh, 2003). Although there are some differences in the findings, characteristics of the difficult temperament profile have consistently shown to be important contributors to children's behavioral adjustment problems (e.g., Klein, 1982; Prior et al., 1999; Telgasi & MacMahon, 1990; Thomas et al., 1968; Wertlieb et al., 1987).

Goodness of Fit

Given that the operationalization of the goodness of fit model has been inconsistent, particularly within school settings, this study did so in terms of demand characteristics of classroom settings based on the early childhood classroom quality literature and on the recommendations of Keogh (2003; Keogh & Speece, 1996). A series of separate HLM analyses or sequential modeling examined the goodness of fit construct in this study. The most important predictors and interactions for the dependent variables

of interest in this study, externalizing behaviors and social skills, were evaluated. The classroom level variables used in the analysis, caregiver sensitivity and program structure, were selected based on correlational data suggesting these two variables reflected disparate constructs related to classroom quality.

Externalizing Behavior. This section will discuss the results of one of this study's outcome variables, externalizing behavior. It will discuss significant predictors as well as those which approached significance and those of interest in the present study.

Consistent with the literature, results indicated that difficult child temperament had a significant positive effect on child externalizing behavior, where higher difficult temperament scores predicted higher levels of student externalizing behavior. These results are consistent with past research that shows that the five major temperament dimensions that comprise the difficult temperament profile are related to behavior problems: negative mood combined with irregularity, nonadaptability, and negative withdrawal responses; excessive persistence; withdrawal and mild negative reactions to new situations combined with slow adaptability; extreme distractibility; and very high or low activity level (Thomas, Chess, & Birch, 1968). A large body of research has followed the NLYS, indicating that these temperament characteristics continue to be related to behavior problems (e.g., Hagekull & Bohlin, 1981; Maziade et al., 1990; Sanson et al., 2004; Teglasi & MacMahon, 1990). As such, the findings of this study are consistent with the existing literature and further illustrate the powerful effect temperament has on behavior.

The results of this study did not find any student demographic variables to significantly affect child externalizing behavior. One control variable, a student's free

lunch status, was found to approach significance in the final model. This finding, that receiving free lunch negatively predicted externalizing behavior is an artifact of limitations in the sample, as all students in two out of the three participating schools received free lunch as part of their educational program.

At the classroom level, all teacher demographic variables, used as controls (i.e., years of teaching, years of education, and teacher ethnicity) were found to have positive effects on child externalizing behavior. However, these effects were not significant. These variables, indicative of structural markers of classroom quality were not expected to reach significance, as the literature indicates these variables to have an indirect influence on child outcomes (e.g., Early et al., 2005; Pianta et al., 2005).

When considering the demand factors of the classroom environment in the goodness of fit model, two process variables at the classroom level were used, caregiver sensitivity and program structure. Results indicated that as predicted, caregiver sensitivity had negative effects on students' overall externalizing behaviors. While not significant, the directionality of the coefficient indicated that such a relationship may be significant with a larger sample size. This finding indicates a relationship between these variables.

This finding, that higher levels of teacher sensitivity was related to more positive child outcomes in behavioral and social domains has been found across age groups and educational settings (e.g., Howes, Matheson, & Hamilton, 1994; Kontos & Wilcox-Herzog, 1997; Pianta, 1999; Pianta et al., 2002). Because the present study was the first to operationalize the goodness of fit in such a manner, this finding is the first of its specific kind. However, it extends previous findings that indicated that child temperament and access to appropriate caregiving behaviors predict affective and

behavioral outcomes (De Schipper et al.; 2004; Hestenes et al., 1993; Volling & Feagans, 1995).

Program structure was not found to have a significant effect on externalizing behavior. This finding is in contrast to research that indicates that classroom organization, a similar construct to that of program structure predicts change in behavioral adjustment and learning behavior (Dominguez, Vitiello, Maier, & Greenfield, 2010). It may be that correctly structuring a classroom allows for the acquisition of new socials skills, versus for the prevention of problem behaviors.

When modeling interactions between child temperament and classroom variables (i.e., caregiver sensitivity and program structure) to predict child externalizing behavior, the investigator found no significant effects. While the present study sought to examine the goodness of fit model by looking at such terms, due to study limitations, the conclusion could not be made that these interactions did not predict child behavioral outcomes. The present study took a novel approach to examine the goodness of fit model by looking at the interactions between child temperament and classroom environment characteristics identified by the classroom quality literature. While no significant effects were found for such terms, a study conducted by De Schipper and colleagues (2004) lends support to such an operationalization. Their results indicated the interaction between child temperament and daycare arrangement stability, a similar construct to that of program structure, significantly predicted child behavior problems. In addition, the study found that the interaction between child temperament and caregiver availability, a similar construct to that of caregiver sensitivity, had significant effects on a child's wellbeing in daycare.

Social Skills. This section will discuss the results that pertain to this study's second outcome variable, social skills. It will discuss significant predictors as well as those which approached significance and those of interest to the present study.

As expected and as with found with externalizing behavior, the difficult child temperament had a significant negative effect on students' social skills. This is consistent with the literature that shows clear linkages between child temperament and socially competent functioning (Sanson et al., 2004). While the relationship between temperament and social skills has received less attention than associations with behavior problems, the vast majority of the literature indicates direct linear effects of temperament on social competence (Rothbart & Bates, 1998; Sanson et al., 2004; Seifer, 2000). For example, preschool students who demonstrate the easy temperament profile are more likely to display socially competent behavior than their more difficult peers (Farver & Branstetter, 1994).

In contrast to the findings for externalizing behavior, one student control (i.e., student disability status) significantly negatively predicted child social skills. This indicates that students with disabilities displayed lower levels of social skills than their nondisabled peers. This study did not consider disabilities by type, as preschool children do not receive a specific classification, and are referred to as "a preschooler with a disability." However, it is generally accepted that children with disabilities often have deficient social skills, but the exact nature of these deficits is unclear (Fussell, Macias, & Saylor, 2005). Research indicates that the acquisition of social skills may differ in children with disabilities. In a series of studies, Guralnick and colleagues found that preschool children with disabilities interacted less when compared to their nondisabled

peers (Guralnick & Weinhouse, 1984), were less successful in their social bids to peers (Guralnick, Connor, Hammond, Gottman, & Kinnish, 1996a), appeared to take the lead in interactions less often (Guralnick, Connor, Hammond, Gottman, & Kinnish, 1996b), and developed fewer friendships (Guralnick, Gottman, & Hammond, 1995; Guralnick & Groom, 1988). As such, the results of this study are consistent with existing literature, and indicated that if children display temperamental characteristics typical of the difficult temperament profile, and have a disability, they are likely to display lower levels of social skills in prekindergarten environments.

At the classroom level, similar to findings in the final externalizing behavior model, all teacher control variables indicative of classroom structural variables (years of teaching, years of education, and teacher ethnicity) were found to be nonsignificant. Again, this is consistent with the classroom quality literature that indicates that structural descriptions of classroom quality are insufficient predictors of children social outcomes, as they are thought to only act as proxy measures (e.g., Early et al. 2005; Pianta et al., 2005). Unexpectedly, teacher ethnicity negatively predicted student social skills; however, this effect was not significant. This finding is most likely due to study limitations and is in contrast to research the demonstrates that teacher minority status either has null (e.g., Pianta et al., 2002) or positive (e.g., Mashburn, Hamre, Downer, & Pianta, 2006) effects on children's behavioral and social outcomes.

When considering the demand factors of the classroom environment in the goodness of fit model, two variables were considered, caregiver sensitivity and program structure. Results indicated that one of the assessed variables, program structure had a positive effect on child social skills that approached significance. Given the study's

sample size, this coefficient would most likely reach significance in a larger sample. This indicates that when an educational program is structured in a developmentally appropriate manner, students were more likely to demonstrate higher levels of social skills. This finding is particularly interesting given that this was not found for externalizing behavior. It may be that more appropriate structure allows for students to acquire and then encourages the use of appropriate social skills.

This study is the first to specifically consider how an educational program's structure contributes to a child's social skills. However, these results are in accord with existing research that suggests that classrooms that score higher on global quality assessments, described in part by how a program is structured are more likely to have students who demonstrate higher levels of social competence (e.g., Bryant et al., 1994; Hestenes et al., 1993; Lambert et al., 2002; Peisner-Feinberg & Burchinal, 1997). In addition, as previously indicated, a recent study conducted by Dominguez and colleagues (2010) found classroom organization, a similar construct to that of program structure, to predicts change in behavioral adjustment and learning behavior. Moreover, it is generally accepted that a classroom that is structured in a developmentally appropriate manner, where children are provided with a general routine, given ample opportunities for free play and small-group work, and include provisions for children with disabilities lends itself to positive child outcomes (Peisner-Feinberg et al., 2001; Pianta, 1999; Pianta et al., 2002; Vandell, 2004)

As with externalizing behavior, none of the interaction terms, assessing how child temperament interacts with classroom variables significantly affected child social skills. However, due to study limitations, the conclusion cannot be made that these interactions

do not predict child behavioral outcomes. In addition, the study conducted by De Schipper and colleagues (2004) suggests that the interaction between child temperament and classroom environment characteristics may have value in predicting social outcomes in children.

Educational Implications

This study sought to operationalize the goodness of fit model in school settings by considering demand factors of the classroom setting. The results of this study indicated that child temperament and disability status as well as school program structure significantly predict child social skills, and that child temperament alone predicted externalizing behavior. While temperament itself is not amenable to intervention, school psychologists can in fact target school program characteristics to improve child outcomes and promote social skill dvelopment. For example, programs may be better structured to permit more small-group interaction, free-play focused and child-directed activities, and teachers may receive in-service training to increase sensitivity and use of appropriate discipline strategies. In addition, the results of this study indicated the fundamental need for students with disabilities to receive explicit social skills training, and in particular, those students who also have difficult temperaments should receive this training. Furthermore, the results of this study provided greater insight into the functional relationship between child temperament, the school environment, and child behavior and social skills, which can allow for temperament-environment interactions to be targeted for intervention in school settings.

Limitations of the Current Study

There are several limitations to the current study, which attempted to operationalize the goodness of fit model and determine to how the relationship between child temperament and behavior is expressed in different types of classroom environments in prekindergarten settings. First, a major study limitation was sample size given this study's purpose and analysis procedures. Having such a low sample size of students, classrooms, and teachers reduced the amount of variability at the second level. This had potential effects on the amount of variance accounted in the model. In addition, this also reduced the power to detect significant effects.

A second limitation, related to sample size and limited level two variance is that classrooms that participated in this study were somewhat homogeneous. There were 11classrooms and 3school sites that participated in this study, of which 2of the sites and 9 of the classrooms were urban Head Start universal prekindergarten settings. In these classrooms, the majority of the students were of Hispanic descent and received free lunch as part of their educational program (all Head Start students receive meals at school). This affected results related to student controls and their effects on child outcomes.

Related to the homogeneous nature of the sample is a third limitation. Given the nature of research procedures and recruitment of this study, the schools who agreed to participate were those who were organized enough to do so. A fundamental question asked in this study was how the relationship between temperament and behavior is expressed in different types of classrooms (i.e., those of varying levels of quality). Because the classrooms that were sampled in this study were (a) homogeneous, and (b) of overall high quality, it was difficult to find effects at the second level.

A fourth limitation stemming from this small sample size is that only two predictors were used at the classroom level. I had measured five classroom variables and initial plans were to model all of these. However, this was not possible due to the small number of classrooms that participated in this study. Given that only two classroom variables were used, program structure and caregiver sensitivity, only limited conclusions can be made about such an operationalization of the goodness of fit model.

A final limitation of this study was that only teacher report was used to measure child temperament and behavior. Accordingly, these reports may be affected by response sets and rater bias, such as the halo effect. These concerns are the result of studies that indicate poor correspondence between caregiver report and direct observation (correlations in the .20-.40 range) (e.g., Vaughn et al., 1987). However, it may be that teachers' views of children's behavior and temperaments are selective, reflect individual expectations for behavior and conduct, and/or change over time (Keogh, 2003).

Directions for Future Research

The future directions of research are aimed to rectify and compensate the identified limitations of the current study. First, future research on this topic should attempt to recruit as many diverse classrooms as possible to increase the power to detect significant effects, particularly the interaction between student- and classroom-level variables. Heterogeneity at the classroom level would mean recruiting classrooms in a variety of different settings (e.g., public, private, urban, rural) that have disparities in the nature and type of available resources and include various types of students and teachers. Specifically, sampled classrooms must vary as much as possible in classroom characteristics and quality. Second, research has suggested that scales measuring

classroom quality measure an unidimensional construct (e.g., Cassidy et al., 2005; Perlman et al., 2004) Therefore, future research may consider using a total scale versus subscale scores to reduce the number of predictors at the classroom level. Third, studies considering the goodness of fit theory should use both caregiver-reported and observational assessments of child temperament and behavior to control for rater biases and response sets. Fourth, future research may consider how child temperament and classroom environment characteristics predict other behavioral and affective outcomes, like internalizing problems. Investigating more disparate child outcomes like internalizing problems might help to decrease concerns of conceptual over lap between child temperament and behavior. Finally, future research may consider how individual temperamental characteristics (versus constellations), classroom environment characteristics, and the interaction of the two predict children's social and behavioral outcomes.

Conclusion

The present study sought to determine how the relationship between child temperament and behavior is expressed in different types of classroom environments in prekindergarten settings. Other goals of the study were to further operationalize the goodness of fit model in school settings and to evaluate possible interactions of process variables indicative of classroom quality with child temperament to see if these interactions predict child behavior and social skills.

Overall, the study further illustrated that child temperament is a highly important correlate and predictor of both student social skills and externalizing behavior. In addition, the study indicated that classroom environment characteristics and quality are

related to child outcomes in behavioral and social domains. Child temperament alone was found to be the sole significant predictor of child externalizing behavior, while child temperament, disability status, and school program structure predicted child social skills. Interestingly, caregiver characteristics, such as the sensitivity with which they interact with their students were more related to externalizing behavior, and program characteristics, such as the manner in which an educational program integrates a developmentally-appropriate structure were related to child social skills. Finally, the study indicated that the goodness of fit model when operationalized in terms of the transactional relationship between temperament and environmental demand factors of characteristics of the classroom setting (as informed by the classroom quality literature) has predictive value and describes child behavioral and social outcomes in prekindergarten settings.

Appendix A

Data from Pilot (Blackwell, 2008)

Table 1

Means and Standard Deviations of Classroom Environment Ratings

	Class 1	Class 2
ECERS-R Activities	4.73	4.50
	(.14)	(.00)
ECERS-R Interaction	5.10	6.50
	(.42)	(.14)
ECERS-R Program	4.00	4.25
Structure	(.00)	(.35)
CIS Sensitivity	2.50	3.18
	(.14)	(.18)
CIS Total Score	3.13	3.43
	(.12)	(.12)

*Note: Standard Deviations appear in parentheses below means.

Table 2

Alpha levels for Environmental Rating Scales

	Inter-rater Reliability Time 1	Inter-rater Reliability Time 2	Test-Retest Reliability
ECERS-R Activities	1.00	0.96	0.96
ECERS-R Interaction	1.00	1.00	0.92
ECERS-R Program Structure	1.00	1.00	1.00
CIS Sensitivity	0.97	0.95	0.96
CIS Total Score	0.96	0.98	0.98

Table 3

Classroom Structural Information

	Class 1	Class 2
Teacher Received Degree	Bachelors	Masters
Teacher Number of Years of Education	16	20
Teacher Number of Years of Teaching	18	21
Child to Teacher Ratio	15:2	14:2

Table 4

Child Behavior and Social Skills Means, Standard Deviations, and Effect Cohen's d

	Class 1	Class 2	*t*	*df*	*Sig.*	*Effect Cohen's d with pooled SD*
Externalizing Problems	85.13 (13.94)	85.14 (23.68)	-.001	27	.999	.002
Social Skills	110.07 (13.63)	104.36 (11.20)	1.23	27	.230	1.62

*Note: Standard Deviations appear in parentheses below means.

Table 5

Means and Standard Deviations of Temperament Variables

	Class 1	Class 2	*t*	*df*	*Sig.*	*Effect Cohen's d with pooled SD*
Activity Level	-0.95 (1.00)	-0.29 (1.41)	-1.46	27	.16	.60
Adaptability	-1.09 (1.10)	-.034 (1.14)	-1.80	27	.08	1.00
Mood	-0.16 (0.76)	-0.17 (1.20)	.04	27	.97	.01
Persistence	-0.99 (0.76)	-0.43 (0.99)	.30	.27	.12	.60

*Note: Standard Deviations appear in parentheses below means. Mean scores are *z*-scores.

Table 6

Temperament and Child Behavior and Social Skills – Class 1

Variables		1	2	3	4	5
1	Activity Level					
2	Adaptability	0.98**				
3	Mood	0.70**	0.71**			
4	Persistence	0.90**	0.92**	0.49		
5	Externalizing Problems	0.94**	0.98**	0.74**	0.86**	
6	Social Skills	-0.84**	-0.83**	-0.67**	-0.77**	-0.81**

** $p < 0.01$.

Table 7

Z-scores from Correlations – Class I

Variables		1	2	3	4	5
1	Activity Level					
2	Adaptability	2.30				
3	Mood	0.87	0.89			
4	Persistence	1.47	1.59	0.54		
5	Externalizing Problems	1.74	2.30	0.95	1.29	0.00
6	Social Skills	-1.22	-1.19	-0.81	-1.02	-1.13

Table 8

Temperament and Child Behavior and Social Skills – Class 2

Variables		1	2	3	4	5
1	Activity Level					
2	Adaptability	0.87**				
3	Mood	0.82**	0.84**			
4	Persistence	0.73**	0.63**	0.57*		
5	Externalizing Problems	0.54**	0.11	0.28	0.39	
6	Social Skills	-0.68**	-0.61**	-0.62*	-0.57*	-0.40

* $p < 0.05$, ** $p < 0.01$.

Table 9

Z-Scores from Correlations – Class 2

Variables		1	2	3	4	5
1	Activity Level					
2	Adaptability	1.33				
3	Mood	1.16	1.22			
4	Persistence	0.93	0.74	0.65		
5	Externalizing Problems	0.60	0.11	0.29	0.41	
6	Social Skills	-0.83	-0.71	-0.73	-0.65	-0.42

Table 10

Difference between Class 1 and 2 Z-scores

Variables		1	2	3	4	5
1	Activity Level					
2	Adaptability	0.96				
3	Mood	-0.29	-0.33			
4	Persistence	0.54	0.85	-0.11		
5	Externalizing Problems	1.13	2.19	0.66	0.88	
6	Social Skills	-0.39	-0.48	-0.09	-0.37	-0.70

Table 11

Differences Divided by the SE (.4174) for Significance

Variables		1	2	3	4	5
1	Activity Level					
2	Adaptability	2.31*				
3	Mood	-0.69	-0.80			
4	Persistence	1.30	2.03*	-0.27		
5	Externalizing Problems	2.72**	5.24**	1.59	2.11*	
6	Social Skills	-0.94	-1.15	-0.21	-0.89	-1.69

>1.96 = .05 level sig, >.258 = .01 level sig.

Table 12

Regression Analyses for Variables Predicting Externalizing Problems

	B	*SE B*	B	*t*	Sig.	R^2
Class 1						
Activity Level	13.08	1.34	.94	9.76	***	.88
Adaptability	12.43	.75	.98	16.51	***	.95
Mood	13.50	3.46	.74	3.91	**	.54
Persistence	15.66	2.61	.86	6.01	***	.74
Class 2						
Activity Level	9.05	4.10	.54	2.21	*	.29
Adaptability	2.34	5.91	.11	.39	ns	.01
Mood	5.42	5.46	.28	.99	ns	.08
Persistence	9.34	6.35	.39	1.47	ns	.15

*$p < .05$, **$p < .01$, ***$p < .001$.

Table 13

Regression Analyses for Variables Predicting Social Skills

	B	*SE B*	B	*t*	Sig.	R^2
Class 1						
Activity Level	-11.44	2.06	-.84	-5.56	***	.70
Adaptability	-10.29	1.94	-.83	-5.32	***	.69
Mood	-12.06	3.69	-.67	-3.26	**	.45
Persistence	-13.75	3.16	-.77	-4.35	***	.59
Class 2						
Activity Level	-5.40	1.69	-.68	-3.20	**	.46
Adaptability	-5.92	2.23	-.67	-2.65	**	.37
Mood	-5.73	2.12	-.62	-2.71	**	.38
Persistence	-6.46	2.68	-.57	-2.41	**	.34

*$p < .05$, **$p < .01$, ***$p < .001$.

Appendix B

Script to Inform School Administrators about the Study

"Hi, my name is Sasha Collins Blackwell and I am a school psychology doctoral student from The Graduate Center of the City University of New York. For my doctoral dissertation, I am conducting a study to determine if children's characteristics are expressed the differently in classroom environments that have different characteristics. I hope to find a number of schools and prekindergarten classrooms to participate in this study. To tell you a little about what this study will entail, I first will need to receive approval for the study from The Graduate Center of the City University of New York's internal review board as well as from your school to begin the study. I then will need to receive written parent consent of all the students in the classrooms who will be participating as well as teacher consent. When I have consent from all the necessary people, I will begin my research. The research entails me and a research assistant observing each classroom once for about an hour and a half. In addition, I will ask the classroom teachers to fill-out two rating scales per child (it will take the teacher about 30 minutes to fill-out both forms per child). One rating scale assesses child temperament; the other rating scale assesses child behavior. I will also collect demographic information from teachers and parents using a brief survey. I will reimburse the teachers for their time and effort. This money can be used towards the classroom budget. I hope your school is interested in participating in this research. I am open to suggestions, and will revise and plan according to your school's concerns and needs."

Appendix C

Student Demographics by Classroom

Table 1

Student Demographics - Classroom 1

Student Demographics	Variable	*N*	%
Gender	Male	3	37.5
	Female	5	62.5
Ethnicity	Black	2	25.0
	Asian	0	0
	Hispanic	5	62.5
	White	0	0
	Other	1	12.5
Disability Status	No	8	100
	Yes	0	0
Free Lunch	No	0	0
	Yes	8	100
English Primary Language	No	3	37.5
	Yes	5	62.5

	N	Range	M	SD
Age in Months	8	36 - 60	49.84	9.34

Table 2

Student Demographics - Classroom 2

Student Demographics	Variable	*N*	%
Gender	Male	3	27.3
	Female	8	72.7
Ethnicity	Black	6	65.5
	Asian	0	0
	Hispanic	4	36.4
	White	0	0
	Other	1	9.1
Disability Status	No	10	90.9
	Yes	1	9.1
Free Lunch	No	0	0.0
	Yes	11	100.0
English Primary Language	No	3	27.3
	Yes	8	72.7

	N	Range	M	SD
Age in Months	11	38 - 62	49.35	6.19

Table 3

Student Demographics - Classroom 3

Student Demographics	Variable	*N*	%
Gender	Male	6	42.9
	Female	8	57.1
Ethnicity	Black	8	57.1
	Asian	0	0
	Hispanic	5	35.7
	White	0	0
	Other	1	7.1
Disability Status	No	13	92.9
	Yes	1	7.1
Free Lunch	No	0	0.0
	Yes	14	100.0
English Primary Language	No	5	35.7
	Yes	9	64.3

	N	Range	M	SD
Age in Months	14	37 - 64	50.12	6.99

Table 4

Student Demographics - Classroom 4

Student Demographics	Variable	*N*	%
Gender	Male	10	55.6
	Female	8	44.4
Ethnicity	Black	1	5.6
	Asian	0	0.0
	Hispanic	17	94.4
	White	0	0.0
	Other	0	0.0
Disability Status	No	10	55.6
	Yes	8	44.4
Free Lunch	No	0	0.0
	Yes	18	100.0
English Primary Language	No	13	72.2
	Yes	5	27.8

	N	Range	M	SD
Age in Months	18	41 - 62	51.83	6.86

Table 5

Student Demographics - Classroom 5

Student Demographics	Variable	*N*	%
Gender	Male	6	54.5
	Female	5	45.5
Ethnicity	Black	3	27.3
	Asian	0	0.0
	Hispanic	7	63.6
	White	0	0
	Other	1	9.1
Disability Status	No	8	72.7
	Yes	3	27.3
Free Lunch	No	1	9.1
	Yes	10	90.
English Primary Language	No	6	54.5
	Yes	5	45.5

	N	Range	M	SD
Age in Months	11	41 - 66	53.68	8.00

Table 6

Student Demographics - Classroom 6

Student Demographics	Variable	*N*	%
Gender	Male	5	55.6
	Female	4	44.4
Ethnicity	Black	3	33.3
	Asian	1	11.1
	Hispanic	4	44.4
	White	1	11.1
	Other	0	0.0
Disability Status	No	6	66.7
	Yes	3	33.3
Free Lunch	No	0	0.0
	Yes	9	100.0
English Primary Language	No	3	33.3
	Yes	6	66.7

	N	Range	M	SD
Age in Months	9	37 - 65	53.40	11.89

Table 7

Student Demographics - Classroom 7

Student Demographics	Variable	*N*	%
Gender	Male	7	53.8
	Female	6	46.2
Ethnicity	Black	2	15.4
	Asian	1	7.7
	Hispanic	9	69.2
	White	0	0.0
	Other	1	7.7
Disability Status	No	10	23.1
	Yes	3	76.9
Free Lunch	No	13	100.0
	Yes	0	0.0
English Primary Language	No	7	53.8
	Yes	6	46.2

	N	Range	M	SD
Age in Months	13	48 - 62	54.60	5.48

Table 8

Student Demographics - Classroom 8

Student Demographics	Variable	*N*	%
Gender	Male	9	60.0
	Female	6	40.0
Ethnicity	Black	5	
	Asian	0	33.3
	Hispanic	10	66.7
	White	0	0.0
	Other	0	0.0
Disability Status	No	14	93.3
	Yes	1	6.7
Free Lunch	No	0	0.0
	Yes	1	100.0
English Primary Language	No	1	6.7
	Yes	14	93.3

	N	Range	M	SD
Age in Months	15	41 - 63	55.27	7.65

Table 9

Student Demographics - Classroom 9

Student Demographics	Variable	*N*	%
Gender	Male	4	28.6
	Female	10	71.4
Ethnicity	Black	3	21.4
	Asian	0	0.0
	Hispanic	10	71.4
	White	0	0.0
	Other	1	7.1
Disability Status	No	13	92.9
	Yes	1	7.1
Free Lunch	No	0	0.0
	Yes	14	100.0
English Primary Language	No	1	7.1
	Yes	13	92.9

	N	Range	M	SD
Age in Months	14	52 - 62	56.86	3.13

Table 10

Student Demographics - Classroom 10

Student Demographics	Variable	*N*	%
Gender	Male	5	55.6
	Female	4	44.4
Ethnicity	Black	1	11.1
	Asian	0	0.0
	Hispanic	0	0.0
	White	6	66.7
	Other	2	22.2
Disability Status	No	0	0.0
	Yes	9	100.0
Free Lunch	No	9	100.0
	Yes	0	0.0
English Primary Language	No	4	44.4
	Yes	5	55.6

	N	Range	M	SD
Age in Months	9	46 - 64	55.33	6.33

Table 11

Student Demographics - Classroom 11

Student Demographics	Variable	*N*	%
Gender	Male	4	50.0
	Female	4	50.0
Ethnicity	Black	0	0.0
	Asian	0	0.0
	Hispanic	0	0.0
	White	6	75.0
	Other	2	25.0
Disability Status	No	7	87.5
	Yes	1	12.5
Free Lunch	No	8	100.0
	Yes	0	0.0
English Primary Language	No	3	37.5
	Yes	5	62.5

	N	Range	M	SD
Age in Months	8	35 - 48	42.88	4.61

Appendix D

PARENT/GUARDIAN DEMOGRAPHIC QUESTIONNAIRE

We would like to ask some questions about you and your child to help interpret the results of the study. Please place a check mark in the appropriate box or write a response on the provided line.

Date Completed:

_____/_____/_____

Month Day Year

1. What is your relationship to the child?

□ Mother
□ Father
□ Other, specify _____

2. Ethnicity

□ Asian, Asian-American, Pacific Islander
□ Black, African-American, Non-Hispanic
□ Hispanic, Hispanic-American, Latino
□ American Indian, Alaskan Native
□ White, Caucasian, Non-Hispanic
□ Other, specify _____

3. How many years of school did the <u>mother</u> complete?

□ Does not apply
□ 6 or less
□ 7
□ 8
□ 9
□ 10
□ 11
□ 12
□ 13
□ 14
□ 15
□ 16
□ 17
□ 18
□ 19
□ 20 or more

4. How many years of school did the <u>father</u> complete?

□ Does not apply
□ 6 or less
□ 7
□ 8
□ 9
□ 10
□ 11
□ 12
□ 13
□ 14
□ 15
□ 16
□ 17
□ 18
□ 19
□ 20 or more

5. Does your child receive free or reduced lunch at school? If yes, which?

_____

6. What languages are spoken at home?

_____

7. Does your child have a documented disability?

_____

8. How many years has your child been in out of home care?

_____

Appendix E

TEACHER DEMOGRAPHIC QUESTIONNAIRE

We would like to ask some questions about you to help interpret the results of the study. Please place a check mark in the appropriate box or write a response on the provided line.

Date Completed:

_____/_____/_____

Month Day Year

1. Ethnicity

□ Asian, Asian-American, Pacific Islander
□ Black, African-American, Non-Hispanic
□ Hispanic, Hispanic-American, Latino
□ American Indian, Alaskan Native
□ White, Caucasian, Non-Hispanic
□ Other, specify _____

2. How many years of school did you complete?

□ Does not apply
□ 6 or less
□ 7
□ 8
□ 9
□ 10
□ 11
□ 12
□ 13
□ 14
□ 15
□ 16
□ 17
□ 18
□ 19
□ 20 or more

3. Received Degrees

□ G.E.D.
□ High school degree
□ Child care credential (e.g., NC, CDA)
□ Early Childhood Certificate
□ Early Childhood Diploma
□ Some college
□ Associate's Degree
□ Bachelor's Degree
□ Some graduate coursework
□ Master's degree
□ Post-Master's degree or coursework

4. How many years have you worked as a teacher?

_____

Appendix F

Teacher Introductory Letter

Date:

Dear Teacher,

My name is Sasha Collins Blackwell and I am a doctoral student in the Ph.D. Program in Educational Psychology at The Graduate Center of the City University of New York. I am conducting a study to determine if children's characteristics are expressed the differently in classroom environments that have different characteristics at your child's school. As part of the study, we will ask you to complete a demographic questionnaire, and two rating scales for every child in your class. In addition, independent observers (my research assistant and I) will do observational ratings of your classroom environment. As compensation for your time and effort, I will provide you with 15 dollars for every student in your classroom for whom you fill out rating scales to be allocated towards your classroom budget. Enclosed is a consent form that asks for your permission to participate in this study.

If you agree to participate in this study, please return the signed consent form to my research assistant or to the school office. If you have any questions about this research, you can contact me at (917) 226-4430 or scollins1@gc.cuny.edu, or you can contact my advisor Professor Georgiana Shick Tryon at (212) 817-8293 or gtryon@gc.cuny.edu. If you have questions about your rights as a participant in this study, you can contact Kay Powell, IRB Administrator, The Graduate Center/City University of New York, (212) 817-7525, kpowell@gc.cuny.edu.

Thank you very much for your time and consideration,

Sasha Collins Blackwell, M.A., M.Phil.

Appendix G

TEACHER CONSENT FORM

My name is Sasha Collins Blackwell and I am a doctoral student in the Ph.D. Program in Educational Psychology at The Graduate Center of the City University of New York (CUNY), and Principal Investigator of a study that will examine if children's characteristics are expressed the differently in classroom environments that have different characteristics. I would like your permission to participate in this study. If you give permission, you will be asked to complete a demographic questionnaire and two rating scales and independent observers will complete ratings of your classroom.

Your will be asked to complete two rating scales for every child in your classroom. It should take approximately 30 minutes to complete the two ratings scales for every child in your class. The first rating scale that you will complete for every child assesses child behavior and adjustment. The second rating scale that you will complete measures how children typically react in their environment. In addition, either research assistants or I will complete simultaneous observational ratings of your classroom. The first rating scale assess general aspects of the classroom environment, like how the program is structured, what types of activities the children are involved in, and what types of interactions occur between teachers and children. The second rating scale measures relational aspects of the classroom environment like teacher warmth and sensitivity. Your name or the child's name will not appear on any of the information gathered and all results will be kept strictly confidential. You and a child's parent can stop participation at any time without any penalties of any kind. As compensation for participation, I will provide you with 15 dollars for every student in your classroom for whom you filled out rating scales to be allocated towards your classroom budget.

The risk involved in this study is minimal, no more than might be encountered in everyday life. The benefits of your participation in this study will be in the contribution made to the understanding of how children's characteristics are expressed differently in different classroom environments. There will be approximately 15 classes consisting of about 20 children each and 15 head teachers that will be taking part in this study. Your name will not appear on any of the information gathered and all results will be kept strictly confidential and all data will be kept in a locked file cabinet. You also can stop participation at any time without any penalties of any kind.

I may publish results of the study, but names of people, or any identifying characteristics, will not be used in any of the publications. If you would like a copy of the study, please provide me with your address and I will send you a copy in the future.

Thank you for your time and consideration. If you agree to participate in this study, please sign and date below. If you have any questions about this research, you can contact me at (917) 226-4430 or scollins1@gc.cuny.edu, or you can contact my advisor Professor Georgiana Shick Tryon at (212) 817-8293 or gtryon@gc.cuny.edu. If you have questions about your rights as a participant in this study, you can contact Kay Powell,

IRB Administrator, The Graduate Center/City University of New York, (212) 817-7525, kpowell@gc.cuny.edu.

CONSENT

This project has been explained to me and I have had an opportunity to ask any questions that I may have and those questions have been answered to my satisfaction. I understand that my participation in this study is totally voluntary and that I may withdraw from this study at any time with no repercussions. I have read this form and I understand this project. I give consent for my participation in this study.

_____ _____

(Signature of Participant) (Date)

_____ _____

(Signature of Principal Investigator) (Date)

Appendix H

Parent Introductory Letter

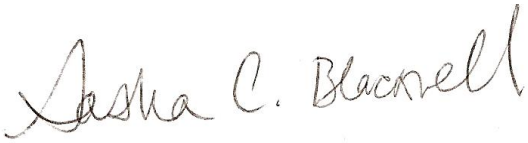
Date:

Dear Parent,

My name is Sasha Collins Blackwell and I am a doctoral student in the Ph.D. Program in Educational Psychology at The Graduate Center of the City University of New York. I am conducting a study to determine if children's characteristics are expressed differently in classroom environments that have different characteristics at your child's school. The study will involve your child's teacher completing two rating scales for every child in the class and independent observers (my research assistants and/or myself) completing observational ratings of your child's classroom environment. In addition, you will be asked to complete a demographic questionnaire that should take you about 5 minutes. Enclosed is a consent form that asks for your permission for your child to participate in this study.

If you give permission for your child to participate in this study, please return the signed consent form to your child's teacher or to the school office. You can also return it in the stamped, self-addressed envelope I provided. If you have any questions about this research, you can contact me at (917) 226-4430 or scollins1@gc.cuny.edu, or you can contact my advisor Professor Georgiana Shick Tryon at (212) 817-8293 or gtryon@gc.cuny.edu. If you have questions about your rights as a participant in this study, you can contact Kay Powell, IRB Administrator, The Graduate Center/City University of New York, (212) 817-7525, kpowell@gc.cuny.edu.

Thank you very much for your time and consideration,



Sasha Collins Blackwell, M.A., M.Phil.

Appendix I

PARENT/GUARDIAN CONSENT FORM

My name is Sasha Collins Blackwell and I am a doctoral student in the Ph.D. Program in Educational Psychology at The Graduate Center of the City University of New York (CUNY), and Principal Investigator of a study that will examine if children's characteristics are expressed differently in classroom environments that have different characteristics. I would like permission for your child to participate in this study. If you give permission, you will be asked to complete a demographic questionnaire, your child's teacher will be asked to complete a demographic questionnaire and two rating scales, and independent observers will complete ratings of your child's classroom environment.

Your child's head teacher will be asked to complete two rating scales. The first rating scale assesses child behavior and adjustment. The second rating scale is a measure of how your child typically reacts in his/her environment. In addition, either my research assistants or I will complete simultaneous observational ratings of your child's classroom using two rating scales. The first rating scale assess general aspects of the classroom environment, like how the program is structured, what types of activities the children are involved in, and what types of interactions occur between teachers and children. The second rating scale measures relational aspects of the classroom environment like teacher warmth and sensitivity. Your child's name will not appear on any of the information gathered and all results will be kept strictly confidential and all data will be kept in a locked file cabinet. You and your child's teacher can stop participation at any time without any penalties of any kind.

The risk involved in this study is minimal, no more than might be encountered in everyday life. The benefits of your child's participation in this study will be in the contribution made to the understanding of how children's characteristics are expressed differently in different classroom environments. There will be approximately 15 classes consisting of 20 children each and 15 head teachers that will be taking part in this study.

I may publish results of the study, but names of people, or any identifying characteristics, will not be used in any of the publications. If you would like a copy of the study, please provide me with your address and I will send you a copy in the future.

Thank you for your time and consideration. If you agree to allow your child to participate in this study, please sign and date below. If you have any questions about this research, you can contact me at (917) 226-4430 or scollins1@gc.cuny.edu, or you can contact my advisor Professor Georgiana Shick Tryon at (212) 817-8293 or gtryon@gc.cuny.edu. If you have questions about your rights as a participant in this study, you can contact Kay Powell, IRB Administrator, The Graduate Center/City University of New York, (212) 817-7525, kpowell@gc.cuny.edu.

CONSENT

This project has been explained to me and I have had an opportunity to ask any questions that I may have and those questions have been answered to my satisfaction. I understand that my child's participation in this study is totally voluntary and that we may withdraw from this study at any time with no repercussions. I have read this form and I understand this project. I give consent for my child to participate in this study.

_____ I agree to participate and to allow my child _____ to participate.

(Child's Name)

_____ _____

(Signature of Parent or Guardian) (Date)

_____ _____

(Signature of Principal Investigator) (Date)

Appendix J

Parent Letter to Explain Demographic Questionnaire

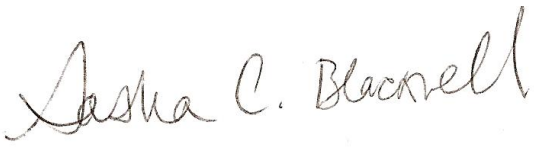
Date:

Dear Parent,

Hello again! Enclosed is a demographic questionnaire for you to complete. This information will help me interpret the results of the study. Please do not write your name or your child's name on this form.

Please return the completed forms to your child's teacher or school office or in the stamped, self-addressed envelope I provided. If you have any questions about this research, you can contact me at (917) 226-4430 or scollins1@gc.cuny.edu, or you can contact my advisor Professor Georgiana Shick Tryon at (212) 817-8293 or gtryon@gc.cuny.edu. If you have questions about your rights as a participant in this study, you can contact Kay Powell, IRB Administrator, The Graduate Center/City University of New York, (212) 817-7525, kpowell@gc.cuny.edu.

Thank you very much for your time and consideration,



Sasha Collins Blackwell, M.A., M.Phil.

Appendix K

Teacher Packet Letter

Date:

Dear Teacher,

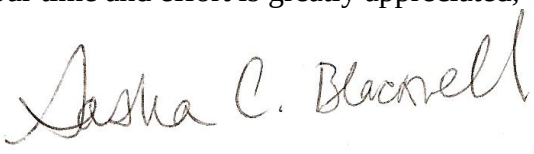
Thank you very much for returning the consent form. This packet contains a demographic questionnaire for you to fill-out. It also contains two types of rating scales for you to complete for every child in your classroom. Every child in your class has an identification number in this study. Since no names or identifying information can be used on any forms, I have prelabeled the forms with the children's identification numbers. Therefore, the packet also includes a list of the children's names and their corresponding identification numbers to enable you to complete the rating scales.

The first type of rating scale is one with which you are already familiar. The Preschool and Kindergarten Behavior Scales (PKBS-2) measures child behavior and social skills. Please complete one PKBS-2 for every student in your class. Section I will already be completed for you. In Section II, just write in the date you fill-out the scale. Section III provides you with instructions on how to fill out the scale. These instructions ask you to rate the each on each of the items on pages 2 and 3 of the rating form. It is not necessary for you to provide additional information in Section IV on page 4.

The second type of rating scale, the Teacher and Caregiver Temperament Inventory for Children (TACTIC) measures how a child typically responds in their environment. Please complete one TACTIC for every child in your class. On page 1, just fill in the date you completed the rating and the length of time you have observed the child. The first page of the scale also provides you with instructions on how to fill-out the rating scale. Please complete all items on the scale.

Please return the completed packet to my research assistant or to the school office. If you have any questions about this research, you can contact me at (917) 226-4430 or scollins1@gc.cuny.edu, or you can contact my advisor Professor Georgiana Shick Tryon at (212) 817-8293 or gtryon@gc.cuny.edu. If you have questions about your rights as a participant in this study, you can contact Kay Powell, IRB Administrator, The Graduate Center/City University of New York, (212) 817-7525, kpowell@gc.cuny.edu.

Your time and effort is greatly appreciated,



Sasha Collins Blackwell, M.A., M.Phil.

Appendix L

Training Procedures for ECERS-R and CIS

1. Two research assistants were familiarized with the ECERS-R and CIS by reading the rating scales and their accompanying documents. The CIS is supported by Arnett's (1989) study, from which it was formulated as well as a document that details item-level interpretation/analysis and scale scoring procedures (Smart Start Evaluation Team, 1999). The ECERS-R is supported by a comprehensive manual and a text, as well as by a resource guide entitled, *All About the ECERS-R* (Cryer, Harms, & Riley, 2003), which provides detailed descriptions of how to complete the scale along with indicator-specific photographs.
2. The investigator reviewed the points highlighted in supporting documents in a training session with the research assistants. During this training session, the investigator discussed how to complete the rating scales at the item level, how to score the rating scales, and how these scores are interpreted.
3. The investigator accompanied the two research assistants to a trial observation of both instruments in a prekindergarten classroom. During this trial observation, the investigator and the two research assistants completed the CIS in entirety and the Activities, Interaction, and Program Structure scales of the ECERS-R.
4. After the trial observation period, the investigator reviewed scoring techniques for both instruments with the research assistants.
5. The proportion of inter-rater agreement was calculated between the 3 raters. Exact percent agreement and agreement within 1 point was calculated. If the proportions of agreement within 1 point were below 85%, steps 1 through 4 were repeated until this

criterion was ascertained (this was however, not the case, as initial agreement was greater than 85%).

Appendix M

Table 1

Correlations between TACTIC Category Scores and PKBS-2 (*N* = 130)

	TACTIC Activity Level	TACTIC Adaptability	TACTIC Approach	TACTIC Intensity	TACTIC Mood	TACTIC Persistence	TACTIC Distractibility	TACTIC Rhythmicity	TACTIC Sensory Threshold	TACTIC Total	PKBS-2 Externalizing Behavior SS	PKBS-2 Social Skills SS
TACTIC Activity Level	1											
TACTIC Adaptability	.81**	1										
TACTIC Approach	.26**	.28**	1									
TACTIC Intensity	.77**	.74**	.17	1								
TACTIC Mood	.61**	.66**	.42**	.78**	1							
TACTIC Persistence	.70**	.54**	.31**	.50**	.40**	1						
TACTIC Distractibility	.66**	.48**	.30**	.48**	.38**	.67**	1					
TACTIC Rhythmicity	.51**	.48**	.20**	.55**	.49**	.39**	.30**	1				
TACTIC Sensory Threshold	.57**	.55**	.27**	.46**	.31**	.48**	.32**	.38**	1			
TACTIC Total	.88**	.82**	.53**	.80**	.75**	.76**	.68**	.64**	.62**	1		
PKBS-2 Externalizing Behavior SS	.52**	.51**	.26**	.63**	.56**	.19*	.29**	.34**	.26**	.54**	1	
PKBS-2 Social Skills SS	-.33**	-.39**	-.53**	-.18*	-.37**	-.27**	-.28**	-.10	-.19*	-.43**	-.45**	1

*$p < .05$; **$p < .0$

Appendix N

HLM Model Summaries

Table 1

Externalizing Behavior

	Model				
	A	B	C	D	E
ICC	18.44%	12.84%	19.41%	24.29%	27.36%
R1		2.19%	32.57%	32.44%	31.74%
R2		36.26%	28.18%	4.15%	-13.72%

Table 2

Social Skills

	Model				
	A	B	C	D	E
ICC	6.05%	0.64%	16.89%	7.75%	12.14%
R1		8.08%	34.59%	35.41%	35.79%
R2		90.86%	-106.30%	15.77%	-37.75%

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist and 1991 Profile*. Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M. (1997). *Guide for the Caregiver-Teacher Report Form Ages 2-5.* Burlington, VT: University of Vermont Department of Psychiatry.

Achenbach, T. M., & Edelbrock, C. (1983). Manual for the child behavior checklist and revised child behavior profile. Burlington, VT: Queen City Printers.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10,* 541-552.

Bates, J. E. (1989). Concepts and measures of temperament. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 1-26). Chichester, England: John Wiley & Sons, Ltd.

Bates, J. E., Freeland, C. A. B., & Loundsbury, M .L. (1979). Measurement of infant difficultness. *Child Development, 50,* 794-803.

Behar, L. B., & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology, 10,* 601-610.

Bergen, J. R., Feld, J. K., Reddy, L. A., Li, F. F., Schwartz, R. D., & Cheng, Y. H. (1992). *MAPS developmental observational guide: Level KL.* Tucson, AZ: ATI.

Billman, J. (1981). The preschool temperament inventory: Construction and standardization of a teacher-rated instrument for assessing temperament of three to six year-old children. (ERIC Document Reproduction Service No. ED224592) Retrieved July 30, 2008, from ERIC database: http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/2f/f4/bd.pdf

Billman, J., & McDevitt, S. C. (1998, October). *TACTIC: A measure of temperament, attention, conduct, and emotion for 2-6 year old children in out-of-home settings.* Paper presented at the twelfth Occasional Temperament Conference, Philadelphia, PA.

Blackwell, S. C. (2008, October). *An examination of the goodness of fit model: Is the relationship between temperament and behavior expressed differently in dissimilar classroom environments?*. Poster session presented at the biannual Occasional Temperament Conference, San Rafael, CA.

Braun, H., Jenkins, F., & Grigg, W. (2006). *A closer look at charter schools using hierarchical linear modeling*. (NCES 2006-460). U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences. Washington, DC: U.S. Government Printing Office.

Bredekamp, S., & Copple, C. (1997). *Developmentally appropriate practices in early childhood programs.* Washington, D.C.: NAEYC.

Bronfenbrenner, U., & Morris, P. A. (1997). The ecology of developmental processes. In W. Damon (Ed.), *Handbook of child psychology* (5th ed., pp. 993-1028). New York: Wiley.

Bryant, D., Burchinal, M., Lau, L., & Sparling, J. (1994). Family and classroom correlates of head start children's developmental outcomes. *Early Childhood Research Quarterly, 9,* 289-309.

Burchinal, M. R., Cryer, D., Clifford, R. M., & Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science, 6*, 2-11.

Burchinal, M. R., Roberts, J. E., Riggins, R., Zeisel, S.A., Neebe, E., & Bryan, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development, 71,* 339-357.

Buss, A. (1989). Temperament as personality traits. In G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp.49-58). West Sussex, England: John Wiley and Sons.

Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development.* New York: John Wiley.

Buss, A. H., & Plomin, R. (1984). Revised temperament survey. In *Temperament: Early developing personality traits.* Mahwah, NJ: Lawrence Erlbaum Associates.

Carey, W. B. (1998). Temperament and behavior problems in the classroom. *School Psychology Review, 27,* 522-533.

Carey, W. B. (2000). *The Carey Temperament Scales.* Scottsdale, AZ: Behavioral-Developmental Initiatives.

Carey, W. B., & McDevitt, S. C. (1978). Revision of the Infant Temperament Questionnaire. *Pediatrics, 61,* 735-739.

Carson, D. K., & Bittner, M. T. (1994). Temperament and school-age children's coping abilities and responses to stress. *The Journal of Genetic Psychology, 155,* 289-302.

Caruso, D A., Dunn, L., Hsu, H., & Hestenes, L. (1991, November). *Preschool observation of classroom experience: Measuring the quality of children's experience in early childhood programs.* Paper presented at the National Association for the Education of Young children, Denver.

Caspi, A., & Silva, P. A. (1995). Temperament qualities at age three predict personality traits in young adulthood: Longitudinal evidence form a birth cohort. *Child Development, 66,* 486-498.

Cassidy, D. J., Hestenes, L. L., Hansen, J. K., Hegde, A., Shim, J., & Hestenes, S. (2005). Revisiting the two faces of child care quality: Structure and process. *Early Education and Development, 16,* 505-520.

Cassidy, D. J., Pugh-Hoese, S., & Buell, M. J. (1995). The effect of education on childcare teachers' beliefs and classroom quality: Year one evaluation of the T.E.A.C.H. early childhood associate degree scholarship program. *Early Childhood Research Quarterly, 10,* 171-183.

Chandler, L. A. (1985). *Assessing stress in children.* New York: Praeger.

Chandler, L. A. (1986). *The Stress Response Scale for Children: Revision.* Pittsburg: University of Pittsburg, The Psychoeducational Clinic.

Chess, S., & Korn, S. (1970). Temperament and behavior disorders in mentally retarded children. *Archives of General Psychiatry*, *23,* 122-130.

Chess, S., Korn, S., & Fernandez, P.B. (1971). *Psychiatric disorders of children with congenital rubella.* New York: Brunner/Mazel.

Chess, S., & Thomas, A. (1991). Temperament. In M. Lewis (Ed.), *Child and adolescent psychiatry: A comprehensive textbook (pp. 145-159).* Baltimore: Williams & Wilkins.

Chess, S., & Thomas, A. (1996). *Temperament: Theory and practice.* New York: Brunner/Mazel.

Chess, S., Thomas, A, & Cameron, M. (1976). Temperament: Its significance for early schooling. *New York University Education Quarterly, 7,* 24-29.

Churchill, S. L. (2003). Goodness of fit in early childhood settings. *Early Childhood Education Journal, 31,* 113-118.

Conners, C. K. (1990). *Conners' Rating Scales Manual.* North Tonawanda, NY: Multi-Health Systems.

Cooper, D. H., & Speece, D. L. (1990). Maintaining at-risk children in regular education settings: Initial effects of individual differences and classroom environments. *Exceptional Children, 57,* 117-126.

Coopersmith, S. (1967). *The antecedents of self-esteem.* San Francisco, CA: Freeman.

Cryer, D., Harms, T. & Riley, C. (2003). *All about the ECERS-R: A detailed guide in words and pictures to be used with the ECERS-R* . Winston-Salem, NC: Kaplan Early Learning, PACT Group.

De Schipper, J. C., Taveecchio, L. W., Van IJzendorrn, M. H., & Jantine, V. Z. (2004). Goodness of fit in center day care: Relations of temperament, stability, and quality of care with the child's adjustment. *Early Childhood Research Quarterly, 19,* 257-272.

De Schipper, J. C.,Van IJzendoorn, M. H., & Tavechio, L. W. C. (2004). Stability in center day care: relations with children's well-being and problem behavior in day care. *Social Development, 13,* 530-551.

De Kruif, R. E. L., McWilliam, R.A., Ridley, S. M., & Wakely, M. B. (2000). Classification of teachers' interaction behaviors in early childhood classrooms. *Early Childhood Research Quarterly, 15,* 247-268.

Doherty-Derkowski, G. (1995). *Quality matters*. Don Mills, Ontario: Addision-Wesley.

Dominguez, X., Vitiello, V.E., Maier, M.F., & Greenfield, D.B. (2010). A longitudinal examination of young children's learning behavior: Child-level and classroom – level predictors of change throughout the preschool year. *School Psychology Review, 29,* 29-47.

Downey, D. B., von Hippel, P. T., & Broh, B. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review, 69,* 613-635.

Dunn, J., & Kendrick, C. (1982). *Siblings: Love, envy and understanding*. Cambridge, M.A: Harvard University Press.

Early, D. M., Bryant, D., Pianta, R., Clifford, R., Burchinal, M., Ritchie, S., et al. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in prekindergarten? *Early Childhood Research Quarterly, 21,* 174-195.

Early, D. M., Maxwell, K. L., Burchinal, M., Bender, R. H., Ebanks, C., Henry, G. T., et al. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*, 558-580.

Epstein, N. B., Bishop, D. S., & Levin, S. (1978). The McMaster model of family functioning. *Journal of Marriage and Family Counseling, 40*, 585-593.

Espinosa, L. M. (2002). High-quality preschool: Why we need it and what it looks like. *Preschool Policy Matters, 1,* 1-11.

Essa, E. L., & Burnham, M. (2001). Child care quality: A model for examining relevant variables. In S. Reifel & M. H. Brown (Eds.), *Early education and care, and reconceptualizing play* (Vol. 11, pp. 59-113). Oxford, England: Elsevier Science Ltd.

Farver, J. & Branstetter, W., (1994). Preschoolers' prosocial responses to their peers' distress. *Developmental Psychology 30,* 334–341.

Fisher, R. A. (1958). *Statistical Methods for Research Workers* (pp. 1-13). Edinburgh, England: Oliver and Boyd.

Fussell, J.L., Macias, M.M., & Saylor, C.F., (2005). Social skills and behavior problems in children with disabilities with and without siblings. *Child psychiatry and human development, 36*, 227-241.

Garrison, W., Earls, F., & Kindlon, D. (1984). Temperament characteristics in the third year of life and behavioral adjustment at school entry. *Journal of Clinical Child Psychology, 13,* 298-303.

Gerber, E. B., Whitebook, M., & Weinstein, R. S. (2007). At the heart of child care: Predictors of teacher sensitivity in center-based child care. *Early Childhood Research Quarterly, 22,* 327-346.

Goldsmith, H. H., Buss, A. H., Plomin, R., Rothbart, M. K., Thomas, A., Chess, S., Hinde, R. A., & McCall, R. B. (1987). Roundtable: What is temperament? Four approaches. *Child Development, 58*, 505-529.

Graham, P., Rutter, M., & George, S. (1973). Temperamental characteristics as predictors of behavior disorders in children. *American Journal of Orthopsychiatry, 43,* 328-339.

Gresham, F. M., & Elliot, S. N. (1990). *Social Skills Rating System.* Circle Pines, MN: American Guidance Service.

Ghazvini, A., & Mullis, R. (2002). Center-based care for young children: examining predictors of quality. *The Journal of Genetic Psychology*, *163,* 112-125

Gray, J. A., (1987). *The psychology of fear and stress*. Cambridge: Cambridge University Press.

Guerin, D. W., Gottfried, A. W., Oliver, P. H., & Thomas, C. W. (1994). Temperament and school functioning during early adolescence. *Journal of Early Adolescence, 14,* 200-225.

Guerin, D. W., Gottfried, A. W., & Thomas, C. W. (1997). Difficult temperament and behavior problems: A longitudinal study from 1.5 to 12 years. *International Journal of Behavior Development, 21,* 71-90.

Guralnick, M.J., Connor, R.T., Hammond, M.A., Gottman, J.M., & Kinnish, K. (1996a). Immediate effects of mainstreamed settings on the social interactions and social integration of preschool children. *American Journal on Mental Retardation, 100,* 359-377.

Guralnick, M.J., Connor, R.T., Hammond, M.A., Gottman, J.M., & Kinnish, K. (1996b). The peer relations of preschool children with communication disorders. *Child Development, 67,* 471-489.

Guralnick, M.J., Gottman, J.M., & Hammond, M.A. (1995). Effects of social setting on the friendship formation of young children differing in developmental status. *Journal of Applied Developmental Psychology, 17,* 625-651.

Guralnick, M.J., & Groom, J.M. (1988). Friendships of preschool children in mainstreamed playgroups. *Developmental Psychology, 24,* 595-604.

Guralnick, M.J., & Weinhouse, E.M., (1984). Peer-related social interactions of developmentally delayed young children: Development and characteristics. *Developmental Psychology, 20*, 815-827.

Hagekull, B. (1994). Infant temperament and early childhood functioning: Possible relations to the five-factor model. In C. J. Halverson, Jr., G. A. Kohnstamm, & R. P. Martin (Eds.), *The developing structure of temperament and personality* (pp. 227-240). Hillsdale, NJ: Erlbaum.

Hagekull, B., & Bohlin, G. (1981). Individual stability in dimensions of infant behavior. *Infant Behavior and Development, 4*, 97-108.

Hagekull, B., & Bohlin, G. (1995). Childcare quality, family and child characteristics and socioemotional development. *Early Childhood Research Quarterly, 10*, 505-526.

Harms, T., & Clifford, R. M. (1980). *The Early Childhood Environment Rating Scale.* New York: Teachers College Press.

Harms, T., Clifford, R. M., & Cryer, D. (2005). *Early Childhood Environment Rating Scale- Revised edition*. New York: Teachers College Press.

Head Start Quality Research Consortium (2003). *Head Start FACES: A whole-child perspective on program performance.* Washington, DC: U.S. Department of Health and Human Services.

Hegvik, R. L., McDevitt, S. C., & Carey, W. B. (1982). *The Middle Childhood Temperament Questionnair*e. Philadelphia: Temple University College of Education.

Hegvik, R. L., McDevitt, S. C., & Carey, W. B. (1982). The Middle Childhood Temperament Questionnaire. *Journal of Developmental and Behavioral Pediatrics, 3*, 197-200.

Helburn, S. W. (Ed.) (1995). *Cost, quality, and child outcomes in child care centers. Technical report.* Denver: University of Denver, Department of Economics, Center for Research in Economic and Social Policy.

Hestenes, L. L., Kontos, S., & Bryan, S. (1993). Children's emotional expression in child care. *Early Childhood Research Quarterly, 8,* 167-192.

Howes, C. (1997). Teacher sensitivity, children's attachment and play with peers. *Early Education and Development, 8,* 41-49.

Howes, C. (1987). Social competency with peers: Contributions from child care. *Early Childhood Research Quarterly, 2,* 155-167.

Howes, C., & Hamilton, C. E. (1993). The changing experience of child care: Changes in teachers' and in teacher-child relationships and children's social competence with peers. *Early Childhood Research Quarterly, 8,* 15-32.

Howes, C., Hamilton, C.E., & Matheson, C.C. (1994). Children's relationships with peers: Differential association with aspects of the teacher-child relationship. *Child Development, 65,* 253-263.

Howes, C., & Stewart, P. (1987). Child's play with adults, toys, and peers: An examination of family and child-care influences. *Developmental Psychology, 23,* 423-430.

Howes, C., Whitebook, M., & Phillips, D. (1992). Teacher characteristics and effective teaching in child care: Findings from the National Child Care Staffing Study. *Child & Youth Care Forum. Special issues: Meeting the child care needs for the 1990s: Perspectives on day care: II, 21,* 399-414.

Hubert, N. C., Wachs, T. D., Peters-Martin, P., & Gandour, M. J. (1982). The study of early temperament: Measurement and conceptual issues. *Child Development, 53,* 571-600.

Hutton, J. B., & Roberts, T. G. (1990). *Stress Impact Scale.* Austin, TX: Pro-Ed.

Jewsuwan, R., Luster, T., & Kostelink, M. (1993). The relation between parents' perceptions of temperament and children's adjustment to preschool. *Early Childhood Research Quarterly, 8,* 33-51

Keogh, B. K. (1982). Children's temperament and teachers' decisions. In R. Porter & G. M. Collins (Eds.), *Temperament differences in infants and young children* (pp. 269-285). London: Pitman Books Ltd (CIBA Foundation Symposium 89).

Keogh, B. K. (1983). Individual differences in temperament: A contribution to the personal, social, and educational competence of learning disabled children. In J. D. McKinney & L. Feagens (Eds.), *Current topics in learning disabilities* (pp. 34-45). Norwood, NJ: Ablex.

Keogh, B. K. (1986). Temperament and schooling: Meaning of goodness of fit. *New Directions for Child Development, 31*, 89-108.

Keogh, B. K. (1989). Applying temperament research to school. In G. A. Kohnstamm, J. E., Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 437-450). Chichester, England: John Wiley & Sons.

Keogh, B. K. (1994). Temperament and teachers' views of teachability. In W. B. Carey & S. C. McDevitt (Eds.), *Prevention and early intervention. Individual differences as risk factors for the mental health of children* (pp. 246-256). New York: Brunner/Mazel.

Keogh, B. K. (2003). *Temperament in the classroom: Understanding individual differences.* Baltimore, MD: Paul H Brookes.

Keogh, B. K., Pullis, M. E., & Cadwell, J. (1982). A short form of the Teacher Temperament Questionnaire. *Journal of Educational Measurement, 19*, 323-329.

Keogh, B. K., & Speece, D. L. (1996). Learning disabilities within the context of school. In D. L. Speece & B. K. Keogh (Eds.), *Research on classroom ecologies: Implications for inclusion of children with learning disabilities* (pp.1-14). Mahwah, N.J: Lawrence Erlbaum Associates.

Klein, K. A. (1982). The relationship between children's temperament and adjustment to kindergarten and Head Start settings. *The Journal of Psychology, 112*, 259-268.

Klein, K. A., & Ballantine, J. H. (2001). The relationship of temperament to adjustment in British infant schools. *The Journal of Social Psychology, 128*, 585-595.

Kontos, S., & Feine, R. (1987). Child care quality, compliance with regulations, and children's development: The Pennsylvania study. In D. Phillips (Ed.), *Quality in child care: What does the relationship tell us?* (pp. 57-79). Washington, DC: National Association for the Education of Young Children.

Kontos, S., Howes, C., & Galinsky, E. (1996). Does training make a difference to quality in family child care? *Early Childhood Research Quarterly, 11*, 427-445.

Kontos, S., & Wilcox-Herzog, A. (1997). Influences on the competence of children's play with objects and peers in early childhood classrooms. *Early Childhood Research Quarterly, 12*, 247-262.

Kornblau, B. W. (1982). The Teachable Pupil Survey: A technique for assessing teachers' perceptions of pupil attributes. *Psychology in the Schools, 19,* 170-174.

Koroloff, N. (1989, February). *Getting a great start: Early intervention demonstration projects (interim report).*

Koroloff, N. (1990, March). *Getting a great start: Part II. Early intervention demonstration projects- January 1998 to June 1989.*

Koroloff, N. (1991, March). *Starting right: An interim report on early identification and prevention services for children at risk of mental or emotional disorders demonstration project- July 1989 to December 1990.*

Koroloff, N. (1992). *Starting right: Part II. A final report on early identification and prevention services for children at risk of mental or emotional disorders demonstration project- July 1989 to June 1991.*

Kreft, I., de Leeuw, J., & Aiken, L. S. (1995). The effects of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, *30*, 1-21.

Kristal, J. (2005). *The temperament perspective: Working with children's behavioral styles.* New York: Paul H Brookes.

Lamb, M. E., Hwang, C., Broberg, A., & Bookstein, F. (1988). The effects of out-of-home care on the development of social competence in Sweden: A longitudinal study. *Early Childhood Research Quarterly, 3*, 379-402.

Lambert, R., Abbott-Shim, M., & McCarty, F. (2002). The relationship between classroom quality and ratings of the social functioning of Head Start children. *Early Childhood Development and Care, 172*, 231-245.

Layzer, J. L., Goodson, B. D., & Moss, M. (1993). *Life in preschool: Vol. 1. Final report to the U.S. Department of Education*. Cambridge, MA: Abt Associates.

Lerner, J. V. (1983). The role of temperament in psychosocial adaptation in early adolescents: A test of the "goodness of fit" model. *The Journal of Genetic Psychology, 143,* 149-157.

Lerner, R. M., & Lerner, J. V. (1977). The effects of age, sex, and physical attractiveness of child-peer relations, academic performance, and elementary school adjustment. *Developmental Psychology, 13*, 585-590.

Lerner, J. V., Lerner, R. M., Windle, M., Hooker, K., Lernerz, K., & East, P .L. (1986). Children and adolescents in their contexts: Tests of a goodness of fit model. In R. Plomin and J. Dunn (Eds.), *The study of temperament: Changes, continuities, and challenges* (pp. 99-114). Hillsdale: NJ: Lawrence Erlbaum Associates.

Lerner, J. V., Lerner, R. M., & Zabski, S. (1985). Temperament and elementary school children's actual and rated academic performance: A test of the "goodness of fit model." *Journal of Child Psychology and Psychiatry, 26*, 125-126.

Lerner, R. M., Palermo, M., Spiro, A., III, & Nesselroade, J.R. (1982). Assessing the dimensions of temperamental individuality across the life-span: The Dimensions of Temperament Survey (DOTS). *Child Development, 53*, 149-159.

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly, 22*, 3-17.

Mann, H. (1965). *Horace Mann on the crisis in education*. Yellow Springs, OH: Antioch Press.

Martin, R. P. (1983). Temperament: A review of research with implications for the school psychologist. *School Psychology Review, 12,* 266-273.

Martin, R. P. (1988). *The Temperament Assessment Battery for Children.* Brandon, VT: The Clinical Psychology Publishing Co.

Martin, R. P. (1996). *The Temperament Assessment Battery for Children-Revised.* Athens, GA: School Psychology Clinic.

Martin, R. P., Nagle, R., & Paget, K. (1983). Relationships between temperament and classroom behavior, teacher attitudes, and academic achievement. *Journal of Psychoeducational Assessment, 1,* 377-386.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79,* 732-749.

Matson, K. L., Rotari, A. F., & Helsel, W. J., (1985). Development of a rating scale to measure social skills in children: The Matson Evaluation of Social Skills with Youngsters (MESSY). *Psychopharmacology Bulletin, 21,* 855-867.

Maxwell, K. L., McWilliam, R. A., Hemmeter, M. L., Ault, M. J., Schuster, J. W. (2001). Predictors of developmentally appropriate practices in kindergarten through third grade. *Early Childhood Research Quarterly, 16,* 431-452.

Maziade, M. (1987). Should adverse temperament matter to the clinician? An empirically based answer. In G. A. Kohnstamm, J. E., Bates, & M. K. Rothbart (Eds.), *Temperament in childhood* (pp. 421-436). Chichester, England: John Wiley & Sons, Ltd.

Maziade, M. Caperaa, P., Laplante, B., Boudreault, M ., Thivierge, J., Cote, R., & Boutin, P. (1985). Value of difficult temperament among 7-year-olds in the general population for predicting psychiatric diagnosis at age 12. *American Journal of Psychiatry, 142*, 943-946.

Maziade, M., Caron, C., Cote, R., Mereete, C., Bernier, H., Laplante, B., & Thivierge, J. (1990). Psychiatric status of adolescents who had extreme temperaments at age seven. *American Journal of Psychiatry, 147*, 1531-1536.

McDermott, P. A., Marston, N.C., & Stott, D. H. (1993). *Adjustment Scales for Children and Adolescents.* Philadelphia, PA: Edumetric and Clinical Science.

McDevitt, S. C. (1994). Assessment of individual differences in the temperament of children: Evaluation of interactions. In W. B. Carey & S. C. McDevitt (Eds.), *Prevention and early intervention: Individual differences as risk factors for the mental health of children* (pp. 193-201). New York: Bruner Mazel.

McDevitt, S. C., & Carey, W. B. (1978). The measurement of temperament in 3-7 year old children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 19*, 245-253.

Merrell, K. W. (1993). *School Social Behavior Scales.* Eugene, OR: Assessment-Intervention Resources (http://www/assessment-intervention.com).

Merrell, K. W. (2002). *Preschool and Kindergarten Behavior Scales- Second Edition Examiner's Manual*. Austin, TX: Pro-Ed.

National Association for the Education of Young Children. (1997). *Developmentally appropriate practices in early childhood programs serving children from birth through age 8: A position statement of the National Association of the Education of Young Children.* Washington DC: Author.

National Association for the Education of Young Children. (2005). *NAEYC early childhood program standards and accreditation criteria: The mark of quality in early childhood education.* Washington DC: Author.

NICHD Early Child Care Research Network. (1998). Early child care and self-control, compliance, and problem behavior at twenty-four and thirty-six months. *Child Development, 69,* 1145-1170.

NICHD Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development, 74,* 976-1005.

NICHD Early Child Care Research Network. (2005). Early child care and children's development in the primary grades: Follow-up results from the NICHD study of early child care. *American Educational Research Journal, 43,* 537-570.

Newfield, J. (1980). Accuracy of teacher reports: Reports and observations of specific classroom behaviors. *Journal of Educational Research, 74,* 78-82.

Oldehinkel, A. J., Hartman, C. A., De Winter, A. F., Veenstra, R., & Ormel, J. (2004). Temperament profiles associated with internalizing and externalizing problems in preadolescence. *Development and Psychopathology, 16,* 421-440.

Paget, K. D., Nagle, R. J., & Martin, R. P. (1984). Interrelationships between temperament characteristics and first-grade teacher-student interactions. *Journal of Abnormal Child Psychology, 12,* 547-560.

Palermo, M. E. (1982). *Child temperament and contextual demands: A test of the goodness-of-fit model.* Unpublished doctoral dissertation, Pennsylvania State University, University Park.

Peisner-Feinberg, E. S., & Burchinal, M. R. (1997). Relations between preschool children's child care experiences and concurrent development: The Cost, Quality, and Outcomes Study. *Merrill-Palmer Quarterly, 43,* 451-477.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72,* 1534-1553.

Pence, A.R., & Goelman, H. (1991). The relationship of regulation, training, and motivation to quality of care in family day care. *Child and Youth Care Forum , 20,* 83-101.

Perlman, M., Zellman, G., & Le, V. (2004). Examining the psychometric properties of the Early Childhood Environment Rating Scale – Revised (ECERS-R). *Early Childhood Research Quarterly, 19,* 398-412.

Phillipsen, L., Burchinal, M., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly, 12 ,* 281-303.

Pianta, R. (1999). *Enhancing relationships between children and teachers*. Washington, DC: American Psychological Association.

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9,* 144-159.

Pianta, R., LaParo, K., & Hamre, B. (2007). *Classroom Scoring Assessment System (CLASS): K-3.* Baltimore, MD: Brookes Publishing Co.

Pianta, R., LaParo, K., Payne, C., Cox, M., & Bradley, R. (2002). Observed quality of the kindergarten classroom environment: Description and relations with teacher, family, school characteristics and child outcomes. *Elementary School Journal, 102,* 225-238.

Plomin, R., & Daniels, D. (1984). The interaction between temperament and environment: Methodological considerations. *Merrill-Palmer Quarterly, 30, 149-162.*

Pullis, M. E. (1989). Goodness of fit in classroom relationships. In W. B. Carey & S. C. McDevitt (Eds.), *Clinical and educational applications of temperament research* (pp. 117-120). Amsterdam/Lisse: Swets & Zeitlinger.

Pullis, M. E., & Cadwell, J. (1982). The influence of children's temperament on teachers' decision strategies. *American Educational Research Journal, 19,* 165-181.

Pullis, M. E., & Cadwell, J. (1985). Temperament as a factor in the assessment of children educationally at risk. *Journal of Special Education, 19,* 91-102.

Raudenbush, S. W. Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods, 2nd ed*. Newbury Park, CA: Sage.

Resnick, G., & Zill, N. (1999, April). *Is Head Start providing high quality education services? "Unpacking" classroom processes.* Paper presented at the Biannual Convention of SRCD. Albuquerque, NM.

Rice, J. K. (2006). Teacher quality: Understanding the effectiveness of teacher attributes. The Economic Policy Institute, Retrieved October 2, 2007, from http://www.epinet.org/content.cfm/books_teacher_quality_execsum_intro

Rice, M. S., & Gaines, S. K. (1992). Measurement of child temperament: Implications for researchers, clinicians, and caregivers. *Children's Health Care, 21*, 177-183.

Robinson, E. A., Eyberg, S. M., & Ross, A.W. (1980).The standardization of an inventory of child conduct problems. *Journal of Clinical Child Psychology, 9*, 22-29.

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development, 52*, 569-578.

Rothbart, M.K., & Bates, J.E. (1998). Temperament. In: W. Damon & N. Eisenberg N (Eds.), *Handbook of child psychology: Social, emotional, and personality development* (Vol. 3, pp. 105-176). New York: Wiley.

Rothbart, M. K., & Putnam, S. P. (2002). Temperament and socialization. In L. Pulkkinen & A. Caspi (Eds.), *Paths to successful development* (pp. 19-45). Cambridge: Cambridge University Press.

Sanson, A., Hemphill, S. A., & Smart, D. (2004). Connections between temperament and social development: A review. *Social Development, 13*, 142-170.

Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly, 9,* 131-151.

Shaefer, E. S., Edgerton, M., & Aaronson, M. (1978). *Classroom Behavior Inventory.* Unpublished manuscript, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Schwartz, I. S., Garfinkle, A. N., & Davis, C. (2002). Arranging preschool environments to facilitate valued social and educational outcomes. In M.R. Shinn, H.M. Walker, & G. Stoner (Eds.), *Interventions for Academic and Behavior Problems II: Preventative and Remedial Approaches* (pp. 455-468). Bethesda, MD: NASP Publications.

Seifer, R. (2000). Temperament and goodness of fit: Implications for developmental psychopathology. In. A. J. Sameroff, M. Lewis, and S. M. Miller (Eds.), *Handbook of Developmental Psychopathology, (2nd ed.,* pp. 257- 276). New York: Kluwer.

Slee, P. (1986). The relationship between *temperament and other factors to children's kindergarten adjustment*. *Child Psychiatry and Human Development, 17,* 104-112.

Smith, B. (1994). The temperament program: Community-based prevention of behavior disorders in children. In W. B. Carey & S. C. McDevitt (Eds.), *Prevention and early intervention: Individual differences as risk factors for the mental health of children* (pp. 257-266). New York: Bruner Mazel.

Snijders, T.A.B., & Bosker, R.J. (1994). Modeled variance in two-level models. *SociologicalMethods and Research, 22,* 342-363.

Speece, D. (1990). *Regular education teachers' attitudes and behaviors toward their mainstreamed learning handicapped students.* Unpublished doctoral dissertation, University of California, Los Angeles.

Sprunger, L. W., Boyce, W. T., & Gaines, J. A. (1985). Family-infant congruence: Routines and rhythmicity in family adaptations to a young infant. *Child Development, 56*, 564-572.

Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., &

Totsika, V. (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly, 21*, 76-92.

Teglasi, H. (1998). Temperament constructs and measures. *School Psychology Review, 27*, 564-585.

Teglasi, H., & MacMahon, B. H. (1990). Temperament and common problem behaviors of children. *Journal of Applied Developmental Psychology, 1,* 331-349.

Terestman, N. (1980). Mood quality and intensity in nursery school children as predictors of behavior disorders. *American Journal of Orthopsychiatry, 50*, 125-138.

Thomas, A., & Chess, S. (1977). *Temperament and development.* New York: Bruner/Mazel.

Thomas, A., & Chess, S. (1982). *Temperament and follow-up to adulthood*. Ciba Foundation Symposium, *89*, 168-75.

Thomas, A., Chess, S., & Birch, H. G. (1968). *Temperament and behavior disorders in children.* New York: New York University Press.

Thomas, A., Chess, S., Birch, H. G., Herzig, M. E., & Korn, S. (1963). *Behavioral individuality in early childhood.* New York: New York University Press.

Thomas, A., Chess, S., Sillen, J., & Mendez, O. (1974). Cross cultural study of behavior in children with special vulnerabilities to stress. In D. F. Picks, A. Thomas, and M. Roff Eds.), *Life History Research in Psychopathology (Vol. 3*, pp. 53-67). Minneapolis: University of Minneapolis Press.

Tschann, J. M., Kaiser, P., Chesney, M. A., Alkon, A., & Boyce, W. T. (1996). Resilience and vulnerability among preschool children: Family functioning, temperament, and behavior problems. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 182-192.

Vandell, D. L. (2004). Early child care: The known and the unknown. *Merrill-Palmer Quarterly, 50*, 387-414.

Van IJzendoorn, M. H., Tavecchio, L. W. C., Stams, G. J., Verhoeven, M., & Reiling, E. (1998a). Attunement between parents and professional caregivers: A comparison of childrearing attitudes in different child-care settings. *Journal of Marriage and the Family, 60*, 771-781.

Van IJzendoorn, M. H., Tavecchio, L. W. C., Stams, G. J., Verhoeven, M., & Reiling, E. (1998b). Quality of center day care and attunement between parents and caregivers: Center day care in cross-national perspective. *Journal of Genetic Psychology, 159*, 437-454.

Vaughn, B. E., Bradley, C. F., Joffe, L. S., Seifer, R., & Barglow, P. (1987). Maternal characteristics measured parentally predict ratings of temperamental "difficulty" on the Carey Infant Temperament Questionnaire. *Developmental Psychology, 23*, 152-161.

Volling, B., & Feagans, L. V. (1995). Infant day care and children's social competence. *Infant. Behavior and Development, 18*, 177-188.

Wallander, J. L., Hubert, N. C., & Varni, J. W., (1989). Child and maternal temperament characteristics, goodness of fit and adjustment in physically handicapped children. *Journal of Clinical and Child Psychology, 17*, 336-344.

Walker, H. M., & McConnell, S. R., (1995). *The Walker-McConnell Scale of Social Competence and Social Adjustment, Elementary Version.* San Diego, CA: Singular.

Wertlieb, D., Weigel, C., Springer, T., & Feldstein, M. (1987). Temperament as moderator of children's stressful experiences. *American Journal of Orthopsychiatry, 54*, 234-245.

Whitebook, M., Howes, C., & Phillips, D. (1990). *Who cares? Child care teachers and the quality of care in America. Final Report of the National Child Care Staffing Study.* Oakland, CH: Child Care Employee Project.

Windle, M., & Lerner, R. M. (1984). The role of temperament in dating relationships among young adults. *Merrill-Palmer Quarterly, 30,* 163-175.

Windle, M., & Lerner, R. M. (1986). Reassessing the dimensions of temperament individuality across the life span: The revised dimensions of temperament survey (DOTS-R). *Journal of Adolescent Research, 1*, 213-230.

Wolk, S., Seanah, C.H., Coll, C. T. G., & Carr, S. (1992). Factors affecting parents' perceptions of temperament in early infancy. *American Journal of Orthopsychiatry, 62*, 71-77.

Zajdeman, H. S., & Minnes, P. M. (1991). Predictors of children's adjustment to daycare. *Early childhood development and care, 74,* 11-28.

Zeitlin, S. (1985). *Coping Inventory*. Bensenville, IL: Scholastic Testing Service, Inc.

Zellman, G. L., Perlman, M., Le, V., & Setodji, C. M. (2008). *Assessing the validity of the Qualistar Early Learning Quality and Improvement System as a tool for improving child-care quality.* Santa Monica, CA: Rand.