MEASURING QUALITY IN RURAL KINDERGARTEN CLASSROOMS:
RELIABILITY AND VALIDITY EVIDENCE FOR THE CLASSROOM ASSESSMENT
SCORING SYSTEM, KINDERGARTEN – THIRD GRADE (CLASS K-3)

A Dissertation in
School Psychology
by
Lia E. Sandilos

© 2012 Lia E. Sandilos

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2012
The dissertation of Lia E. Sandilos was reviewed and approved* by the following:

James C. DiPerna
Associate Professor of Education
Professor-in-Charge, School Psychology Program
Dissertation Adviser, Chair of Committee

James Douglas Coatsworth
Associate Professor of Human Development and Family Studies

Pamela M. Cole
Professor of Psychology

Robert L. Hale
Professor of Education

*Signatures are on file in the Graduate School.
Abstract

The purpose of the current study was to evaluate the structural validity and stability of scores on a measure of global classroom quality, the Classroom Assessment Scoring System, Kindergarten - Third Grade (CLASS K-3; Pianta, La Paro, & Hamre, 2008). Using data from a sample of 417 kindergarten classrooms in the rural Southern and Mid-Atlantic regions of the United States, confirmatory factor analysis was used to determine the structural validity of CLASS K-3. Factor analytic findings supported a 3-factor and 10-dimension structure for the CLASS K-3; however, modifications were made to the original CLASS model. In addition, stability of CLASS scores was assessed through intraclass correlations with a subset of classrooms (n = 30). Results indicated that the Emotional Support and Classroom Organization domains demonstrated higher levels of 1-year stability than the Instructional Support domain, with Positive Climate being the most consistent dimension. Finally, hierarchical multiple regression was used to determine if professional development and classroom resources positively predicted CLASS K-3 scores. Regression analyses did not yield a statistically significant relationship between professional development and CLASS K-3 scores, and though the relationship between classroom resource variables and CLASS scores displayed statistical significance, the magnitude of effect was consistently negligible.
# Table of Contents

List of Tables ........................................................................................................ vi
List of Figures .......................................................................................................... vii

Chapter 1. Introduction and Literature Review ....................................................... 1
  National Policies and Legislation ........................................................................... 3
  Theoretical Perspectives on Early Education ....................................................... 5
  Early Childhood Observation Scales .................................................................... 13
  Overall Limitations of Observation Scales ......................................................... 28
  The Classroom Assessment Scoring System, K-3 .............................................. 31
  Reliability of CLASS Scores .............................................................................. 35
  Validity of CLASS Scores .................................................................................. 36
  Factors Influencing Classroom Quality ............................................................... 41
  Rural Education ................................................................................................... 45
  Conclusions and Rationale .................................................................................. 47
  Project Aims and Research Questions ................................................................. 48

Chapter 2. Method .................................................................................................... 50
  Participants .......................................................................................................... 50
  Measures ............................................................................................................. 51
  Procedure ............................................................................................................ 54
  Design and Data Analyses .................................................................................. 54

Chapter 3. Results ................................................................................................... 59
  Structural Validity (Hypothesis 1) ...................................................................... 59
  Stability (Hypothesis 2) ....................................................................................... 69
<table>
<thead>
<tr>
<th>Chapter/Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relationships Between Resources, Professional Development &amp; Quality (Hypothesis 3)</td>
<td>77</td>
</tr>
<tr>
<td>Chapter 4. Discussion</td>
<td>86</td>
</tr>
<tr>
<td>Structural Validity</td>
<td>87</td>
</tr>
<tr>
<td>Stability</td>
<td>89</td>
</tr>
<tr>
<td>Relationships Between Resources, Professional Development &amp; Quality</td>
<td>90</td>
</tr>
<tr>
<td>Limitations</td>
<td>92</td>
</tr>
<tr>
<td>Directions for Future Research</td>
<td>93</td>
</tr>
<tr>
<td>Implications</td>
<td>94</td>
</tr>
<tr>
<td>Conclusion</td>
<td>98</td>
</tr>
<tr>
<td>References</td>
<td>101</td>
</tr>
<tr>
<td>Appendix A. Early Childhood Measures Excluded from Literature Review</td>
<td>119</td>
</tr>
<tr>
<td>Appendix B. CLASS K-3 Protocol</td>
<td>120</td>
</tr>
<tr>
<td>Appendix C. Current Development and Training Questionnaire</td>
<td>121</td>
</tr>
<tr>
<td>Appendix D. Classroom Resources Scale - Variable List</td>
<td>122</td>
</tr>
</tbody>
</table>
List of Tables

Table 1. Characteristics of Observation Scales of Early Childhood Classroom Quality .......... 16

Table 2. Reliability and Validity Evidence for Observation Scales of Early Childhood Classroom Quality ........................................................................................................................................ 19

Table 3. Definitions and Examples of CLASS K-3 Domains and Dimensions ..................... 32

Table 4. Means, Standard Deviations, Skew, and Kurtosis Values for CLASS K-3 Dimension and Domain Scores ........................................................................................................................................ 60

Table 5. Correlations for CLASS K-3 Domains and Dimensions for Factor Analytic Sample .... 61

Table 6. Fit Indices for Pianta et al. (2008) Original CLASS K-3 Model and Subsequent Modifications ........................................................................................................................................ 63

Table 7. Fit Indices for Pakarinen et al. (2010) CLASS K-3 Model (Negative Climate Removed) and Subsequent Modifications ........................................................................................................................................ 68

Table 8. Fit Indices for Alternative CLASS K-3 Models .............................................................. 70

Table 9. Means, Standard Deviations, Skew, Kurtosis, Intraclass Correlations, and P-Values for Dimension and Domain Scores for Stability Analyses ........................................................................................................................................ 75

Table 10. Means, Standard Deviations, Skew, Kurtosis, and Missing Data Values for Continuous Predictor Variables for Regression Analyses ........................................................................................................................................ 78

Table 11. Correlations for Predictor Variables and CLASS K-3 Domains in Regression Analyses ........................................................................................................................................ 80

Table 12. Results of Hierarchical Multiple Regression Analysis for the Prediction of Emotional Support Scores ........................................................................................................................................ 81

Table 13. Results of Hierarchical Multiple Regression Analysis for the Prediction of Classroom Organization Scores ........................................................................................................................................ 83

Table 14. Results of Hierarchical Multiple Regression Analysis for the Prediction of Instructional Support Scores ........................................................................................................................................ 84
List of Figures

Figure 1. CLASS K-3 original model (Pianta et al., 2008). ......................................................... 2
Figure 2. CLASS K-3 original model with factor loadings (Pianta et al., 2008). ......................... 64
Figure 3. Revised CLASS K-3 model. ......................................................................................... 66
Figure 4. Revised Pakarinen et al. (2010) CLASS K-3 model .................................................... 67
Figure 5. One-factor model of CLASS K-3................................................................................ 71
Figure 6. Two-factor model of CLASS K-3................................................................................ 72
Figure 7. Atheoretical model of CLASS K-3. .......................................................................... 73
Chapter 1. Introduction and Literature Review

Children’s experiences in early childhood classrooms have been identified as an important predictor of future academic and social-emotional functioning (La Paro & Pianta, 2000; Zaslow, Martinez-Beck, Tout, & Halle, 2011). Instructional theory emphasizes educators’ ability to interact with their students, employ consistent behavior management techniques, and utilize instructional materials in order to engage children in learning (Rutter & Maughan, 2002). In addition, continuing education and professional development for teachers has been identified as a valuable component of classroom quality (Copple & Bredekamp, 2009). Developing instruments that can accurately evaluate characteristics of high quality teaching and document the contextual factors that may influence effective instruction could improve the quality of education.

The focus of the current study was to assess the structural validity and stability of the Classroom Assessment Scoring System, Kindergarten - Third grade (CLASS K-3); a scale developed to evaluate the quality of early classroom environments. Additional factors that potentially affect classroom quality (i.e., professional development, classroom resources) will also be explored. The following review will examine the current literature regarding high quality early education that has set the stage for the CLASS K-3 framework (Figure 1). Specifically, key educational policy and legislation, theoretical perspectives regarding early childhood development, and empirical studies of early childhood observation systems will be reviewed. There will also be a discussion of the relevant literature pertaining to professional development and classroom resources.
Figure 1. CLASS K-3 original model (Pianta et al., 2008).
National Policies and Legislation

Early childhood education (ECE) has received increased attention and scrutiny over the past three decades (Zaslow, Martinez-Beck, Tout, & Halle, 2011). Nationally, the enrollment of children in pre-kindergarten and kindergarten programs has steadily increased from 25% in 1979 to 60% in 2009 (The Anne E. Casey Foundation, 2009). In response to the growing need for ECE programming, standards emerged for ECE licensing and accreditation, and the federal government invested funds in pre-kindergarten education through large-scale programs such as Head Start, Early Head Start, and the Child Care and Development Fund (Zaslow et al., 2011). Simultaneously, educational researchers furthered efforts to establish a substantive operational definition of “high quality early education” by examining theoretical models of education and developing instruments to measure the “quality” of early childhood classroom experiences. In conjunction with the increased availability of, and interest in, ECE, legislative policies, national “best practice” guidelines, and key models of learning and development have accelerated the focus on improving early childhood learning environments and contributed to the rising expectations for the quality of ECE.

Significant events in the high-quality ECE movement include the establishment of the National Education Goals Panel (1991), the creation of the No Child Left Behind Act (2001) and the issuance of the National Association for the Education of Young Children (NAEYC) Developmentally Appropriate Practices (DAP). In 1991, the National Educational Goals Panel (NEGP) was established by the federal government to monitor state and national progress toward educational goals over a 10-year period. The first goal emphasized in NEGP addressed school readiness by stating, "By the year 2000, all children in America will start school ready to learn" (National Education Goals Report, 1991, “Goal 1: Ready to Learn,” para. 1). However, the
NEGP was discontinued when the No Child Left Behind Act (NCLB) became law in 2001. NCLB shaped the current view of education by specifying performance standards and increasing accountability. The emphasis on accountability in education led to a surge in research examining teaching practices in the early years of education that may influence later academic success of children, particularly for those children considered to be at risk for poor outcomes (Zaslow et al., 2011). Concurrently, the NAEYC released multiple DAP publications advocating for high quality education for typically developing children (birth through 8 years) and children with disabilities (e.g., NAEYC, 1987, 1997, 2009). Each DAP publication represents an application of longstanding empirical and theoretical studies of early learning and more recent scientific and educational literature. As a result, the DAP statement has influenced the advancement of early learning standards, development of policy, and the creation of scales that measure the quality of ECE (Copple & Bredekamp, 2009). A recent DAP publication (Copple & Bredekamp, 2009) specified that ECE should foster skill development and learning, positive teacher-child interactions, appropriate assessment practices, and the use of curricular materials in the classroom. These guidelines also propose that learning is best facilitated by teachers who implement clear routines, foster positive interactions, provide informative and encouraging feedback, and embed a combination of direct instruction, reasoning, and problem solving activities into daily instruction. The DAP guidelines specifically identify the kindergarten year as an essential time for children’s physical, social-emotional, and cognitive development (Copple & Bredekamp, 2009).

In addition to legislation and professional guidelines, evolving theoretical and philosophical models of ECE also have influenced didactic practices and provided guiding frameworks for conceptualizing indicators of high quality ECE. Several significant theoretical
perspectives regarding early development and effective practices in education include behavioral theory, constructivism, Piagetian developmental-stage theory, attachment theory, social cognitive theory, cognition, and the developmental systems model of child development and learning (Copple & Bredekamp, 2009; Pianta, 1999). These theories were selected for this literature review because they reflect the progression of models of early learning over the past century, and can be linked to domains often used to define high quality ECE. Additionally, aspects of each theory can be identified in the conceptual frameworks of observation scales that have been developed over the past 30 years, as well as in the scale that is the primary focus of this study, the CLASS K-3.

Theoretical Perspectives on Early Education

**Behavioral Theory.** The use of behavior management techniques in the school setting has been strongly influenced by behavioral theory. Behavioral theory is widely considered to have been developed by Watson (1919), who believed that psychology would advance as a science if researchers focused on observable and measurable behaviors in animals and humans. Skinner (1950), a successor of Watson, advanced behavioral theory by proposing that social behaviors are learned through a process of operant conditioning, in which the provision of reinforcement results in behavioral change. Two components of operant conditioning are the reinforcing stimulus and the response. A reinforcing stimulus is something that increases the likelihood that a response will occur. In other words, if a reinforcing stimulus follows a behavioral response, that response is more likely to reoccur. There are two types of reinforcement, positive and negative. Positive reinforcement occurs when the addition or presence of a stimulus increases the occurrence of the behavioral response. Conversely, reinforcement can be viewed as negative when its removal increases the likelihood that a
response will reoccur. An intervention technique that utilizes reinforcement procedures is shaping. Shaping is a method in which successive approximations (i.e., systematic rewards) of a specified behavioral outcome are reinforced (Olson & Hergenhahn, 2009; Skinner, 1950, 1971).

According to the behavioral theory paradigm, teachers can help young students learn appropriate behaviors in school by reinforcing those specific actions. Providing positive reinforcement to increase desired behaviors is considered a highly effective behavior management technique particularly for young children. However, behaviorism has been criticized for its lack of emphasis on cognition (Woolfolk, 2008).

**Constructivist Theory.** Constructivist theory stems from the research of philosophers and educators such as Lev Vygotsky (Olson & Hergenhahn, 2009). Although there are many sub-theories within constructivism, two overarching beliefs that are common in constructivist educational practices are: (a) students actively construct their own knowledge, and (b) the interaction of people and events/situations is a key component of the acquisition of knowledge and the refinement of skills (Cobb & Bowers, 1999). Because constructivism-based teaching practices subscribe to a student-centered theory of learning, constructivist classroom practices encourage the learner to interpret information and construct meaning by engaging in unstructured inquiry or critical thinking activities. In a constructivist classroom, the teacher acts as a facilitator whose primary goal is to aid students in arriving at their own understanding of new concepts. Additionally, teachers implementing constructivist techniques do not typically utilize traditional teacher-delivered instruction. Instead, they facilitate situations that will aid students in becoming actively involved in the content of the lesson (Schunk, 2008; Woolfolk, 2011).

Vygotsky was considered a social constructivist philosopher who believed that human learning, particularly the acquisition of language, is intertwined with culture and socialization.
Although Vygotsky’s original writings were published in Russian in the 1920s and 1930s, his works were not translated into English until the late 1970s and 1980s (Kozulin, 1986). Vygotsky proposed the zone of proximal development, a concept that has been embraced in the fields of developmental psychology and education. The zone of proximal development is a crucial period during which children can master specific tasks if provided with appropriate support. Vygotsky termed the support provided to children during this learning-sensitive period as scaffolding, which can manifest as the provision of steps or cues, as well as demonstration and encouragement (Kozulin, 1986).

Vygotsky’s perspective of constructivism ultimately was limited because his writings lacked a detailed application to education, making his hypotheses difficult to empirically evaluate. However, interpretation of his theory suggests that teachers should not only create a physical environment that is conducive to learning, but they should engage in assisted or guided learning techniques (Ormrod, 2011). Further, scaffolding in ECE classrooms has been identified as an important component of instructional support for young children (O’Connor & McCartney, 2007).

**Piagetian Developmental-Stage Theory.** Another renowned theorist, Jean Piaget, had a particular impact on the way in which ECE and child development were originally conceptualized. Piaget theorized that children’s cognitive development occurred in a series of stages. He hypothesized that as children move through each distinct stage, their thought processes progress from being simplistic, concrete, and egocentric to more complex, empathetic, and abstract (Piaget, 1952). Further, through his observations of children, he concluded that cognitive development results from the interaction of environmental and genetic factors. In addition, Piagetian theory states that the degree to which individuals can learn through, and
benefit from, interactions and the social transmission of knowledge is dependent upon their level of intellectual development. Therefore, children cannot learn something until they have the cognitive prerequisites to do so. Thus, children who are at one stage of development cannot be taught concepts of a higher stage (Johnson, Christie, & Wardle, 2005; Piaget, 1952; Schunk, 2008).

Although Piagetian theory was widely accepted initially, psychologists and educational researchers began to question Piaget’s hypothesized stages because research indicated that children develop in a more continuous manner, as opposed to distinct stages (Woolfolk, 2008). Research indicating that preschool children understand concepts such as number sense and object permanence has also led to the belief that Piaget underestimated the cognition of young children (Gelman, 2000; Woolfolk, 2008). Neo-piagetians have rejected the stage-like model of development, but have retained and expanded Piaget’s beliefs regarding the impact of environment and genetics on cognitive development (Ormrod, 2011).

**Attachment Theory.** Although not considered a learning theory, attachment theory has been integral in research regarding the development of emotion and the importance of early caregiver relationships. Attachment theory emerged from the works of John Bowlby and Mary Ainsworth in the mid-1900s, and continues to influence perspectives regarding early child-caregiver relationships to this day (Bretherton, 1992). Attachment is a term referring to the emotional bond formed between children and their parents or other caregiving adults. The strength and quality of this bond has implications for the emotional stability of individuals throughout their lives and the types of relationships individuals form in adulthood (Ainsworth & Bowlby, 1991; Bowlby, Ainsworth, Boston, & Rosenbluth, 1956). Early parenting style directly impacts whether children will form a secure or insecure attachment with a caregiver. Four types
of parenting styles have been identified (i.e., authoritative, authoritarian, permissive, and neglectful), with the authoritative parenting style having the most positive outcomes for children (Baumrind, 1991).

Research on attachment and parenting styles has implications for ECE teachers, who may act as the secure caregiver for children who have not developed secure attachments with their own parents. Further, research has indicated that strong early emotional relationships can enhance children’s academic achievement (Burchinal et al., 2008; O’Connor & McCartney, 2007; Pianta, La Paro, Payne Cox, & Bradley, 2002).

Social Cognitive Theory. The role of observation and modeling in education originates with social cognitive theorists. Social cognitive theory, particularly the work of Albert Bandura (e.g., Bandura, 1982; Bandura, Ross, & Ross, 1961), asserts that a triadic relationship exists between human beings, their environments, and their behaviors. Social cognitive theorists hypothesized that individuals extract information from a variety of social experiences, such as observations of peers and adults, and verbal discourse. According to this theory, two essential components of early learning that affect the acquisition of information are observation and modeling. Research in social cognitive theory has demonstrated that children can accrue a wide range of skills simply by watching other people perform those skills, and that this approach can be a positive and effective modality for teaching (Bandura, 1986; Bandura et al., 1961; Bellini & Akullian, 2007). In addition to observation and modeling, the concept of self-efficacy is elaborated upon in social cognitive theory. Specifically, beliefs about competency in a certain area directly affect performance. Thus, positive or negative feedback can have a significant impact on the outcome of one’s actions (Bandura, 1982).
The application of social cognitive theory in the classroom environment places teachers in a central position as live models from which children can learn social behaviors and academic skills. Social cognitive theory also sheds light on the importance of feedback and encouragement in relation to student performance. Research has corroborated the importance of the construct of self-efficacy by demonstrating that effective instruction provided in the classroom should be specific, encouraging, and developmentally appropriate (Bransford, Brown, & Cocking, 1999).

**Cognitive Theory and Neuroscience.** Information processing theory emerged as research on the etiology of learning shifted to internal mechanisms in the mind. While not excluding the influence of the environment on the acquisition of knowledge, information processing theory focused on the cognitive processes occurring within the individual (Newell & Simon, 1972). Information processing theorists concluded that short-term, long-term, and working memory impact the way in which human beings mentally manipulate and interpret information from the environment. Information processing theorists also emphasized the role of attention in learning. Techniques such as rehearsal, elaboration, and visual imagery have been identified as methods for strengthening working memory the long-term storage of information (Baddeley, 2003).

More recently, neuroscientific research has contributed to the understanding of the complex relationship between cognition and learning. Early brain development requires environmental stimulation, and critical periods have been identified during which synaptic connections are formed for language, emotions, and auditory and sensorimotor skills (Schunk, 2008). The implication for aiding the process of early brain development is to provide children with rich learning experiences and stimulating environments at an early age, as those experiences
and environments will augment the formation of neural networks and synaptic connections (Schunk, 2008). As such, early childhood educators are in a position to interact with children at a particularly vulnerable developmental time, and have the capacity to influence the shaping and molding of the brain during these critical periods.

**Developmental Systems Model.** The developmental systems model evaluates the impact of the child’s environment on early learning. This multi-systemic or transactional view of learning considers early relationships and the conduciveness of children’s home and school environments (Mashburn & Pianta, 2006). Two theoretical models that incorporate a developmental systems perspective and provide useful frameworks for better understanding the development of young children in relation to their early childhood environments are the Bioecological Model (Bronfenbrenner & Morris, 1998) and General Systems Theory (Pianta, 1999; Sameroff, 1995).

In the Bioecological Model, development is viewed as a process that occurs not only within the child, but also via interaction within the child’s environment. Specifically, proximal and distal environmental processes influence children’s social and cognitive development. Proximal processes consist of children’s immediate interactions with people, as well as with their physical environment and with sources of information (e.g., literature, media). The proximal processes are moderated by distal processes, which include characteristics of the educational system or the culture (e.g., family, society, media, religion, etc.; Bronfenbrenner & Morris, 1998). Based on this perspective, the environment and social processes involving teachers, peers, and schools may, in turn, influence children’s growth (Rimm-Kaufman & Chiu, 2007).

General Systems Theory, a framework addressing the complex relationships between biological, ecological, social and other living systems, also can be applied to experiences within
the classroom setting (Pianta, 1999). In this theory, teachers’ relationships are critical to children’s development and these relationships are vehicles by which children’s needs can be addressed. Teachers serve as role models and regulate behavior through interactions, relationships, and behavior management strategies. As a result, children’s strengths and needs are not just defined as their academic abilities; rather they also are seen as the result of the educator’s teaching methods in the classroom setting (Rimm-Kaufman & Chiu, 2007).

The aforementioned legislation, guidelines, and theoretical models have impacted the conceptualization of ECE environments, and have contributed to increased emphasis on the pivotal role that teachers can play in children’s academic and social development. As our understanding of, and expectations for, ECE classrooms have increased, a key challenge has been identifying method(s) and measures that can be used to assess the quality of ECE environments and promote development. One method of data collection that has often mirrored the changing perspective of ECE, and that has been used by educators and educational researchers for assessing the quality of a classroom environment is systematic observation. Systematic observation systems have the potential to provide valuable details regarding didactic techniques that may be predictive of positive academic and social-emotional outcomes (Zaslow et al., 2011).

In the following section, systematic observation scales are reviewed chronologically (oldest first). In addition to providing a brief description of the content and constructs intended to be assessed by each scale, connections are made to theories that guided their development. Additionally, the psychometric properties of each scale are examined in an attempt to identify strengths and limitations of observation scales measuring global classroom quality.
Early Childhood Observation Scales

Over the past 30 years, a number of observation scales have been created to examine various aspects of teaching and classroom quality. These observation systems have directly measured a wide range of teaching strategies, classroom and curricular resources, and administrative characteristics hypothesized to promote positive academic, social, and emotional growth in children. In this section, observation scales that assess global classroom quality and pre-date CLASS K-3 will be reviewed.

Several methods were used to identify observation scales for this review. First, literature was identified through the following search engines: PsychInfo, Proquest, and ERIC. Examples of descriptive terms used to conduct the literature searches included: *classroom, classroom quality, childhood, early childhood, ecological, elementary, observation scale, observation schedule, observation system, teacher behavior, teaching methods, teaching quality*. These searches yielded a number of relevant studies (e.g., scale validation, empirical studies using classroom observation systems). In addition, these searches yielded three compendium reports on early childhood measures (Grinder, 2007; Malone et al., 2010; Halle & Vick, 2007), which also were consulted to identify prospective scales. Finally, the Mental Measurements Yearbook was utilized to identify prospective scales that were not located through the initial search strategies, and to gain additional content-related information and psychometric data for several specific scales.

Based on these search strategies, a total of 51 early childhood classroom observation scales were identified for initial consideration. Those retained for this literature review met the following criteria:

1. Text of the scale is written in English.
2. Scale is not limited for use in a specific state/district, but appropriate for use nationally.

3. Identified as appropriate for use in kindergarten classrooms.

4. Primary purpose of the scale is to examine global aspects of the classroom environment and teacher quality.

5. Format requires direct observation.

6. Direct observation is focused on classroom-level variables (i.e., whole class, teachers/caregivers).

Of the 51 instruments identified initially, 42 were excluded from this literature review for the following reasons: Instrument was only appropriate for use in the state of Pennsylvania ($n = 1$); instrument was not in direct observation format ($n = 4$); instrument was not developed for use in kindergarten ($n = 8$); instrument focused on a specific academic subject as opposed to global classroom quality ($n = 12$); instrument’s primary focus was at the home-, individual student-, school- or district-level ($n = 17$). A complete list of the measures that did not meet criteria for selection can be found in Appendix A.

Ultimately, nine scales were selected for inclusion, as they were most consistent with the criteria. The selected scales consisted of the following measures: Classroom Practices Inventory, Kindergarten Practices (CPI-KP; Hyson, Hisch-Pasek, & Rescorla, 1990; Vartuli, 1992), Preschool Classroom Implementation Rating Scale (PCI; Frede & Miller, 1990), Early Childhood Classroom Observation Measure (ECCOM; Stipek & Byler, 2004), Arnett Caregiver Interaction Scale (CIS; Arnett, 1989); Child-Caregiver Interaction Scale (CCIS; Carl, 2007), Assessment Profile for Early Childhood Programs: Research Edition (Assessment Profile; Abbott-Shim, Neel, & Sibley, 1992, 1998), Classroom Observation System (COS; National Institute of Child Health and Human Development Early Child Care Research Network, 1998),

The overarching theoretical framework, purpose, domains assessed, age range, and rating format of each observation scale are displayed in Table 1. Synopses of published reliability and validity evidence for scores on each scale are located in Table 2. In an effort to maintain uniformity among the descriptions of observation scales, specific reliability and validity criteria were established for this literature review. Further, the scales have been organized according to the theoretical framework(s) that form the basis of each scale’s content.

Observation Systems Based on Constructivist Theory

Classroom Practices Inventory - Kindergarten Practices (CPI-KP; Hyson, Hisch-Pasek, & Rescorla, 1990; Vartuli, 1992). The CPI-KP is an observation scale intended for use in kindergarten classrooms. The constructivist philosophy and the NAEYC DAP (1987) served as the frameworks for the content of the scale (Hyson et al., 1990). The CPI-KP has two primary sections, which consist of Program/Instructional Practices and Emotional Support. The Program/Instructional Practices section has a number of items that assess both developmentally appropriate practices and developmentally inappropriate practices in elementary classrooms. The Emotional Support section assesses the degree of warmth, assistance, and encouragement that the teacher provides to students in the classroom.

---

1 Criteria for reliability were as follows: < .60 = unacceptable, .60 to .69 = marginally acceptable, .70 to .79 = relatively acceptable, and ≥ .80 = acceptable (Nunally & Bernstein, 1994; Salvia & Ysseldyke, 2007; Sattler, 2001). Criteria for validity were as follows: < .30 = weak, .30 to .70 moderate, and > .70 = strong.
### Table 1

**Characteristics of Observation Scales of Early Childhood Classroom Quality**

<table>
<thead>
<tr>
<th>Observation Scale</th>
<th>Framework</th>
<th>Purpose</th>
<th>Domains Assessed</th>
<th>Age Range</th>
<th>Rating Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom Practices Inventory, Kindergarten Practices (CPI-KP; Hyson, Hisch-Pasek, &amp; Rescorla, 1990; Vartuli, 1992)</td>
<td>Constructivist Theory</td>
<td>Measures developmentally appropriate practices and the fidelity of instructional transactions in kindergarten</td>
<td>Factors: Program or Instructional focus &amp; Emotional Climate</td>
<td>Kindergarten</td>
<td>36 items rated in time-sampling format to determine if behavior occurred in time frame – 1 hour</td>
</tr>
<tr>
<td>Preschool Classroom Implementation Rating Scale (PCIS; Frede &amp; Miller, 1990)</td>
<td>Constructivist Theory</td>
<td>Measures general quality factors for a cognitive-developmental classroom</td>
<td>Twelve subscales: Room Arrangements, Routine, Planning, Work/Free Play, Clean-up, Small Group, Outside, Circle, Teacher/Child Interactions, Classroom Management, Team Evaluation &amp; Planning</td>
<td>Ages: 3 – 6 (preschool &amp; kindergarten)</td>
<td>52 items rated as not observed, not evident, evident, or optimal; Authors suggest observers spend one full day in a classroom</td>
</tr>
<tr>
<td>Early Childhood Classroom Observation Measure (ECCOM; Stipek &amp; Byler, 2004)</td>
<td>Constructivist Theory</td>
<td>Evaluate academic instruction, social climate, &amp; resources; assess constructivist (child-centered) and didactic (teacher-centered) instructional approaches</td>
<td>Constructivist subscales: Instruction, Management, Social Climate Didactic subscales: Instruction, Management, Social Climate</td>
<td>Ages 3 – 7 (preschool, kindergarten, &amp; first grade)</td>
<td>32 items rated from 1 (practices are rarely seen) to 5 (practices predominate)</td>
</tr>
<tr>
<td>Scale (CIS; Arnett 1989)</td>
<td>Theory</td>
<td>Interaction</td>
<td>Subscales</td>
<td>Center-based and home-based early childhood environments</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------</td>
<td>-------------</td>
<td>-----------</td>
<td>----------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Caregiver Interaction</td>
<td></td>
<td>Assess teacher-child interactions via teacher’s emotional tone, discipline style, and responsiveness</td>
<td>Four subscales: Positive Interaction, Punitiveness, Permissiveness, Detachment</td>
<td>26 items rated from 1 (never true) to 4 (often true); recommended observation time is a minimum of 2 hours</td>
<td></td>
</tr>
<tr>
<td>Child-Caregiver Interaction Scale (CCIS; Barbara, 2007)</td>
<td>Attachment Theory</td>
<td>Assess caregiver interaction across age groupings and settings</td>
<td>Three domains: Emotional, Cognitive/Physical, Social</td>
<td>Infancy through school age (center- and home-based early childhood programs)</td>
<td></td>
</tr>
<tr>
<td>Assessment Profile for Early Childhood Programs (Assessment Profile; Abbott-Shim &amp; Sibley, 1992; 1998)</td>
<td>Developmental Systems</td>
<td>Evaluate Classroom Quality</td>
<td>Classroom: Learning Environment, Scheduling, Curriculum Methods, Interacting, Individualizing</td>
<td>Ages: 3 - 7 years (center-based child care, preschool programs, kindergarten classrooms)</td>
<td></td>
</tr>
<tr>
<td>Classroom Observation System (COS K; NICHD 1998)</td>
<td>Developmental Systems</td>
<td>Examines the characteristics of the classroom by observing interactions and rating qualitative characteristics of the teacher, child, and classroom.</td>
<td>Classroom Scales: Classroom Management, Overcontrol, Positive Emotional Climate, Negative Emotional Climate, Literacy Instruction, Evaluative Feedback, Instructional Conversation</td>
<td>Version available for kindergarten, first, second, and third grade</td>
<td></td>
</tr>
<tr>
<td>Assessment Practices in Early Elementary Classrooms (APEEC; Hemmeter, Maxwell, Ault, &amp; Schuster, 2001)</td>
<td>NAEYC DAP</td>
<td>Measure developmentally appropriate practices in elementary school</td>
<td>Three subscales: Physical Environment, Instructional Context, Social Context</td>
<td>Kindergarten – third grade general education classrooms</td>
<td>16-item Likert-type scale ranging from 1 (inadequate) to 7 (excellent)</td>
</tr>
</tbody>
</table>
Table 2

Reliability and Validity Evidence for Observation Scales of Early Childhood Classroom Quality

<table>
<thead>
<tr>
<th>Observation Scale</th>
<th>Sources</th>
<th>Interrater (+/-1)</th>
<th>Internal Consistency (α)</th>
<th>Test-Retest</th>
<th>Concurrent</th>
<th>Predictive</th>
<th>Construct</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPI-KP, 1990; 1992</td>
<td>Vartuli &amp; Rohs, 2009</td>
<td>.87 - .91</td>
<td>.81</td>
<td>-</td>
<td>.41</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PCI, 1990</td>
<td>Barnett et al., 1998; in press.</td>
<td>.94 - 1.00</td>
<td>.89</td>
<td>.93</td>
<td>.60</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ECCOM, 2004</td>
<td>Stipeck &amp; Byler, 2004</td>
<td>.80 - .92&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.73 - .91</td>
<td>-</td>
<td>.23&lt;sup&gt;g&lt;/sup&gt;</td>
<td>.41&lt;sup&gt;j&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td>CIS, 1989</td>
<td>Arnett, 1989; Resnick &amp; Zill, 1998</td>
<td>1.00&lt;sup&gt;1&lt;/sup&gt;; .27 - .89&lt;sup&gt;b&lt;/sup&gt;</td>
<td>.93 - .98</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Factor structure confirmed</td>
</tr>
<tr>
<td>CCIS, 2007</td>
<td>Carl, 2007</td>
<td>.88 - .93&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.94</td>
<td>-</td>
<td>.75&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.62&lt;sup&gt;k&lt;/sup&gt;</td>
<td>.67&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Assessment Profile, 1993</td>
<td>Abott-Shim, 2000; Wilkes, 1989</td>
<td>.83 - .91</td>
<td>.83 - .91&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-</td>
<td>.74-.64&lt;sup&gt;f&lt;/sup&gt;</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------------------</td>
<td>-----------</td>
<td>----------------</td>
<td>---</td>
<td>---------------</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>COS, 1998</td>
<td>Hamre&amp;Pianta, 2005; NICHD, 2004</td>
<td>.63&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-</td>
<td>.71&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.67&lt;sup&gt;b&lt;/sup&gt;; .61&lt;sup&gt;i&lt;/sup&gt;</td>
<td>-</td>
<td>Factor structure confirmed</td>
</tr>
<tr>
<td>ECERS-R, 1998</td>
<td>Perlman et al., 2004; Harms et al., 1998</td>
<td>.71</td>
<td>.71 - .88</td>
<td>.69&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.45 - .58&lt;sup&gt;i&lt;/sup&gt;</td>
<td>.49&lt;sup&gt;j&lt;/sup&gt;</td>
<td>-</td>
</tr>
<tr>
<td>APEEC, 2001</td>
<td>Maxwell et al., 2001</td>
<td>.50 – 1.00</td>
<td>.77</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.67&lt;sup&gt;b&lt;/sup&gt;; .61&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Intraclass correlation coefficient  
<sup>b</sup>Exact agreement  
<sup>c</sup>Item response theory  
<sup>d</sup>Pearson correlation  
<sup>e</sup>Outside Expert Fidelity  
<sup>f</sup>ECERS, ECERS-R  
<sup>g</sup>Reported teacher-child closeness  
<sup>h</sup>Assessment Profile  
<sup>i</sup>CIS  
<sup>j</sup>Basic Literacy Skill  
<sup>k</sup>Keystone Stars Pre-Kindergarten Quality Rating  
<sup>l</sup>Peabody Picture Vocabulary Test
Acceptable interrater reliability indices were found during CPI-KP training sessions and data collection, and the internal consistency reliability coefficient also was in the acceptable range. Concurrent validity was assessed by examining the relationship between the CPI-KP and scores from fidelity checks conducted for a constructivism-based curriculum called Project Construct. A moderate correlation was found between the CPI-KP and the curriculum fidelity evaluation (Vartuli & Rohs, 2009).

One major limitation of CPI-KP is a lack of evidence supporting its psychometric properties. Specifically, no test-retest reliability information or construct validity evidence has been reported in research. Another limitation of the CPI-KP scale, and potential reason for its moderate correlation with the curriculum fidelity expert, may be that the CPI-KP is based on the 1987 NAEYC DAP guidelines, and it has not been revised to reflect the most recent NAEYC practice guidelines.

**Preschool Classroom Implementation Rating Scale (PCI; Frede & Miller, 1990).** The PCI was originally created as a measure of fidelity for the High/Scope Perry Preschool curriculum, but it was later revised for use in all preschool and kindergarten programs (Frede & Miller, 1990; Halle & Vick, 2007). The framework for both the High/Scope Perry Preschool curriculum and the PCI was based on Vygotsky’s work in constructivist philosophy and Piaget’s research on cognitive development. The PCI is composed of twelve subscales (i.e., Room Arrangements, Routine, Planning, Work/Free Play, Clean-up, Small Group, Outside, Circle, Teacher/Child Interactions, Classroom Management, Team Evaluation & Planning).

Interrater reliability coefficients reported from training and real-time observations fell in the acceptable range. The test-retest and internal consistency reliability estimates also fell in
acceptable ranges (Barnett, Frede, Mobasher, & Mohr, 1988; Barnett et al., 2008). Convergent validity estimates revealed that classrooms with constructivist-based curricula (i.e., High/Scope Perry Preschool, Tools of the Mind) scored higher on the PCI than non-constructivist-based classrooms (Barnett et al., 1988; Barnett et al., 2008). Concurrent validity for the PCI was examined with the ECERS-R and a moderate relationship was identified (Barnett et al., 2008).

There are some limitations to the extant reliability and validity evidence for the PCI. Although the original PCI measure demonstrated acceptable reliability and validity estimates in constructivist classrooms, there is no empirical evidence of acceptable reliability and validity in non-constructivist-based classrooms. This limits the scope of the scale’s use in ECE, as not all classrooms implement constructivist curricula. Additionally, the psychometric information reported in Barnett et al. (2008) only reflected one unidentified subscale of the PCI.

**Early Childhood Classroom Observation Measure (ECCOM; Stipek & Byler, 2004).**

The ECCOM was originally developed for educational research. The theoretical framework for the scale is based on both constructivist (child-centered) and didactic (teacher-centered) theories of learning. The ECCOM measures classroom quality through Instruction, Social Climate, and Management subscales (Halle & Vick, 2007).

Interrater agreement for the ECCOM fell in the acceptable range. Internal consistency reliability indices for each subscale ranged from relatively acceptable to acceptable. Concurrent validity analyses revealed that teachers who scored higher on constructivist subscales of the ECCOM focused on higher-order thinking skills, and teachers who had higher scores on didactic subscales placed more emphasis on basic, foundational math and literacy skills (Stipek & Byler, 2004).
The primary limitation to the ECCOM is that further reliability and validity evidence is needed, as there is no evidence of test-retest reliability or construct validity. Another potential limitation of ECCOM is that the scores from constructivist and didactic subscales of this measure appeared to be considerably divergent. As such, the way in which global classroom quality is evaluated with the ECCOM remains unclear.

**Observation Systems Based on Attachment Theory**

**Arnett Caregiver Interaction Scale (CIS; Arnett, 1989).** The Caregiver Interaction Scale is a global observation rating measure of teacher interactions with children. Theoretical foundation for the development of the scale was based on research regarding the socialization practices that occur during parent-child interactions (i.e., attachment theory), as well as the NAEYC 1987 DAP statement (Arnett, 1989). The four primary subscales are Positive Interaction, Punitiveness, Detachment, and Permissiveness.

An analysis of construct validity revealed the presence of four primary factors in the CIS (i.e., Positive Interaction, Punitiveness, Detachment, Permissiveness; Arnett, 1989). Interrater reliability evidence reported from training sessions and real-time observations ranged from unacceptable to acceptable. Acceptable internal consistency estimates were also reported (Resnick & Zill, 1999).

The reliability evidence for scores on the CIS is limited because interrater agreement estimates fall into the unacceptable range, and no test-retest data are available. Additionally, there is minimal evidence regarding the concurrent or predictive validity of scores on the CIS. Another limitation is that the scale was created two decades ago, and it has not been revised since the original publication. Carl (2007) also indicated that another limitation of the CIS is the lack of clear operational definitions for the four subscales, which could lead to users’ misinterpretation of the constructs.
Child-Caregiver Interaction Scale (CCIS; Carl, 2007). The framework for the CCIS was developed to address limitations of the CIS (Carl, 2007). As such, the CCIS employs the same theoretical framework of child-caregiver attachment upon which the CIS was based, but it places emphasis on updated NAEYC recommendations for developmentally appropriate practices. The CCIS assesses three primary interaction domains: Emotional, Cognitive/Physical, and Social.

An acceptable level of interrater reliability for CCIS scores was reported from training sessions. The internal consistency reliability coefficient also fell in the acceptable range. Convergent validity was established through the strong positive correlation between the CCIS and the “Interactions” subscale of the ECERS-R. Predictive validity analyses indicated that STAR level scores (quality rating for early childhood classrooms) and teacher years of experience were found to be positive predictors of scores on the CCIS (Carl, 2007).

Some limitations were identified in the preliminary validation of the CCIS. First, there is no evidence to support adequate interrater reliability of CCIS scores beyond training sessions (Carl, 2007). There is also no evidence of construct validity for the CCIS. Since the 2007 validation study of the CCIS, no other data have been collected for this measure.

Observation Systems Based on Developmental-Systems Model

Assessment Profile for Early Childhood Programs: Research Edition (Assessment Profile; Abbott-Shim & Sibley, 1992, 1998). The Assessment Profile is a global measure of the quality of an early childhood learning environment. The Assessment Profile has been created in both a summative (classroom-level) and formative version (program-level). The summative version is called the Assessment Profile: Research Edition, and the framework of the scale appears to reflect the developmental systems model learning perspective, as it emphasizes the interaction between the child, the teacher, and the environment. The scale contains five primary
classroom domains (i.e., Learning Environment, Scheduling, Curriculum Methods, Interacting, and Individualizing; Abbott-Shim & Sibley, 1998).

The Assessment Profile: Research Edition was originally developed through Item Response Theory (IRT; Abbott-Shim, Neel, & Sibley, 1993). With the revision of the Assessment Profile (Research Edition II), second-order factor analysis was conducted indicating that the five observed scales stem from a single underlying construct of global classroom quality (Abbott-Shim, Lambert, & McCarty, 2000). Average interrater reliability estimates recorded for the Assessment Profile from formal trainings fell in the acceptable range (Abbott-Shim et al., 2000). Concurrent validity analysis between the Assessment Profile and the ECERS yielded moderate to strong correlations (Wilkes, 1989).

One limitation regarding the reliability of the Assessment Profile is that the stability of scores over time has not been examined. Another limitation arises because the majority of published research available for the Assessment Profile utilizes the summative Research Edition of the scale, yet the summative version was adapted from the longer formative version, for which no reliability and validity evidence was identified through the search methods used in this literature review.

Classroom Observation Scale (COS; National Institute of Child Health and Human Development, 1998). The COS was developed for use in the National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network. Theoretical underpinnings of the COS reflect the developmental systems model of early behavioral and academic growth (Bronfenbrenner & Morris, 1998; Pianta, 1999), as well as a focus on the formation of early emotional bonds between caregiver and student, and the use of scaffolding and modeling in instruction. The COS was based on a theoretical model that, through later
revisions, ultimately became the framework for CLASS. The two primary constructs measured on the COS consist of Emotional Climate and Instructional Climate. The Emotional Climate is composed of seven subscales consisting of Positive Climate, Negative Climate, Teacher Sensitivity, Intrusiveness, Detachment, Classroom Management, and Over-control. The Instructional Support construct consists of Literacy Instruction, Evaluative Feedback, Instructional Conversation, and Encouragement of Child Responsibility (Hamre & Pianta, 2005).

Average interrater agreement for the COS fell in the marginally acceptable range, and test-retest data fell in the relatively acceptable level range (NICHD Early Child Care Research Network, 2004). Factor analytic findings confirmed that the COS measured two primary factors (Emotional Climate and Instructional Climate; Hamre & Pianta, 2005).

One limitation of the COS is that it demonstrates inadequate levels of reliability, as none of the estimates reach the threshold for acceptable reliability, and no internal consistency reliability estimates have been reported. Additionally, information regarding the structural validity analysis of COS is limited, and a more detailed account of the analyses is needed to confirm the construct validity of the scale.

**Observation Systems Based on NAEYC DAP Guidelines**

The remaining observation scales did not reference specific theoretical frameworks. Instead, the scales were characterized primarily by the NAEYC DAP guidelines, which have been informed by both theory and practice.

**Early Childhood Environment Rating Scale – Revised (ECERS-R; Harms & Clifford, 1980; Harms, Clifford, & Cryer, 1998).** The ECERS-R is a widely used instrument that assesses characteristics of preschool, kindergarten, and child-care programs. The framework of the original ECERS (Harms & Clifford, 1980) was based on research regarding
developmentally appropriate practices during a time when the licensing and accreditation process for ECE programs was being established (Sakai, Whitebook, Wishard, & Howes, 2003). The 1998 revision of the ECERS-R utilized the NAEYC 1997 DAP guidelines as the primary conceptual framework (Harms et al., 1998). The ECERS-R measure assesses seven distinct components of the classroom: Space and Furnishings, Personal Care Routines, Language-Reasoning, Activities, Interaction, Program Structure, and Parents and Staff.

Despite its widespread use, few data have been published concerning the psychometric properties of the ECERS-R (Perlman, Zellman, & Vi-Nhuan, 2004). Internal consistency reliability estimates for scores on the ECERS-R subscales ranged from relatively acceptable to acceptable. Interrater reliability agreement fell in the relatively acceptable range (Harms et al., 1998). Average test-retest reliability fell in the marginally acceptable range (Clifford, 2005). Construct validity of scores on the ECERS-R, was established by a panel of experts who reviewed the revised scale (Zaslow et al., 2011). However, further factor analytic findings regarding the structural validity of the ECERS-R indicated that the scale assesses one overall aspect of classroom quality, as opposed to the seven distinct aspects of classroom quality suggested by the authors (Perlman et al., 2004).

A major limitation of the ECERS-R is that construct validity evidence indicates that it may not be measuring seven main components of early childhood environments (Perlman et al., 2004). In addition, the ECERS-R comprehensively evaluates physical environmental features, but provides minimal information regarding effective teaching practices. Douglas (2004) noted that a limitation of the ECERS-R is that the scale does not capture the quality of teacher-child and peer interactions in the classroom.

Interrater reliability agreement for scores derived from the APEEC ranged from unacceptable to acceptable. Internal consistency reliability fell in the marginally acceptable range. No stability evidence was reported for the APEEC. Concurrent validity analyses of the APEEC indicated a moderate correlation with the Assessment Profile: Research Edition II and a moderate correlation with the Caregiver Interaction Scale (Maxwell et al., 2001). No structural validity evidence was reported for the APEEC.

A major limitation of the validation study for the APEEC is that factor analytic methods to assess the structural validity of the scale were not conducted. As a result, the authors only report correlations to substantiate the structural validity of the scale. Additionally, a portion of reliability indices reported for the scale fall in the unacceptable range, and no stability evidence is available.

Overall Limitations of Observation Scales

In reviewing the nine observation scales that met criteria for inclusion in this review, several overarching limitations can be addressed. With regard to psychometric properties, each scale is lacking some reliability or validity evidence. Test-retest stability and construct validity
data were most consistently missing for many of the observations scales. Specifically, test-retest data were missing for four of the scales (i.e., CIS, PCI, ECCOM, CPI-KP), and construct validity evidence also was missing for five of the scales (i.e., CPI-KP, PCI, ECCOM, APEEC, CCIS). In several instances, interrater reliability data were reported as a minimum certification level from trainings as opposed to actual data from real-time observations (i.e., CIS, PCI, CCIS). Additionally, concurrent validity evidence was most frequently demonstrated via a correlation with the ECERS/ECERS-R observation scale (i.e., PCI, Assessment Profile, APEEC). However, recent validity evidence indicates that the ECERS-R may not be a structurally valid measure (Perlman et al., 2004).

Regarding theory, the most commonly referenced frameworks were constructivist learning theory, attachment theory, and the developmental-systems model. However, two scales did not reference empirically based theory (i.e., ECERS-R, APEEC), rather they were broadly based on versions of the NAEYC DAP guidelines. A dated version of the DAP guidelines (i.e., 1987) was also incorporated into the frameworks of the CIS and the CPI-KP. The 1987 DAP statement was criticized for excluding pertinent information regarding early brain research (Carl, 2007). Further, five of the scales have not been revised in over a decade (i.e., CPI-KP, PCI, CIS, COS, ECERS-R). As early childhood research and practice have been evolving rapidly during the past two decades, these measures are limited and inconsistent with current perspectives in the field. Additionally, the theoretical frameworks reported in validation studies for the reviewed observation scales often were weak or nonexistent. As such, the theoretical basis for the domains assessed on several scales had to be inferred through a critical evaluation of the scale conducted by the author of the current study.

Classroom-level factors measured by each of the scales are another key area to consider.
In order to evaluate the quality of a classroom environment, one must first identify the key components that should be measured. All of the scales emphasize caregiver/teacher emotional valence as a subcomponent of the observation; concepts that are theoretically linked to the attachment theory and the developmental-systems model of learning. The inclusion of social-emotional climate is a strength for each of the scales, as teacher-child relationships have been identified as an important aspect of early academic success (La Paro & Pianta, 2000; Pianta, La Paro, Payne, Cox, & Bradley, 2002). However, there is less consistency across measures with regard to other domains (e.g., instruction, behavior management). While some observation systems place emphasis on daily scheduling, material resources, and physical structure of the environment (i.e., ECERS-R, PCI), other scales focus on behavior management or didactic techniques (i.e., COS, CPI-KP, ECCOM). Because there are many elements to a quality classroom environment, each observation scale allows investigators to examine the classroom through a slightly different lens. However, if observation data are to be used in formal teacher evaluation, the development of clear operational definitions regarding a “high quality” classroom environment is crucial.

Of equal importance, is the consideration that data from observation systems may be used to promote instructional changes for children. If such changes are to be made, educators need to be confident that the changes are grounded in empirically sound data collection methods. As such, scales used for the purpose of teacher and classroom environment evaluation need to possess a strong theoretical foundation, and have acceptable psychometric properties. Recent ECE literature has identified exemplary early education programs as establishing positive student-teacher relationships, using consistent routine and classroom management methods, and providing instructional supports through language modeling and scaffolding (Pianta, 1999;
The Classroom Assessment Scoring System, Kindergarten through third grade (CLASS K-3) is a relatively new observation system that was created with the intent of addressing the potential gaps and limitations identified in previous scales, and effectively evaluating research-supported components of high quality education.

In the next section, the theoretical background, psychometric properties, and relevant empirical research examining the CLASS K-3 observation scale is reviewed. The strengths and limitations of the scale are assessed, as well as the unique contribution this scale has made to systematic observation systems and research regarding high quality early childhood education.

**The Classroom Assessment Scoring System, K-3**

The goal of the Classroom Assessment Scoring System (CLASS) framework is to examine the relationship between teachers and their students, as well as to evaluate the quality of instruction and behavior management in the classroom (La Paro, Pianta, & Stuhlman, 2004). Pre-kindergarten, elementary, and secondary versions of CLASS are available; however, the focus of the current study is the kindergarten through Grade 3 version. Teaching quality on all versions of CLASS is assessed in terms of three major domains consisting of Emotional Support, Classroom Organization, and Instructional Support, and 10 dimensions consisting of Positive Climate, Negative Climate, Regard for Student Perspective, Teacher Sensitivity, Behavior Management, Productivity, Instructional Learning Formats, Concept Development, Quality of Feedback, and Language Modeling. (Brief descriptions of CLASS dimensions and domains are reported in Table 3).

The primary theoretical focus of the CLASS framework is the developmental systems model of early learning, which includes the works of Bronfenbrenner and Morris (1998) and Pianta (1999), as children’s interactions with their teacher and the classroom environment (i.e.,
### Table 3

**Definitions and Examples of CLASS K-3 Domains and Dimensions**

<table>
<thead>
<tr>
<th>Domain/Dimension</th>
<th>Definition</th>
<th>Examples of Observed Classroom Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Support</td>
<td>Support for social-emotional functioning in classroom</td>
<td></td>
</tr>
<tr>
<td>Positive Climate</td>
<td>Level of positive teacher-student, peer-peer interaction</td>
<td>Using verbal or physical affection (e.g., Nice job!)</td>
</tr>
<tr>
<td>Negative Climate</td>
<td>Level of teacher-student, peer-peer negativity</td>
<td>Raising voice, rolling eyes</td>
</tr>
<tr>
<td>Teacher Sensitivity</td>
<td>Awareness/responsiveness to academic/emotional needs</td>
<td>Expressing concern for well-being; reflecting emotions</td>
</tr>
<tr>
<td>Regard for Student</td>
<td>Emphasis on student interests/autonomy</td>
<td>Providing opportunity for autonomy (e.g., class helper)</td>
</tr>
<tr>
<td>Perspectives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom Management</td>
<td>Behavior management and classroom organization</td>
<td></td>
</tr>
<tr>
<td>Behavior Management</td>
<td>Capacity to manage student behavior</td>
<td>Using behavioral techniques (e.g., positive reinforcement)</td>
</tr>
<tr>
<td>Productivity</td>
<td>Management of time during school day</td>
<td>Consistent provision of activities; rapid transitions</td>
</tr>
<tr>
<td>Instructional Learning</td>
<td>Degree of student interest and learning</td>
<td>Using diverse materials; student engagement</td>
</tr>
<tr>
<td>Formats</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instructional Support</td>
<td>Promoting language skills and cognitive development</td>
<td></td>
</tr>
<tr>
<td>Concept Development</td>
<td>Techniques used to promote analytical thinking skills</td>
<td>Linking material to students’ lives</td>
</tr>
<tr>
<td>Quality of Feedback</td>
<td>Use of feedback to strengthen skill development</td>
<td>Scaffolding new skills</td>
</tr>
<tr>
<td>Language Modeling</td>
<td>Techniques used to increase language development</td>
<td>Using and defining new vocabulary</td>
</tr>
</tbody>
</table>

*Note.* Definitions and examples were based on the CLASS K-3 manual (Pianta et al., 2008).
proximal processes) are considered crucial for academic success. The development of CLASS factors also included empirical research on high quality teaching and an extensive review of preceding observation tools used in early childhood and elementary classrooms (La Paro et al., 2004). The framework of CLASS emphasizes that the interaction of students with their environment is a crucial mechanism for early development and learning (Pianta et al., 2008). A closer examination of the constructs assessed within CLASS reveals that the scale incorporates aspects of many theoretical models of learning.

**Emotional Support.** Research regarding effective didactic practices in early childhood has emphasized that the quality of teacher-student relationships in early education has a significant influence on student learning and future academic success (La Paro & Pianta, 2000; Pianta et al., 2002). Previous literature also indicates that children often enter school without important social-emotional skills. Rimm-Kaufmann, Pianta, and Cox (2000) found that 20% of kindergarten teachers reported that approximately half of their students lack the social skills needed to achieve early academic success. Fostering social-emotional support in ECE classrooms is particularly crucial, as children with behavioral and emotional issues have been found to be less receptive to intervention as early as age 8 (Eron, 1990). Within the Emotional Support domain of CLASS, attachment and social cognitive theories are evidenced, as positive feedback and child-caregiver interaction are key components of this construct. Specifically, this domain assesses the level of positive/negative teacher-student and peer-peer interaction, as well as the degree to which the teacher demonstrates awareness/responsiveness to students’ academic and emotional needs, and the teacher’s emphasis on student interest and autonomy (Pianta et al., 2008).
**Classroom Organization.** Recent literature indicates that consistent behavior management and classroom organization techniques have been found to positively impact learning (Buyse, Verschueren, & Doumen, 2011; Domínguez, Vitiello, Maier, & Greenfield, 2010). Further, proactive redirection of misbehavior has long been considered more effective than reactive behavior management strategies (Yates & Yates, 1990). Effective classroom management strategies have been linked to higher levels of self-regulatory and adaptive behaviors in kindergarten students (Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). The Classroom Organization domain of CLASS draws from research on behavioral theory, as the use of proactive reinforcement strategies yields higher scores on this construct. Within this domain, the teachers are evaluated on their ability to proactively manage behavior, effectively make use of learning time, and maintain student engagement in instruction (Pianta et al., 2008).

**Instructional Support.** High quality instruction and feedback have long been considered to have a significant impact on the development of higher-order thinking skills (Yates & Yates, 1990). Additionally, frequent conversation and exposure to literacy has a profound impact on children’s early language development (Copple & Bredekamp, 2009). However, kindergarteners enter school with wide variation in their level of exposure to language and literacy (Hart & Risely, 2003). Thus, the Instructional Support domain emphasizes teachers’ use of techniques to promote analytical thinking skills, provision of feedback to strengthen skills, and facilitation of language development. Because the materials available in early education programs can vary widely, the Instructional Support domain is distinctive in that it assesses what teachers do with what they have, and it does not evaluate the quantity or quality of the curricular materials accessible in the physical environment (Pianta et al., 2008). The behavioral, social
cognitive, and information processing theories appear to be incorporated into the Instructional Support domain through the evaluation of scaffolding, modeling, rehearsal, and elaboration.

The CLASS observation system was designed to provide a research-based framework to assess various characteristics of teaching quality in early childhood and elementary classrooms. Since the scale was first published in 2007, CLASS has been widely adopted for use in research and evaluation in over 3,000 early childhood classrooms (Hamre, Goffin, & Kraft-Sayre, 2009). As a part of The Improving Head Start for School Readiness Act of 2007, the Office of Head Start selected CLASS as one of the primary observation scales piloted to assess the quality of Head Start classrooms nationwide (Early Childhood Learning and Knowledge Center, 2008). The growing popularity of the CLASS framework has led to national and international assessments of the psychometric properties of the CLASS, as well as the relationship between CLASS scores and a variety of academic and behavioral outcomes.

Reliability of CLASS Scores

Interrater reliability derived from scores on CLASS during training sessions and real-time observations has varied widely with indices falling in the unacceptable to acceptable ranges (.53 to 1.00; Hamre, Mashburn, Pianta, & LoCasale-Crouch, 2008; Pianta et al., 2008; Sandilos & DiPerna, 2011). Internal consistency score reliabilities for 2, 3, and 4 cycles of CLASS for preschool and third grade classrooms range from relatively acceptable to acceptable (.76 to .89; Pianta et al., 2008).

Stability of CLASS Scores. Observations conducted in the early elementary grades have indicated that there is substantial variation in children’s emotional and instructional classroom experiences. To enhance early childhood education, it is important to examine aspects of the classroom environment that are relatively stable, as well as those that are more likely to
change both within and across school years (La Paro, Rimm-Kaufman, & Pianta, 2006). Research regarding CLASS stability across a school year has indicated that data from Emotional Support factors are relatively stable in early childhood classrooms across fall and spring observations. The Classroom Organization and Instructional Support factors have demonstrated lower levels of stability across the school year (Curby, Rimm-Kaufman, & Ponitz, 2009; Pianta et al., 2008).

Prior research has focused on the stability of classroom quality throughout 1 day, over the course of 1 year, or across grade levels (Cadima, Leal, & Burchinal, 2010; Curby et al., 2009; La Paro et al., 2009). More research is needed regarding the stability of an individual teacher’s CLASS scores from one school year to the next, as this would provide information regarding the stability of didactic quality as the composition of children in a classroom changes.

Validity of CLASS Scores

Structural Validation of CLASS. Because CLASS is a new, complex system with multiple levels and dimensions, it is vital to examine the structural validity of the scale using scores from diverse educational populations. In a 2007 validation study, the fit of the CLASS framework was assessed via one-, two-, and three-factor models using data from a sample of 4,000 pre-kindergarten through 5th grade U.S. classrooms. The sample of classrooms was taken from several large-scale studies that occurred between 1998 and 2005. The scales used during the observations were versions of the CLASS pre-kindergarten through third grade frameworks, and the Classroom Observation System (COS-1, -3, -5). Results of a confirmatory factor analysis indicated that the three-factor model (Emotional Support, Classroom Organization, and Instructional Support) demonstrated a better overall fit in pre-kindergarten through third grade
classrooms than the one- and two-factor models that were tested (Hamre, Pianta, Mashburn, & Downer, 2007).

There were some notable limitations to this study, however. Most significantly, some fit indices examined for the 3-factor and 10-dimension published CLASS model, which was tested on pre-kindergarten classrooms (N=152), did not meet less stringent thresholds for fit indicating a presence of error in the model (Browne & Cudeck, 1993; Hu & Bentler, 1999). The root mean squared error of approximation (RMSEA) of .14 was above the recommended .08 threshold for acceptable fit and above the .10 threshold for mediocre fit, and the goodness-of-fit index (GFI) of .86 did not meet the .90 threshold for good fit. Tucker-Lewis index (TLI=.88) was also just below the .90 criterion for an acceptable fitting model (Browne & Cudeck, 1993; Hu & Bentler, 1999; Schermelleh-Engle, Moosbrugger, & Muller, 2003).

In addition, although the framework of three overarching domains (i.e., Emotional Support, Classroom Organization, and Instructional Support) was consistently evaluated across grades, there was variation in the types of dimensions included within those domains because different, revised versions of the CLASS and COS frameworks were used during the longitudinal data collection. As a result, direct comparison of the internal structure of CLASS across grade-levels was not possible. For example, an earlier version of CLASS, which excluded the Language Modeling and Regard for Student Perspective dimensions, was tested on the kindergarten sample in this study (Hamre et al., 2007). As such, further research is needed regarding the fit of the current CLASS K-3 model with classrooms in the primary grades.

The CLASS framework also has been examined internationally. A study conducted in Finland examined the structural validity of CLASS Pre-K using data from 49 Finnish kindergarten classrooms (Pakarinen et al., 2010). Results of an initial confirmatory factor
analysis (CFA) indicated that the three-factor model hypothesized by Pianta et al. (2008) did not fit the Finnish classroom data (CFI=.87, TLI=.82, RMSEA=.23, SRMR=.08). Additionally, Negative Climate displayed poor discriminant validity within the Emotional Support domain. A final CFA was conducted with Negative Climate removed from the model and with the residuals correlated between Behavior Management and Productivity, and between Concept Development and Quality of Feedback. Although most of the indices demonstrated good fit in the final revised model (CFI=.96, TLI=.94, SRMR=.04), the RMSEA value was still inflated above acceptable fit (RMSEA =.14). Further, the three domain factors (Emotional Support, Classroom Organization, and Instructional Support) had high inter-correlations (> .90). Because the three domains correlated highly, a one-factor model of global classroom quality was also tested. Results, though, indicated that the one-factor model also did not fit the data.

A significant limitation of this study was the small sample number of classrooms (n = 49). The sample of 49 classrooms is well below an appropriate sample size for factor analytic research and may have negatively impacted the results of the study (Cattell, 1978; Comrey & Lee, 1992; Gorusch, 1983). Another limitation of this study is that the researchers used the CLASS pre-kindergarten (Pre-K) manual in kindergarten classrooms. Although the CLASS framework relies on the underlying assumption that the same factors should be present in high quality classrooms from pre-kindergarten to third grade, the dimensions manifest differently depending on the developmental level. Thus, it would be important for the CLASS K-3 version to be used with a sample of kindergarten classrooms.

In sum, the two previous structural validity studies published to date have some significant limitations, and consequently provide limited justification for the use of CLASS K-3 in kindergarten classrooms. Thus, further research examining the internal structure of the 3-
factor, 10-dimension model of CLASS K-3 is warranted, especially considering the already widespread use of this scale in early elementary classrooms.

**Concurrent and Predictive Validity of CLASS.** Criterion-related validity estimates examining the relationship between CLASS and ECERS-R fell in the moderate range for “interaction” factors (.45 to .65; extent to which classrooms promote teacher-child interactions, facilitate effective discipline, and encourage children to communicate) and the low moderate range for “provision” factors (.33 to .36; availability of materials). The lower correlations between “provision” factors on the CLASS and ECERS-R could be attributed to differences in the way each scale assesses materials. The ECERS-R evaluates “provision” factors by assessing the cleanliness of furnishings, the abundance of books accessible in the classroom, and the types of toys and gross motor equipment available to the children (Harms, et al., 1998). In contrast, CLASS assesses “provision” factors solely based on how the teacher utilizes materials in the classroom to facilitate learning (Hamre et al., 2008; Pianta et al., 2008).

The CLASS framework was developed with the intention of identifying key aspects of teaching quality that can be related to socio-emotional and academic growth in young children (Pianta et al., 2008). Based on the framework of the scale, teachers with higher CLASS scores should have a positive influence on the academic and socio-emotional development of their students. Previous research has indicated that the constructs measured in CLASS may be related to changes in classroom behavior and academic outcomes for children in the pre-kindergarten and early elementary grades. Higher scores on the Emotional Support domain were linked to growth in sound awareness skills (Cadima et al., 2010); whereas lower scores were identified as being predictive of poor teacher-child relationships (Buyse et al., 2011). Elevated scores on the Classroom Organization domain were related to growth in learning behaviors (e.g., initiative,
engagement, problem solving; Dominguez et al., 2010), as well as an improvement in early math skills and adaptive behaviors (Cadima et al., 2010; Rimm-Kaufman et al., 2009). Interestingly, higher levels of Instructional Support were identified as being prevalent in classrooms where teachers reported having students with significantly lower levels of cognitive self-control (e.g., persistent, works toward goals), positive work habits, and engagement (Rimm-Kaufman et al., 2009).

Although correlational relationships between CLASS domains and academic/behavioral data appear to exist, further investigation of strong, long-term linkages between CLASS scores and student outcomes is warranted. As such, recommendations for future research have highlighted the necessity to explore additional classroom-level factors not measured by the CLASS framework that may contribute to classroom variability and affect CLASS scores (Dominguez et al., 2010). Additionally, because research indicates that empirically supported didactic techniques and warm teacher-student interactions have a positive influence on student outcomes, increased research on effective professional development for teachers has also been proposed as a key avenue for future research (Hamre et al., 2010, Rimm-Kaufman et al., 2009).

In light of the findings from the previous validation studies it appears that there are still some significant gaps in the structural validity of CLASS framework, as well as in the predictive nature of classroom-level variables. Limitations to the construct validity studies (i.e., Hamre et al., 2007; Pakarinen et al., 2010) warrant further exploration of the structure of CLASS across elementary grades and with diverse populations in order to provide a more comprehensive evaluation of the scale’s factor structure. Likewise, an examination of additional classroom-level variables and professional development characteristics that improve global classroom quality
could ultimately help to identify ways to better facilitate the socio-emotional and academic development of young children.

Factors Influencing Classroom Quality

Early childhood classrooms in the United States have demonstrated substantial variability with regard to classroom quality (LoCasale-Crouch et al., 2007). The degree of classroom quality has been linked not only to proximal characteristics (e.g., teacher-child interactions) but also the distal factors (e.g., teacher subject-matter knowledge, teacher beliefs and attitudes, community poverty, parent education level) affecting the school environment (La Paro et al., 2009; LoCasale-Crouch et al., 2007; Pianta, 2003). To fully understand the multi-level construct that is classroom quality, these potential factors should also be further explored. Two factors that have been identified as potentially affecting educational quality are professional development opportunities (Cohen & Hill, 2000; Pianta, 2003) and the amount and quality of material resources in the classroom (Bloom, 1998; Kainz & Vernon-Feagans, 2007).

Professional Development. The primary mechanism for changing the instructional practices of in-service teachers is professional development. The NAEYC position statement highlighted ongoing professional development as an effective method for maintaining and improving teacher quality (Copple & Bredekamp, 2009). Garet, Porter, Desimone, Birman, Kwang, and Yoon (2001) evaluated the professional development experiences of over 1,000 teachers across the United States. The core characteristics (e.g., content, substance) of effective professional development opportunities consisted of focusing on content knowledge, providing opportunities for active learning, and establishing consistency and coherence with other learning activities. Key structural features of professional development (e.g., characteristics of the
organization and design of the professional development) were identified as group activity components, collective participation, and sustained duration (Garet et al., 2001).

Follow-up studies corroborated the findings of Garet et al. and indicated that the most influential professional development opportunities focused on specific teaching practices (e.g., use of specific types of technology), included collective participation and active learning, and built on previous knowledge (Desimone, Porter, Garet, Yoon, & Birma, 2002; Quick, Holtzman, & Chaney, 2009).

In a 2009 technical report by the National Staff Development Council, professional development was defined as “a product of both externally-provided and job-embedded activities that increase teachers’ knowledge and change their instructional practice in ways that support student learning” (Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009, p. 9).

Current areas of strength for educational professional development and learning were the availability of mentoring programs for beginning teachers and the increased focus on content knowledge-based professional development activities. Limitations to professional development included its short duration, low ratings of perceived usefulness from teachers, inability to generate collaboration between teachers, and lack of provision of opportunities for teachers to influence school policy. However, the creation of specific, empirically-based professional development models was identified as having a potentially positive impact on professional learning and education (Wei et al., 2009).

Guskey (1986, 2002) developed the “Model of Teacher Change,” which proposed that a change in teacher attitudes and classroom practices comes only after teachers see first-hand evidence of improvement in outcomes for their students. Thus, the potential implications for this model were recognition that teacher change may be better facilitated if regular feedback
regarding student progress is provided, and that following up with, supporting, and providing mild pressure and/or incentive for teachers to use certain teaching techniques or strategies is essential after an initial professional development training (Guskey, 2002). Although this model provided an overarching framework for professional development, the simplicity of the model yielded little detail regarding effective ways to improve specific teaching practices (e.g., literacy, mathematics instruction).

The MyTeachingPartner professional development model was created to be used in conjunction with the CLASS observations in order to provide ECE teachers with structured, data-based feedback regarding their didactic techniques (Hamre et al., 2010). Teachers independently enrolled in the program and submitted videotaped segments of their language and literacy lessons. The participating teachers were then provided with detailed feedback in accordance with their scores on CLASS. As compared to a control group of teachers, the MyTeachingPartner participants taught in classrooms that yielded higher scores on measures of pre-literacy skill and phonological awareness (Hamre et al., 2010).

MyTeachingPartner demonstrates characteristics of a strong professional development program because of the lengthy duration, the provision of specific feedback to teachers regarding teaching techniques, and the collection of student academic data. However, the voluntary and costly nature of the program limits the number of teachers who will participate in this type of professional development.

It is essential to examine the types of professional development available to school districts, as well as the impact of professional development on teaching quality. Further research assessing professional development in a data-based context, with linkages made between types of professional development opportunities and classroom processes, can assist educators in
improving both the selection and provision of teacher training opportunities that will have the most powerful impact on classroom environment, instructional techniques, and the use of curricular resources (Mitchem, Wells, & Wells, 2003).

**Classroom Resources.** In the Bioecological theoretical model, development is viewed as a process that occurs not only within the child, but also via interaction within the child’s environment (Bronfenbrenner & Morris, 1998). An enriching home environment is fundamental to the development of children’s school readiness capabilities, as stimulating settings and experiences in early childhood (i.e., resources, relationships) are essential to brain development and knowledge construction (Shonkoff & Phillips, 2000). Household income and resources have been directly linked to children’s early cognitive stimulation (Votruba-Drzal, 2003). Similarly to the home setting, children also benefit from an enriching school environment (Kainz & Vernon-Feagans, 2007). Instructional resources in classrooms can affect the performance of both the teachers and students. High quality curricular materials are particularly important for beginning teachers, as they tend to rely closely on the curricular resources as means of instruction (Grossman & Thompson, 2004; Schwarz et al., 2008). Curricular resources facilitate instruction and affect the individual learning process of each student (Ball & Cohen, 1996; Schwarz et al., 2008). Through the effective use of resources, teachers can organize the physical classroom environment to encourage exploration, cognitive growth, and peer interaction (Tomlinson, 2009).

Research regarding resources in ECE has touched upon an array of material and subjects. High quality literacy resources in the classroom have been shown to have a positive influence on the academic growth of economically disadvantaged children in kindergarten through third grade (Kainz & Vernon-Feagans, 2007). Early familiarity with science and mathematics materials plays a key role in fostering interest and confidence in those subject areas (Remillard, 2005;
Schwarz et al., 2008). The use of socio-dramatic play materials with young children has been associated with the development of emotion regulation, empathy, role-taking, and communication skills (Hughes, 1999; Lindsey & Colwell, 2003). However, much of the available research on material resources is in the form of narrative accounts from teachers, and more quantitative research is needed regarding the attributes (e.g., quantity, quality, availability) of a wide range of classroom material resources and their effect on teaching quality (Grossman & Thomson, 2008).

Classroom resources can have a significant positive impact on education. However, current observational evidence indicates that there is a high degree of inter-classroom variability in early elementary environments with regard to the physical environment, instructional practices, and learning opportunities to which children are exposed (NICHD Early Childhood Research Network, 2002). Thus, further empirical exploration of the relationship between characteristics of physical classroom environment and teaching quality may provide additional insight into the influence of classroom resources on overall classroom quality.

**Rural Education**

The environmental context of a school system should also be considered when exploring factors that influence the quality of education. In the United States, approximately 20% of families live in rural communities (Vernon-Feagans, Cox, & Family Life Project Key Investigators, 2011). The 2010 U.S. census defined a rural area very broadly as, “all population, housing, and territory not included within an urban area”, which at maximum would include a municipality with less than 50,000 inhabitants (U.S. Census Bureau, 2010). More specific definitions for rural counties and school districts also exist. For example, The Center for Rural Pennsylvania defines any county or school district as rural when the number of people per square mile within the county/school district is less than the average population density of the state (i.e.,
284 persons per square mile; Center for Rural Pennsylvania, 2010). Because rural communities are recognized as a unique segment of the U.S. population, research regarding this demographic is warranted (Stelmach, 2011); however, limited research has been conducted on education in rural school systems (Witherspoon & Ennett, 2011).

Education programs could present opportunities to promote academic and socioemotional growth in students growing up in rural communities. For example, caregivers (e.g., preschool & early elementary teachers) can potentially play an even more significant role in children’s cognitive and social development in rural communities as families in this population are often geographically isolated. Such isolation can result in fewer influential adults in a child’s early development, thus potentially magnifying the nurturing role of teachers (Marotz-Baden, Hennon, & Brubaker, 1988). In addition, general school climate may be more positive in rural education systems, as the National Center for Education Statistics (NCES) documented that rural districts report lower incidences of school crime, threats made to teachers, student bullying, and gang-related activities, than in urban or suburban districts (NCES, 2008).

Growing up in rural communities also presents some obstacles and risks. Rural areas have been associated with higher rates of poverty since nearly a quarter of rural America currently lives below the poverty line. This demographic has been identified as having limited educational resources and fewer professional development opportunities (Vernon-Feagans et al., 2011; Mitchem et al., 2003; O’Hare, 2009). Rural schools also face a variety of obstacles with regard to the provision of evidence-based career development for educators. These challenges include the necessity of delivering services to a large geographic area, the burden of high transportation costs, and lack of professionals with expertise in low-incidence populations (e.g., vision, hearing, cognitive impairments) and early intervention (Berry, Petrim, Gravelle, &
Farmer, 2011; Mitchem et al., 2003). Moreover, because access to resources (e.g., job opportunities, technological advancements) is more limited, young professionals have increasingly migrated out of rural towns to seek employment in urban centers (Stelmach, 2011). Research on rural education has indicated that teachers and principals working in these communities were found to be less experienced, had a lower likelihood of finishing advanced degrees, and earned smaller salaries than educators in non-rural schools (Joyce, 1994; Weber, Jensen, Miller, Mosley, & Fisher, 2005). Taking these factors into account, it appears that individuals living in rural U.S. regions are potentially at risk for poor academic outcomes.

Considering the benefits and challenges facing rural communities and their education systems along with the lack of evidence-based literature regarding this population, it is important that more research regarding the quality of education in rural settings be conducted. Specifically, it may be informative to explore the impact of available career development opportunities and material resources on teacher practices, as well as to investigate how classroom quality may manifest when observations are conducted in rural school systems.

Conclusions and Rationale

Classroom quality is a construct that has received increasing attention in ECE research, as high quality teaching can have a powerful effect on student academic and behavioral outcomes. The current review of literature highlighted a variety of scales that have been developed over the past three decades to assess global classroom quality. However, many of these scales have not been revised in the past decade, lack theoretical frameworks, or have limited reliability and validity evidence. The CLASS K-3 is a relatively new and widely used measure of classroom quality that has the potential both to provide valuable data regarding effective teaching techniques and to act as a means of accountability and professional development in school
districts. However, due to limitations of previous studies (e.g., Curby et al., 2009; Hamre et al., 2007; Pakarinen et al., 2010) more research is warranted regarding the stability and structural validity of the scale.

To further understand the complex construct of classroom quality, it is also important to explore additional factors that may affect differences in classroom environments and teaching quality. Certain models of teacher professional development and types of material classroom resources have been found to positively impact behavioral and academic growth in children (e.g., Desimone et al., 2002; Hamre et al., 2010; Kainz & Vernon-Feagans, 2007; Remillard, 2005). However, more empirical research is needed regarding the relationships between professional development opportunities, attributes of classroom resources, and global teaching quality.

**Project Aims and Research Questions**

The primary purpose of this dissertation is to examine the internal structure of the Classroom Assessment Scoring System K-3 (CLASS K-3), as well as the long-term stability of CLASS K-3 scores across school years. In addition, the impact of professional development and classroom resources on teaching quality will be explored. Specifically, the following research questions and hypotheses will be addressed in the proposed study.

1. Does a three-factor model best represent the internal structure of CLASS scores?

   *Hypothesis 1.* The best-fitting CLASS K-3 model will demonstrate 10 dimensions loading on three domains (Figure 1; Pianta et al., 2008).

2. Do scores from the CLASS K-3 demonstrate stability over 1-year?

   *Hypothesis 2.* The CLASS K-3 scores will demonstrate moderate stability (.30 ≤ r < .50; Cohen 1988, 1992) across classroom observations conducted over a 1-year period.
3. Do professional development and classroom resources positively predict scores on CLASS K-3?

*Hypothesis 3.* Professional development and classroom resources scores will positively predict scores on CLASS K-3.
Chapter 2. Method

Participants

The Family Life Project (FLP) is a longitudinal study of the cognitive, social, and emotional development of young children. The FLP study has followed 1,292 children and their families since the participating children were born in 2004. Participants in the FLP sample reside in two major rural areas of high poverty in the Southern and Mid-Atlantic regions of the United States. Specifically, participants were recruited from three counties in North Carolina (NC) and three counties in Pennsylvania (PA) that met the following qualifications: (a) less than 50,000 residents; (b) not adjacent to a major metropolitan city; (c) approximately half of school-age residents eligible for free or reduced lunch. Families in the FLP study were recruited from the Obstetrics and Gynecology units of the hospitals in the six counties because the goal of the study was to recruit mothers and their infants at birth (Vernon-Feagans et al., 2011).

Stratified random sampling was used to over-sample for families with low socio-economic status (SES) in both regions. Low SES families were defined as having an annual income at or below 200% of the poverty threshold (NC=85%; PA=66%). Additionally, oversampling of African American families (67%) was conducted in the NC region, as the PA participants were 95% Caucasian (Vernon-Feagans et al., 2011).

Data for the current study were drawn from a sample of 429 classrooms (NC=234; PA=195) that enrolled participating children during their kindergarten year. Schools were recruited in the year prior to data collection and monetary incentives were provided to teachers and principals in the form of gift cards. Data were collected at two time points (i.e., fall 2008 and fall 2009). Demographic information on teachers was collected through a self-report questionnaire. Across the total classroom sample, 415 teachers were women and 14 were men.
Teachers ranged from 22 to 66 years of age (m = 41.5, σ = 11.2), with an average of 9.4 years of teaching experience (range = < 1 year - 38 years). In addition, 88% of teachers were Caucasian, 98% spoke English as their first language, and 92% were certified in elementary education.

**Measures**

**Classroom Assessment Scoring System (CLASS K-3).** CLASS K-3 assesses the quality of the classroom environment in terms of three major domains: Emotional Support, Classroom Organization, and Instructional Support. These domains are further divided into 10 dimensions. Emotional Support consists of Positive Climate, Negative Climate, Teacher Sensitivity, and Regard for Student Perspective. Classroom Organization consists of Behavior Management, Productivity, and Instructional Learning Formats. Finally, Instructional Support is comprised of Concept Development, Quality of Feedback, and Language Modeling. (Previously published psychometric data for CLASS are reported in the literature review; examination of additional psychometric evidence is the primary focus of the current study).

Each dimension is rated on a 7-point scale ranging from Low (1 - 2) to Middle (3 - 5) to High (6 - 7; see Appendix B). Raters can observe a classroom for one to four “cycles”, and CLASS guidelines recommend observing for a minimum of two cycles (Pianta et al., 2008). A cycle consists of 20 minutes of observation and 10 minutes of coding. Individual dimensions are calculated by averaging scores across cycles within an observation. The three total domain scores are calculated by summing the aggregated dimensions and dividing by the number of dimensions within that domain. In the current study, each teacher observation consisted of two cycles. The averaged dimension scores and/or total domain scores were used for each statistical analysis.

**Kindergarten Teacher Questionnaire - Current Development and Training.** FLP investigators developed the kindergarten teacher questionnaire. The questionnaire has 10 major
sections consisting of Teacher Demographics, Classroom Demographics, Assessment Use, Curricular Use, Instructional Activities and Curricular Focus, Current Development and Training, Thoughts about Teaching, Threats to Students’ Academic Success, and Difficulties for Teachers. For the purpose of the current study, data from the Teacher Demographics, Classroom Demographics, and Current Development and Training sections were analyzed (see Appendix C).

Items in the Current Development and Training section were adapted from a professional development survey administered by the U.S. Department of Education for the Reading First Implementation Evaluation: Interim Report. The professional development section of the Interim Report reviewed the curricular supports and peer coaching offered in Reading First and Title I programs established through NCLB. The actual survey questions were available in an appendix of the report (U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, 2006, Appendix D2).

The Kindergarten Teacher Questionnaire - Current Development and Training section consists of 12 questions divided into three sections. The first section poses questions regarding the frequency with which teachers have participated in specific professional development-related activities since the beginning of the school year (Never = 1, Once a month or less = 2, Two or three times a month = 3, Once or twice a week = 4, Three or four times a week = 5, Daily = 6). The second section inquires about the availability of professional development opportunities and whether or not the teacher has received the available staff development and training opportunities during the current school year (I have received this type of assistance this year = 1, Available but I did not receive = 2, Not available at my school = 3). The third section requires teachers to report the total number of hours of professional development opportunities provided
to them during the current school year in reading, mathematics, and social-emotional
development. For the purpose of this study, the sum of the items on this measure was used to
create a Total Professional Development score.

**Classroom Resources Scale.** The Classroom Resources Scale was developed as part of
the Early Childhood Longitudinal Study, Kindergarten cohort (ECLS-K), in which
characteristics of school composition and classroom instruction were used to create multi-level
models of influential child ecologies (Kainz & Vernon-Feagans, 2007). The scale examines
whether specific resources including reading, writing, and listening materials are present in the
classroom (yes = 1, no = 0), as well as if resources are accessible to the children (yes = 1, no =
0). The scale also assesses the quality (high = 3, moderate = 2, low quality = 1) and quantity
(small/insufficient = 1, moderate = 2, more than enough = 3) of literacy and other academic
resources (e.g., trade books, leveled books, word walls, headphones).

Items on this scale yield composite scores in four primary domains: Writing Center, Total
Literacy, Environmental Print, and Listening Center (Kainz, 2009). The Total Literacy Text
composite is made up of items that evaluate the presence, accessibility, quality, and quantity of
types of books (i.e., trade books, leveled books, big books, magazines). Items on the
Environmental Print composite variable assess the presence, quality, quantity, and accessibility
of additional literacy-related materials (i.e., alphabet banner, word wall, literacy chart, print
tags). The Writing Center composite is composed of items assessing the presence, accessibility,
and quality of a writing center in the classroom. Listening Center composite is composed of
items evaluating the presence, accessibility, and quality of listening materials (i.e., headphone,
CD player, etc.; see Appendix D). In the current study Writing Center and Listening Center
were dummy coded as dichotomous variables (present =1; not present = 0).
Procedure

Kindergarten teachers of participating children were recruited and observed in the fall of 2008 (Time 1) and 2009 (Time 2). A total of 461 observations were conducted, with 32 teachers observed twice over the 1-year data collection period. Observers consisted of part-time graduate assistants and full-time data collectors. Prior to the start of data collection, the graduate assistants and full-time data collectors were formally trained and certified in all measures used in FLP research. The observers traveled to various classroom sites in North Carolina and central Pennsylvania over a 16-week period. All raters used the CLASS K-3 observation scale to assess the kindergarten classroom environment. The maximum set of four cycles used in CLASS K-3 was shortened to two cycles in order to allow for multiple observations to be completed in a day. Thus, raters observed for two 30-minute cycles, totaling 1 hour.

After conducting the 1-hour CLASS K-3 observation, data collectors then completed the Classroom Resource Scale. The scale took approximately 10-15 minutes to complete and data collectors were permitted to move freely around the classroom in order to accurately analyze the various resources that were present, as well as the state or quality of the resources. If the data collector could not locate a particular resource, that material was not counted as being present in the classroom. Teachers were also given a computerized questionnaire (i.e., Kindergarten Teacher Questionnaire) that they were asked to complete on their own time in exchange for monetary compensation.

Design and Data Analyses

Structural Equation modeling using confirmatory factor analytic methods (CFA) was used to evaluate the structural validity of the CLASS K-3. The stability of CLASS K-3 scores was investigated by examining scores for the kindergarten teachers who were observed during
both fall 2008 (Time 1) and fall 2009 (Time 2) data collection periods. In addition, hierarchical multiple linear regression analyses were used to determine whether professional development and classroom resource variables were positively related to classroom quality as measured by CLASS K-3 scores.

**Hypothesis 1 (Structural Validity).** Analysis of Moment Structures (AMOS) 19.0 software was used to estimate each model through CFA. In CFA, the relationship among latent factors is tested through the *a priori* specification of models that reflect hypothesized patterns of associations between indicators (observed scores) and latent factors. To test a model with CFA, it must be “identified” (i.e., the number of free parameters is less than or equal to the number of observations; Kline, 2005). All of the models tested in the present study were over-identified, indicating that the number of observations exceeded the number of parameters. Thus it was theoretically possible to estimate a unique solution for each model (Kline, 2005).

After the estimation procedure converges to a solution, model fit must be assessed (Schermelleh-Engle et al., 2003). Tanaka (1993) recommended using multiple fit indices to assess the overall fit of a model. As such, a variety of model fit and model comparison indices were examined in the present study. Specifically, the three model fit indices assessed in this study were the root mean squared error of approximation (RMSEA), the goodness-of-fit index (GFI), and the standardized root mean square residual (SRMR). The RMSEA is an absolute fit index that represents the lack of fit of the model to the population covariance matrix. Browne and Cudeck (1993) recommended that RMSEA values less than .05 be considered as a good fit, values between .05 and .08 as an adequate fit, and values between .08 and .10 as a mediocre fit, whereas values greater than .10 are not acceptable. The most widely accepted criterion for RMSEA is that values less than .06 are indicative of good fit (Hu & Bentler, 1999). The
goodness-of-fit index (GFI) is another absolute fit index for which values greater than .90 may indicate a good fit (Hu & Bentler, 1999; Kline 2005). The standardized root mean square residual (SRMR) is an index based on covariance residuals. The SRMR is considered favorable if the value is less than or equal to .08 (Hu & Bentler, 1999), but values as high as .10 can be interpreted as acceptable (Schermelleh-Engel et al., 2003).

The model comparison fit indices examined in this study were the comparative fit index (CFI), Bentler-Bonett normed fit index (NFI), Tucker-Lewis index (TLI), Akaike Information Criterion (AIC), and the Consistent Akaike Information Criterion (CAIC). The CFI, NFI, and TLI (also known as non-normed fit index) are incremental fit indices that assess the improvement in fit of a proposed model relative to a baseline (null model). CFI, NFI, and TLI values greater than or equal to .90 are regarded as evidence for an acceptable-fitting model (Bentler & Bonett, 1980; Hu & Bentler, 1999; Kline, 2005), while values greater than .95 are considered a good fit (Schermelleh-Engel, 2003). The AIC and CAIC are predictive fit indices that assess model fit in terms of hypothetical replications from the same population. Lower values of AIC and CAIC are preferred, as smaller values are more likely to replicate (Kline, 2005). Modification indices were used to construct the model that best fit the data.

**Hypothesis 2 (Stability Analyses).** To examine stability of CLASS K-3 domains and dimensions across a 1-year period, intraclass correlations (ICC) were analyzed. Because ICC coefficients provide an estimate of the relationship between two variables of the same unit or construct, they provided an estimate of stability for each dimension and domain over time (Field, 2009). The *a priori* alpha level for statistical significance of intraclass correlations was set at $p < .05$. Statistically significant differences for dimensions and domain scores between fall 2008
(Time 1) and fall 2009 (Time 2) were calculated using paired-samples t-tests. Sample means for
dimension and domain scores were calculated by aggregating across all observations.

**Hypothesis 3 (Relationships Between Resources, Professional Development &
Quality).** Hierarchical (blockwise) entry was used to enter predictors into the multiple
regression models. The first block consisted of three sociodemographic variables (i.e., annual
teaching income, number of years of teaching experience in district, elementary education
certification) included as covariates in each model (Malmberg, Hagger, Burn, Mutton, & Colls,
2010; Palardy & Rumberger, 2008; Pianta et al., 2002). In the second block, professional
development was entered as a predictor, as a relationship between the provision of professional
development and an improvement in CLASS scores has been identified in empirical research
(Hamre et al., 2009). Next, four scores from the Classroom Resources Scale (i.e., total literacy
text sum, environmental print sum, presence of writing center, presence of listening center) were
added as the final block in the model. Each domain (i.e., Emotional Support, Classroom
Organization, and Instructional Support) was regressed upon the covariates and the primary
variables of interest. The following equation was tested across the three CLASS domains:

\[ Y_i = [b_0 + (b_1X_{i1} + b_2X_{i2} + b_3X_{i3}) + b_4X_{i4} + (b_5X_{i5} + b_6X_{i6} + b_7X_{i7} + b_8X_{i8}) + \varepsilon_i]. \]

\( Y_i = \) Domain (Emotional Support, Classroom Organization, Instructional Support)

\( X_{i1} = \) Annual Income

\( X_{i2} = \) Number of years of teaching experience in district

\( X_{i3} = \) Elementary education certification

\( X_{i4} = \) Professional Development

\( X_{i5} = \) Total Literacy Text

\( X_{i6} = \) Environmental Print
\(X_{i7} = \text{Writing Center}\)

\(X_{i8} = \text{Listening Center}\)

An *a priori* alpha level was set at \(p \leq .05\) for statistical significance, and the level of practical significance was set to reflect a moderate effect size (Adjusted \(R^2 \geq .09\)) for total variance explained by the model. Correlation coefficients and squared semi-partial correlations were interpreted based on the following criteria: small \((r < .30)\), moderate \((.30 \leq r < .50)\), and large \((r \geq .50)\). Squared semi-partial correlations \((R^2)\) were interpreted using the following criteria: low \((R^2 < .09)\), moderate \((.09 \leq R^2 < .25)\), and large \((R^2 \geq .25)\); Cohen, 1988; 1992). The \(R^2_{\text{change}}\) was examined to determine the improvement in the model as new predictors were entered \((p \leq .05)\). Standardized beta weights \((\beta)\) were assessed to evaluate the contribution of each predictor to the model. Standardized residuals with an absolute value of 3.29, more than 1% of standardized residuals with values greater than +/- 2.58, and more than 5% of standardized residuals with values greater than +/- 1.96 were identified as cause for concern, as they may be indicative of a model that does not adequately represent the data (Field, 2009).
Chapter 3. Results

Structural Validity (Hypothesis 1)

The primary focus of the current study was to evaluate the structural validity of CLASS K-3 scores with rural kindergarten classrooms. CLASS K-3 data ($N = 429$) were screened for missing values. Only three cases (< 1% of the total sample) had one or more missing data points. Using the Mahalanobis distance test (Tabachnick & Fidell, 1996), nine multivariate outliers ($p < .001$) were identified. No systematic patterns of missing data or outliers were evident among the identified cases; thus, the cases were deleted listwise (Field, 2009). Despite the removal of 12 cases, the final sample size ($N=417$) was sufficient for the planned analyses.

Descriptive statistics for each CLASS K-3 dimension and domain are provided in Table 4. Dimension scores aggregated across two cycles were used as items or indicators in the CFA. Prior to conducting CFA, the data were examined to determine if the necessary assumptions had been met. Normality was determined through a visual inspection of normal probability plots and an examination of skew and kurtosis values. Standardized indices were considered highly skewed or kurtotic at > 2.0 and > 7.0, respectively (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Based on these criteria, only one CLASS domain, Negative Climate, demonstrated severe skew and severe kurtosis (Table 4). A log-linear transformation was used to successfully normalize the Negative Climate data (transformed skew = 2.4; transformed kurtosis = 5.8).

Linearity of the data, as examined via a visual inspection of scatter plots, was met. The presence of multicollinearity or singularity (i.e., correlations greater than or equal to .90), which was assessed with the correlation matrix, was not a significant concern as only one correlation had a value of .90. (range = -.03 to .90; Table 5).
Table 4

*Means, Standard Deviations, Skew, and Kurtosis Values for CLASS K-3 Dimension and Domain Scores*

<table>
<thead>
<tr>
<th>Dimension/Domain</th>
<th>M</th>
<th>SD</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Climate</td>
<td>5.35</td>
<td>1.03</td>
<td>-.22</td>
<td>-.46</td>
</tr>
<tr>
<td>Negative Climate*</td>
<td>1.16</td>
<td>3.43</td>
<td>3.43</td>
<td>13.97</td>
</tr>
<tr>
<td>Teacher Sensitivity</td>
<td>4.90</td>
<td>1.11</td>
<td>-.09</td>
<td>-.50</td>
</tr>
<tr>
<td>Regard for Student Perspectives</td>
<td>3.97</td>
<td>1.11</td>
<td>-.13</td>
<td>-.36</td>
</tr>
<tr>
<td>Behavior Management</td>
<td>5.41</td>
<td>.98</td>
<td>-.52</td>
<td>.19</td>
</tr>
<tr>
<td>Productivity</td>
<td>5.42</td>
<td>.96</td>
<td>-.46</td>
<td>.10</td>
</tr>
<tr>
<td>Instructional Learning Formats</td>
<td>4.64</td>
<td>.96</td>
<td>-.16</td>
<td>-.28</td>
</tr>
<tr>
<td>Concept Development</td>
<td>2.60</td>
<td>1.07</td>
<td>.32</td>
<td>-.82</td>
</tr>
<tr>
<td>Quality of Feedback</td>
<td>3.39</td>
<td>1.08</td>
<td>.12</td>
<td>-.71</td>
</tr>
<tr>
<td>Language Modeling</td>
<td>3.09</td>
<td>.98</td>
<td>.21</td>
<td>-.33</td>
</tr>
<tr>
<td>Emotional Support</td>
<td>5.26</td>
<td>.76</td>
<td>-.33</td>
<td>-.07</td>
</tr>
<tr>
<td>Classroom Organization</td>
<td>5.16</td>
<td>.80</td>
<td>-.57</td>
<td>.31</td>
</tr>
<tr>
<td>Instructional Support</td>
<td>3.03</td>
<td>.92</td>
<td>.13</td>
<td>-.59</td>
</tr>
</tbody>
</table>

*Note. *Negative Climate was log transformed.*
Table 5
Correlations for CLASS K-3 Domains and Dimensions for Factor Analytic Sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Positive Climate</td>
<td>-</td>
<td>-.43**</td>
<td>.74**</td>
<td>.58**</td>
<td>.57**</td>
<td>.46**</td>
<td>.57**</td>
<td>.18**</td>
<td>.31**</td>
<td>.31**</td>
<td>.88**</td>
<td>.65**</td>
<td>.30**</td>
</tr>
<tr>
<td>2. Negative Climate</td>
<td>-</td>
<td>-.37**</td>
<td>-.27**</td>
<td>-.47**</td>
<td>-.28**</td>
<td>-.24**</td>
<td>-.03</td>
<td>-.14**</td>
<td>-.09</td>
<td>-.53**</td>
<td>-.40**</td>
<td>.10*</td>
<td></td>
</tr>
<tr>
<td>3. Teacher Sensitivity</td>
<td>-</td>
<td>.61**</td>
<td>.51**</td>
<td>.42**</td>
<td>.64**</td>
<td>.19**</td>
<td>.37**</td>
<td>.33**</td>
<td>.89**</td>
<td>.63**</td>
<td>.34**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Regard for Student</td>
<td>-</td>
<td>.24**</td>
<td>.19**</td>
<td>.56**</td>
<td>.36**</td>
<td>.36**</td>
<td>.33**</td>
<td>.82**</td>
<td>.40**</td>
<td>.40**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perspective</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Behavior Management</td>
<td>-</td>
<td>.63**</td>
<td>.39**</td>
<td>.07</td>
<td>.23**</td>
<td>.17**</td>
<td>.54**</td>
<td>.82**</td>
<td>.18**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Productivity</td>
<td>-</td>
<td>.56**</td>
<td>.07</td>
<td>.27**</td>
<td>.26**</td>
<td>.42**</td>
<td>.88**</td>
<td>.22**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Instructional Learning</td>
<td>-</td>
<td>.37**</td>
<td>.52**</td>
<td>.46**</td>
<td>.67**</td>
<td>.79**</td>
<td>.51**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formats</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Concept Development</td>
<td>-</td>
<td>.67**</td>
<td>.62**</td>
<td>.27**</td>
<td>.21**</td>
<td>.87**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Quality of Feedback</td>
<td>-</td>
<td>.70**</td>
<td>.39**</td>
<td>.41**</td>
<td>.90**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Language Modeling</td>
<td>-</td>
<td>.36**</td>
<td>.36**</td>
<td>.87**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Emotional Support</td>
<td>-</td>
<td>.65**</td>
<td>.38**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Classroom Organization</td>
<td>-</td>
<td>.37**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Instructional Support</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* **p < .01. *p < .05.
Confirmatory factor analyses (maximum likelihood extraction) were conducted for five models. The first model tested in this study was identical to the factor structure tested by Hamre et al. (2007) with the CLASS standardization sample and reported in the published version of the CLASS manual (Pianta et al., 2008). Then, based on the abbreviated model proposed by Pakarinen et al. (2010), a model excluding the Negative Climate dimension altogether was tested. Next, a model with 10 dimensions loading on one global domain, a model with 10 dimensions loading on two domains (Emotional Support and Instructional Support), and an atheoretical model with 10 dimensions randomly forced to load on three domains were tested.

As noted previously, multiple fit indices were used to evaluate goodness of fit, and, if the initial CLASS model did not demonstrate adequate fit, modification indices were used to improve model fit. Given that chi-square is particularly sensitive to sample size (Kline, 2005; Schermelleh-Engel et al., 2003), this statistic was de-emphasized when evaluating the fit of each model.

**CLASS Model.** The original CLASS K-3 framework (Hamre et al., 2007; Pianta et al., 2007) was tested with the current sample, and results indicated that the model did not fit the data well (Table 6; Figure 2). Based on modification indices, five changes were made to improve the model. First, the residuals of Productivity and Behavior Management were correlated. Second, the residuals of Behavior Management and Negative Climate were correlated. Third, the residuals of Regard for Student Perspective and Concept Development were correlated. Fourth, a direct pathway was inserted from Emotional Support to Behavior Management. Finally, the pathway from Classroom Organization to Behavior Management was removed, as the weight of this pathway was negligible after completing the other modifications. In the revised CLASS model
Table 6

*Fit Indices for Pianta et al. (2008) Original CLASS K-3 Model and Subsequent Modifications*

<table>
<thead>
<tr>
<th>Modification</th>
<th>Fit Indices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>1) CLASS (Pianta et al., 2008)</td>
<td>359.9</td>
</tr>
<tr>
<td>2) e5 ⇔ e6</td>
<td>280.1</td>
</tr>
<tr>
<td>3) e5 ⇔ e2</td>
<td>238.0</td>
</tr>
<tr>
<td>4) e8 ⇔ e4</td>
<td>208.9</td>
</tr>
<tr>
<td>5) Emotional Support → Behavior Management</td>
<td>142.2</td>
</tr>
<tr>
<td>6) Classroom Organization → Beh Mgmt removed</td>
<td>142.3</td>
</tr>
</tbody>
</table>
Figure 2. CLASS K-3 original model with factor loadings (Pianta et al., 2008).
(Figure 3), the changes improved all of the fit indices. The GFI, SRMR, CFI, NFI, and TLI all fell within the range of acceptable to good fit, and the RMSEA fell just within the .10 threshold for mediocre fit (Browne & Cudeck, 1993). The AIC and CAIC produced the lowest values once all five modifications were made to the revised CLASS model (Table 6).

**Pakarinen Model.** The second model tested in this study was based on the results of the Pakarinen et al. (2010) study. First, Negative Climate was removed from the CLASS model; however the resulting model did not demonstrate good fit with the data (Table 7). Next, the residual errors of Behavior Management and Productivity were correlated, which created a slight improvement in the model. However, the third modification recommended by Pakarinen et al. (i.e., correlated residuals of Quality of Feedback and Concept Development) resulted in a covariance matrix that was not positive definite, so the exact Pakarinen model could not be tested as part of the current study.

Instead, three additional revisions were made to the model based on modification indices: (a) correlating residuals of Regard for Student Perspective and Concept Development, (b) correlating residuals of Positive Climate and Behavior Management, (c) and setting a direct pathway from Emotional Support to Productivity (Figure 4). In the revised Pakarinen model (Figure 4), GFI, SRMR, CFI, and NFI met criteria for acceptable to good fit. Additionally, AIC and CAIC produced the lowest values once all modifications were made to the Pakarinen model. The TLI was just below the criterion threshold for acceptable fit, and the RMSEA was inflated beyond the mediocre threshold (Table 7).
Figure 3. Revised CLASS K-3 model.

SRMR = .060
RMSEA = .097
GFI = .936
CFI = .948
AIC = 194.3
Figure 4. Revised Pakarinen et al. (2010) CLASS K-3 model.

SRMR = .065
RMSEA = .123
GFI = .927
CFI = .939
AIC = 196.4
### Table 7

*Fit Indices for Pakarinen et al. (2010) CLASS K-3 Model (Negative Climate Removed) and Subsequent Modifications*

<table>
<thead>
<tr>
<th>Modification</th>
<th>$\chi^2$</th>
<th>DF</th>
<th>SRMR</th>
<th>RMSEA</th>
<th>GFI</th>
<th>CFI</th>
<th>NFI</th>
<th>TLI</th>
<th>AIC</th>
<th>CAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Pakarinen (No Negative Climate)</td>
<td>308.0</td>
<td>24</td>
<td>.084</td>
<td>.169</td>
<td>.853</td>
<td>.862</td>
<td>.853</td>
<td>.793</td>
<td>350.6</td>
<td>456.2</td>
</tr>
<tr>
<td>2) e4 &lt;-&gt; e5</td>
<td>221.8</td>
<td>23</td>
<td>.075</td>
<td>.144</td>
<td>.893</td>
<td>.903</td>
<td>.894</td>
<td>.849</td>
<td>265.8</td>
<td>376.5</td>
</tr>
<tr>
<td>3) e7 &lt;-&gt; e3</td>
<td>193.8</td>
<td>22</td>
<td>.072</td>
<td>.137</td>
<td>.911</td>
<td>.917</td>
<td>.907</td>
<td>.863</td>
<td>239.8</td>
<td>355.5</td>
</tr>
<tr>
<td>4) e1 &lt;-&gt; e4</td>
<td>157.4</td>
<td>21</td>
<td>.064</td>
<td>.125</td>
<td>.926</td>
<td>.934</td>
<td>.925</td>
<td>.886</td>
<td>205.4</td>
<td>326.2</td>
</tr>
<tr>
<td>5) Emotional Support -&gt; Productivity</td>
<td>146.4</td>
<td>20</td>
<td>.065</td>
<td>.123</td>
<td>.927</td>
<td>.939</td>
<td>.930</td>
<td>.889</td>
<td>196.4</td>
<td>322.2</td>
</tr>
</tbody>
</table>

*Note.* All $\chi^2$ values are statistically significant at $p < .01$. 
**Alternative Models.** Three alternative factor structures (i.e., one-factor, two-factor, and atheoretical) also were tested (Table 8). Given the moderate to high correlations observed between the latent factors and dimensions in this study, as well as previous research (e.g., Hamre et al., 2007; Pakarinen et al., 2010), a one-factor model was tested with the CLASS data (Figure 5). Results of this CFA, however, did not support such a model (Table 8).

A two-factor model (Figure 6) was tested with Emotional Support and Classroom Organization combined into a single factor while the Instructional Support domain remained intact as the second factor. Results indicated that the two-factor model improved upon the one-factor model, but it still demonstrated worse fit than the original three-factor CLASS model and the Pakarinen model. Finally, an atheoretical model (Figure 7) was tested, and fit indices did not demonstrate good fit (Table 8).

**Stability (Hypothesis 2)**

To assess the stability of CLASS K-3 scores, data were examined for teachers who had been observed at two time points (fall 2008 and fall 2009; \(N=32\)). There were no missing data among these cases, and assumptions for a paired \(t\)-test and correlational analyses were tested. An outlier was classified as any case 3 or more standard deviations from the mean (Hair, Black, Babin, Andersen, & Tatham, 2006). Two teachers’ scores on the Emotional Support and Instructional Support domains were outliers and subsequently deleted, resulting in 30 cases being included in the final sample. Data were assessed to determine if the sampling distribution of the differences between scores was normal (Field, 2009). Examination of standardized skewness and kurtosis values was conducted (Fabrigar et al., 1999), as well as a visual inspection of histograms
Table 8

*Fit Indices for Alternative CLASS K-3 Models*

<table>
<thead>
<tr>
<th>Modification</th>
<th>$\chi^2$</th>
<th>DF</th>
<th>SRMR</th>
<th>RMSEA</th>
<th>GFI</th>
<th>CFI</th>
<th>NFI</th>
<th>TLI</th>
<th>AIC</th>
<th>CAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Factor</td>
<td>839.6</td>
<td>35</td>
<td>.140</td>
<td>.235</td>
<td>.684</td>
<td>.633</td>
<td>.625</td>
<td>.528</td>
<td>879.6</td>
<td>980.3</td>
</tr>
<tr>
<td>2 Factors (Emotional &amp; Instructional)</td>
<td>412.2</td>
<td>34</td>
<td>.091</td>
<td>.164</td>
<td>.835</td>
<td>.828</td>
<td>.816</td>
<td>.772</td>
<td>454.2</td>
<td>559.5</td>
</tr>
<tr>
<td>Atheoretical</td>
<td>342.9</td>
<td>32</td>
<td>.209</td>
<td>.151</td>
<td>.867</td>
<td>.549</td>
<td>.533</td>
<td>.366</td>
<td>388.9</td>
<td>505.2</td>
</tr>
</tbody>
</table>

*Note.* All $\chi^2$ values are statistically significant at $p < .01$. 
Figure 5. One-factor model of CLASS K-3.

SRMR = .140
RMSEA = .235
GFI = .684
CFI = .633
AIC = 879.6
Figure 6. Two-factor model of CLASS K-3.

SRMR = .091
RMSEA = .164
GFI = .835
CFI = .828
AIC = 454.2
Figure 7. A theoretical model of CLASS K-3.

SRMR = .209
RMSEA = .151
GFI = .867
CFI = .549
AIC = 388.9
with superimposed normal curves. Results indicated that the distribution of the data approximated a normal curve. Descriptive statistics are located in Table 9.

**Emotional Support.** Intraclass correlations are reported in Table 9. The intraclass correlation for the Emotional Support domain was high (ICC=.62; p =.01). Intraclass correlations for individual dimensions on the Emotional Support domain (i.e., Positive Climate, Negative Climate, Teacher Sensitivity, and Regard for Student Perspective) varied widely, ranging from -.03 to .72. Upon closer examination of the raw data, it became apparent that, although it was a highly stable dimension, Negative Climate had no correlation because of limited variability in the ratings (i.e., nearly all teachers received a score of 1 at both time points; Suen, 1999). The paired t-test indicated that there was a statistically significant difference between average Emotional Support scores for fall 2008 and fall 2009. On average, teachers’ Emotional Support scores were higher during the 2009 observation (M = 5.55 SE = .15) than during the 2008 observation (M = 4.85, SE = .13), t (29) = -4.61, p = .00, r = .46.

**Classroom Organization.** The intraclass correlation for Classroom Organization (Table 9) was high (ICC=.56; p =.02). Individual dimensions within the Classroom Organization domain (i.e., Behavior Management, Productivity, and Instructional Learning Formats) demonstrated moderate to high correlations (.35 - .65). The paired-samples t-test indicated that there was not a statistically significant difference between average Classroom Organization scores in 2008 and 2009. On average, teachers received similar Classroom Organization scores during the 2008 (M = 5.14 SE = .16) and 2009 observations (M = 5.29, SE = .15), t (29) = -.927, p =.362, r = .39).

**Instructional Support.** The intraclass correlation for the Instructional Support domain (Table 9) was small and non-significant (ICC =.18; p =.30). Intraclass correlations for
### Table 9

*Means, Standard Deviations, Skew, Kurtosis, Intraclass Correlations, and P-Values for Dimension and Domain Scores for Stability Analyses*

<table>
<thead>
<tr>
<th>Dimension/Domain (difference score)</th>
<th>$M$</th>
<th>$SD$</th>
<th>Skew</th>
<th>Kurtosis</th>
<th>ICC</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Climate</td>
<td>-.67</td>
<td>.99</td>
<td>-.66</td>
<td>.31</td>
<td>.72</td>
<td>.00</td>
</tr>
<tr>
<td>Negative Climate</td>
<td>.03</td>
<td>.66</td>
<td>-1.41</td>
<td>7.85</td>
<td>-.03</td>
<td>.53</td>
</tr>
<tr>
<td>Teacher Sensitivity</td>
<td>-.90</td>
<td>1.26</td>
<td>-.16</td>
<td>1.66</td>
<td>.58</td>
<td>.01</td>
</tr>
<tr>
<td>Regard for Student Perspectives</td>
<td>-2.23</td>
<td>1.34</td>
<td>.09</td>
<td>.15</td>
<td>.23</td>
<td>.24</td>
</tr>
<tr>
<td>Behavior Management</td>
<td>.13</td>
<td>1.08</td>
<td>-.68</td>
<td>2.02</td>
<td>.65</td>
<td>.00</td>
</tr>
<tr>
<td>Productivity</td>
<td>.10</td>
<td>1.15</td>
<td>1.10</td>
<td>3.43</td>
<td>.50</td>
<td>.04</td>
</tr>
<tr>
<td>Instructional Learning Formats</td>
<td>-.70</td>
<td>1.23</td>
<td>.53</td>
<td>.17</td>
<td>.35</td>
<td>.13</td>
</tr>
<tr>
<td>Concept Development</td>
<td>-.88</td>
<td>1.24</td>
<td>-.95</td>
<td>.00</td>
<td>.03</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>-37</td>
<td>1.14</td>
<td>0.00</td>
<td>0.35</td>
<td>0.33</td>
<td>0.15</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Quality of Feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language Modeling</td>
<td>-0.43</td>
<td>1.14</td>
<td>-0.19</td>
<td>-0.41</td>
<td>-0.03</td>
<td>0.53</td>
</tr>
<tr>
<td>Emotional Support</td>
<td>-0.69</td>
<td>0.82</td>
<td>-0.17</td>
<td>0.43</td>
<td>0.62</td>
<td>0.01</td>
</tr>
<tr>
<td>Classroom Organization</td>
<td>-0.16</td>
<td>0.92</td>
<td>0.40</td>
<td>1.41</td>
<td>0.56</td>
<td>0.02</td>
</tr>
<tr>
<td>Instructional Support</td>
<td>-0.56</td>
<td>1.02</td>
<td>-0.63</td>
<td>0.08</td>
<td>0.18</td>
<td>0.30</td>
</tr>
</tbody>
</table>
dimensions on the Instructional Support domain (i.e., Concept development, Quality of Feedback, and Language Modeling) demonstrated low to moderate correlations ranging from -.03 to .33. On average, teachers demonstrated higher Instructional Support scores during the 2009 observation ($M = 2.74$, $SE = .10$) than during the 2008 observation ($M = 2.18$, $SE = .17$), $t(29) = -3.02$, $p = .01$, $r = .54$.

**Relationships Between Resources, Professional Development & Quality (Hypothesis 3)**

Regression analyses were conducted using SPSS version 19.0. Separate hierarchical multiple regression analyses were conducted for each CLASS domain (i.e., Emotional Support, Classroom Organization, and Instructional Support) to examine whether professional development and classroom resources scores positively predicted scores on CLASS K-3. None of the variables were missing more than 7% of the data points and no systematic pattern of missing data was evident. Thus, a total of 43 cases were deleted listwise (Field, 2009; Hair et al., 2006). Two extreme outliers were identified and removed from the data. The resulting sample size ($N=384$) was sufficient based upon Tabachnick and Fidell’s (2007) recommendation of $N > 50 + 8m$ ($m = $ the number of predictor variables).

Descriptive statistics for each continuous predictor variable are provided in Table 10. Prior to conducting regression analyses, data were examined to determine that the necessary assumptions had been met. Two categorical predictor variables (Writing Center and Listening Center) were dummy coded as present (1) or not (0).

Normality was determined through an examination of standardized skew and kurtosis values as well as a visual inspection of histograms with superimposed normal curves. All continuous predictor variables were approximately normal. Visual inspection of the standardized residual plots for each of these analyses revealed that the assumption of linearity was met. All
Table 10

*Means, Standard Deviations, Skew, Kurtosis, and Missing Data Values for Continuous Predictor Variables for Regression Analyses*

<table>
<thead>
<tr>
<th>Predictor</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Skew</th>
<th>Kurtosis</th>
<th>% Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Years Teaching</td>
<td>407</td>
<td>9.37</td>
<td>8.46</td>
<td>1.17</td>
<td>.62</td>
<td>5%</td>
</tr>
<tr>
<td>Annual Income</td>
<td>400</td>
<td>4.54</td>
<td>1.30</td>
<td>.135</td>
<td>.548</td>
<td>7%</td>
</tr>
<tr>
<td>Total Professional Development</td>
<td>397</td>
<td>18.62</td>
<td>2.91</td>
<td>.414</td>
<td>.034</td>
<td>7%</td>
</tr>
<tr>
<td>Classroom Resources Total Text</td>
<td>424</td>
<td>18.52</td>
<td>6.57</td>
<td>-.46</td>
<td>-.10</td>
<td>.1%</td>
</tr>
<tr>
<td>Classroom Resources Environmental Print</td>
<td>424</td>
<td>14.58</td>
<td>4.51</td>
<td>-.65</td>
<td>.174</td>
<td>.1%</td>
</tr>
</tbody>
</table>

*Teacher income is rated on a Likert scale where 4 = $30,001-40,000.
observations were independent. The assumption of homoscedasticity was tested by examining a scatterplot of residuals against each of the predictor variables. This indicated that the variance was evenly distributed around zero, and the assumption of homoscedasticity was not violated. The presence of multicollinearity or singularity was ruled out by examining the correlation matrix, which had values ranging from .00 to .69 (Table 11). Variance inflation factor and tolerance were also examined, and values were less than 10 and approaching 1.00 (.96 - .99), respectively. Standardized residuals were assessed for each model. In model 1, three cases (.8%) were greater than +/- 3.29, four cases (1%) were greater than +/- 2.58, and 16 values (4%) were greater than 1.96. In model 2, two cases (.5%) were greater than 3.29, 14 cases (3.6%) were greater than +/-1.96, and five cases (1.3%) were greater than 2.58. In model 3, five cases (1.3%) were greater than 1.96, and 1 case (.3%) was greater than 2.58, indicating little cause for concern.

**Model 1.** A hierarchical (blockwise) entry was used to enter predictors into Model 1, and Emotional Support was regressed upon the covariates and the primary variables of interest. The first block consisted of three sociodemographic variables (i.e., teacher annual income, number of years teaching in the district, elementary education certification) that were entered as covariates. In the second block, the total professional development score was entered as a predictor. Four variables created from the Classroom Resources Scale (i.e., total literacy text, environmental print, presence of writing center, presence of listening materials) were added as the final block in the model.

The results for Model 1 are presented in Table 12. The omnibus test for hierarchical regression was statistically significant ($p \leq .05$) for the third block only, but results did not reach the threshold for practical significance for any block: Block 1 $F(3, 383) = 1.18, p = .32$; Block 2
Table 11

*Correlations for Predictor Variables and CLASS K-3 Domains in Regression Analyses*

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Income</td>
<td>-</td>
<td>.46**</td>
<td>.14**</td>
<td>.02</td>
<td>.01</td>
<td>.00</td>
<td>.05</td>
<td>.03</td>
<td>.01</td>
<td>.07</td>
<td>.09</td>
</tr>
<tr>
<td>2. Years teaching</td>
<td>-</td>
<td>.05</td>
<td>.06</td>
<td>.05</td>
<td>-.09</td>
<td>.03</td>
<td>-.03</td>
<td>.08</td>
<td>.15**</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>3. Certification</td>
<td>-</td>
<td>.09</td>
<td>.18**</td>
<td>.16**</td>
<td>.12*</td>
<td>.07</td>
<td>.02</td>
<td>-.04</td>
<td>-.03</td>
<td>.15**</td>
<td>.01</td>
</tr>
<tr>
<td>4. Professional Development</td>
<td>-</td>
<td>.02</td>
<td>.02</td>
<td>.04</td>
<td>-.04</td>
<td>.01</td>
<td>.00</td>
<td>-.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Total Text</td>
<td>-</td>
<td>.32**</td>
<td>.03</td>
<td>.22**</td>
<td>.01</td>
<td>.00</td>
<td>.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Environmental Print</td>
<td>-</td>
<td>.16**</td>
<td>.27**</td>
<td>-.07</td>
<td>-.03</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Writing Center</td>
<td>-</td>
<td>.15**</td>
<td>.05</td>
<td>-.03</td>
<td>.14**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Listening Center</td>
<td>-</td>
<td>.10*</td>
<td>.05</td>
<td>.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Emotional Support</td>
<td>-</td>
<td>.69**</td>
<td>.38**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Classroom Organization</td>
<td>-</td>
<td>.37**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Instructional Support</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* $n = 387$. *$p < .05$. **$p < .01$. 
Table 12

Results of Hierarchical Multiple Regression Analysis for the Prediction of Emotional Support Scores

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>T</th>
<th>p-value</th>
<th>Δ R²</th>
<th>R² adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>Teacher Income</td>
<td>-.04</td>
<td>.04</td>
<td>-.06</td>
<td>-1.01</td>
<td>.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years Teaching</td>
<td>.01</td>
<td>.01</td>
<td>.12</td>
<td>2.03</td>
<td>.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certification</td>
<td>.04</td>
<td>.15</td>
<td>.01</td>
<td>.27</td>
<td>.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Block 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Professional Development</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
<td>.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Block 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.03</td>
<td>.02</td>
</tr>
<tr>
<td>Total Text</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
<td>.16</td>
<td>.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental Print</td>
<td>-.01</td>
<td>.01</td>
<td>-.07</td>
<td>-1.23</td>
<td>.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing Center</td>
<td>.17</td>
<td>.09</td>
<td>.10</td>
<td>1.99</td>
<td>.05*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Center</td>
<td>.26</td>
<td>.10</td>
<td>.14</td>
<td>2.71</td>
<td>.01*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * p ≤ .05.
In Block 1, the covariates accounted for 1% of the variance in Emotional Support scores on CLASS K-3. Adding the total professional development that a teacher had received that year (Block 2) produced negligible change in $R^2$, as less than 1% of the variance was explained. In Block 3, the change in $R^2$ when adding classroom resource variables into the model was statistically significant but of limited practical significance, explaining only 3% of the variance in Emotional Support scores. Specifically, Writing Center and Listening Center variables were statistically significant in the third block.

**Model 2.** In Model 2, Classroom Organization was regressed upon the covariates and the primary variables of interest. The results for Model 2 are presented in Table 13. The omnibus test for hierarchical regression was statistically significant for the first and second blocks, but results did not demonstrate practical significance: Block 1 $F(3, 383) = 4.12, p = .01$; Block 2 $F(3, 383) = 3.08, p = .02$; Block 3 $F(8, 383) = 1.80, p = .08$. In Block 1, the covariates accounted for only 3% of the variance in Classroom Organization scores on CLASS K-3. Adding the total annual professional development (Block 2) resulted in a negligible change in $R^2$, as less than 1% of the variance was explained. In Block 3, adding classroom resource variables into the model, also yielded a negligible change in $R^2$ (1%). In each block, number of years of teaching experience was statistically significant.

**Model 3.** In Model 3, Instructional Support was regressed upon the covariates and the primary variables of interest. The results for Model 3 are presented in Table 14. The omnibus test for hierarchical regression was statistically significant for only the third block, but results did not demonstrate practical significance: Block 1 $F(3, 383) = 2.24, p = .08$; Block 2 $F(4, 383) = 1.69, p = .15$; Block 3 $F(8, 383) = 2.42, p = .02$. In Block 1, the covariates accounted for 1% of
Table 13

Results of Hierarchical Multiple Regression Analysis for the Prediction of Classroom Organization Scores

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>T</th>
<th>p-value</th>
<th>Δ R²</th>
<th>R² adj</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Block 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher Income</td>
<td>-.02</td>
<td>.04</td>
<td>-.02</td>
<td>-.40</td>
<td>.69</td>
<td>.03</td>
<td>.02</td>
</tr>
<tr>
<td>Years Teaching</td>
<td>.02</td>
<td>.01</td>
<td>.18</td>
<td>3.27</td>
<td>.00*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certification</td>
<td>-.13</td>
<td>.15</td>
<td>-.04</td>
<td>-.83</td>
<td>.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Block 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professional Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Block 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Text</td>
<td>-.01</td>
<td>.01</td>
<td>-.03</td>
<td>-.65</td>
<td>.52</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>Environmental Print</td>
<td>-.01</td>
<td>.01</td>
<td>-.02</td>
<td>-.44</td>
<td>.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing Center</td>
<td>.00</td>
<td>.09</td>
<td>.00</td>
<td>.04</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Center</td>
<td>.13</td>
<td>.10</td>
<td>.07</td>
<td>1.33</td>
<td>.18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* *p* ≤ .05.
Table 14

Results of Hierarchical Multiple Regression Analysis for the Prediction of Instructional Support Scores

<table>
<thead>
<tr>
<th>Predictor</th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>T</th>
<th>p-value</th>
<th>Δ R²</th>
<th>R² adj</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Block 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>Teacher Income</td>
<td>.05</td>
<td>.04</td>
<td>.08</td>
<td>1.30</td>
<td>.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years Teaching</td>
<td>.01</td>
<td>.01</td>
<td>.08</td>
<td>1.36</td>
<td>.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certification</td>
<td>-.22</td>
<td>.16</td>
<td>-.07</td>
<td>-1.34</td>
<td>.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Block 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
<td>.02</td>
</tr>
<tr>
<td>Professional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Block 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>Total Text</td>
<td>.01</td>
<td>.01</td>
<td>.04</td>
<td>.75</td>
<td>.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental Print</td>
<td>.01</td>
<td>.01</td>
<td>.03</td>
<td>.61</td>
<td>.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing Center</td>
<td>.29</td>
<td>.10</td>
<td>.16</td>
<td>3.05</td>
<td>.00*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening Center</td>
<td>.08</td>
<td>.11</td>
<td>.04</td>
<td>.75</td>
<td>.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. * p ≤ .05.
the variance for Instructional Support scores on CLASS K-3. Adding the total professional
development that a teacher had received that year (Block 2) resulted in negligible change in $R^2$,
such that less than 1% of the variance for Instructional Support scores was explained. In Block 3,
the change in $R^2$ when adding classroom resource variables into the model was statistically
significant but practical significance was limited, explaining an additional 3% of the variance in
Instructional Support scores. Writing Center variable achieved statistical significance in Block 3.
In sum, although there were statistically significant predictors in each regression model, all of the
predictors demonstrated negligible practical significance.
Chapter 4. Discussion

The purpose of this project was to examine the internal structure of the Classroom Assessment Scoring System K-3 (CLASS K-3), the stability of CLASS K-3 scores, and the impact of professional development and classroom resources on teaching quality, as measured by CLASS K-3. The internal structure of CLASS K-3 was assessed through confirmatory factor analyses. Stability of scores was examined through intraclass correlations and paired t-tests. The relationship of professional development and classroom resources with CLASS K-3 scores was evaluated through hierarchical multiple regression analyses. All analyses were conducted using data from a sample of classrooms in high-poverty rural areas in the Southern and Mid-Atlantic regions of the United States.

Factor analytic findings supported a 3-factor and 10-dimension structure for the CLASS K-3; however, modifications were made to the original CLASS model and fit indices were interpreted using less stringent criteria (i.e., “good, acceptable, and mediocre criteria”, Browne & Cudeck, 1993; Schermelleh-Engel et al., 2003) than those recommended for testing structural equation models (Hu & Bentler, 1999; Kline, 2005). Results also indicated that the Emotional Support and Classroom Organization domains demonstrated higher levels of 1-year stability than the Instructional Support domain, with Positive Climate being the most consistent dimension. Regression analyses did not indicate a statistically significant relationship between professional development and CLASS K-3 scores, and though the relationship between classroom resource variables and CLASS scores displayed some statistical significance, the effect was consistently negligible in magnitude.
Structural Validity

The primary hypothesis for this study was that the best-fitting CLASS K-3 model would demonstrate 10 dimensions loading on three domains (Figure 1; Pianta et al., 2008). The findings revealed that the selected fit indices for the original CLASS model did not meet either the recommended (Hu & Bentler, 1999; Kline, 2005) or the less stringent criterion thresholds (Browne & Cudeck, 1993; Schermelleh-Engel, 2003). This finding is somewhat consistent with previous CLASS structural validity research (e.g., Hamre et al., 2007; Pakarinen et al., 2010).

Among the primary fit indices considered in this and previous studies, RMSEA, an index measuring the presence of error in a model, most consistently demonstrated poor fit. Specifically, across all models tested in the current study, the 3-factor model from the standardization study, and the revised model based on the Pakarinen study, the RMSEA index never met the criterion for good fit, and often did not meet the less stringent criterion for mediocre fit. The RMSEA of the original CLASS model tested in the current study (.157) also was fairly similar to the RMSEA produced with the CLASS model in the standardization study (.14; Hamre et al., 2007). In the standardization study, Hamre et al. (2007) hypothesized that RMSEA may not have met the criterion threshold because this index is sensitive to model size and may demonstrate slightly better fit as the number of observed variables increases (Kenny & McCoach, 2003). However, previous research on structural equation modeling has indicated that sufficient sample sizes (N >200) yield very minor changes in RMSEA even with a large increase in variables (i.e., RMSEA change of .002 with an increase from 10 to 25 observed variables; Kenny & McCoach, 2003).

The GFI and TLI also did not meet the most stringent thresholds for the original CLASS model in the current study and in the previous standardization study. The limited fit of RMSEA, GFI, and TLI suggests a pattern of non-negligible error emerging across studies. The presence of
such error indicates that the factor structure does not fit varying samples based on the criteria recommended in empirical literature for evaluating structural models. Although the overall structure has been generally consistent across studies, some facets of the model may be less stable than others.

In the current study, several minor modifications (i.e., correlating residuals) and one substantive modification (i.e., placement of a direct pathway between Emotional Support and Behavior Management) resulted in a revised CLASS model that demonstrated reasonable fit with the data. One potential explanation for the substantive modification made to CLASS K-3 is that the Behavior Management dimension yields higher scores when teachers utilize positive strategies that enable students to self-regulate, assist students in understanding the feelings of others (e.g., perspective taking), and employ subtle cues to redirect behavior, as opposed to overt disciplinary actions (Pianta et al., 2008). The emotionally-supportive strategies encompassed within the operationalization of Behavior Management may be contributing to its relationship with the Emotional Support domain in the current study. Mosier (2001) recommended that developmentally appropriate classroom management techniques foster understanding of social consciousness and prosocial behavior (e.g., perspective taking, sharing). The implementation of such behavior management techniques requires a degree of emotional sensitivity on the part of the teacher. Also, teachers with high Emotional Support scores may generally be regarded by raters as being better teachers overall, which could have contributed to the strong correlation between Emotional Support and Classroom Management.
Stability

The second hypothesis was that CLASS K-3 scores would demonstrate moderate stability across two separate observations conducted over a 1-year period. Intraclass correlations for dimension and domain scores varied widely. Positive Climate exhibited the highest level of consistency for a dimension, and the Emotional Support and Classroom Organization domains displayed higher levels of consistency than the Instructional Support domain. The consistency of the Emotional Support domain, particularly the Positive Climate dimension, is in agreement with previous research on the reliability of CLASS K-3 scores (Curby et al., 2009; Sandilos & DiPerna, 2011). Results of paired t-tests also suggested that Classroom Organization scores constituted the most stable domain, as this was the only domain for which there was not a significant difference between average classroom scores of teachers from fall 2008 to fall 2009.

The inconsistency of the Instructional Support domain over time may be a reflection of raters’ difficulty with interpretation of this domain. In a previous reliability study of CLASS Pre-K, Sandilos and DiPerna (2011) found Instructional Support to have significantly lower levels of interrater agreement than the Emotional Support domain, with Concept Development and Language Modeling being two of the least reliable dimensions on the CLASS Pre-K scale. Sandilos and DiPerna hypothesized that Instructional Support dimensions may be more difficult to rate in early childhood classrooms because there is less opportunity to observe rich examples of whole-group instruction and the facilitation of higher-order thinking skills. The low level of interrater agreement could contribute to long-term instability of scores within this domain.

The relative lack of stability of the Instructional Support domain scores may also be related to the variety of academic needs present in the transition to kindergarten, given that students enter their first year of formalized schooling with varying academic skills (Rimm-
Kaufman et al., 2000). The nature of kindergarten (i.e., young age, extended time away from parents, first time in a structured setting) may require higher levels of Emotional Support and Classroom Organization from teachers on a consistent basis; whereas the academic needs of students in the primary grades may be more variable depending on the composition of the class from one year to the next (Copple & Bredekamp, 2009; La Paro et al., 2009). In addition, the time of year in which an observation occurs also may affect the stability of CLASS scores. For example, teachers’ didactic methods may be less established during the first few months of the school year while they adjust to the potentially new and unique needs of the students in their classroom. These reliability findings have potential implications for training, research, and practice, as noted in subsequent sections. As such, future research is necessary to explore variations in the stability of CLASS scores across the primary grades (i.e., kindergarten through Grade 3) and examine the short-term stability of scores within the course of a school year (e.g., 1 week, 1 month, etc.).

**Relationships Between Resources, Professional Development & Quality**

The third hypothesis was that professional development and classroom resources scores positively predict scores on CLASS K-3. A hierarchical regression model was tested for each of the CLASS K-3 domains. Significant covariates and predictors varied depending on the domain upon which the variables were regressed. In all of the models, however, covariates and predictors individually accounted for negligible amounts of variance, and the total variance explained by each model also was minimal. Based on the current data, it appears that general professional development and classroom resources do not play a meaningful role in predicting CLASS K-3 scores with the present sample.
The observed lack of a relationship between professional development and teaching quality is consistent with previous research, which has indicated that professional development practices in schools typically are short in duration, discourage collaboration, and generally perceived as ineffective by teachers (Wei et al., 2009). In addition, the rural location of the sample also could have contributed to the current findings, as limited access to resources and a lack of professionals with unique expertise are two challenges that rural schools often encounter (Berry et al., 2011; Mitchem et al., 2003).

Despite statistical significance, the negligible relationship between classroom resources and CLASS scores supports Pianta et al.’s (2008) position that material resources in the classroom do not influence scores on CLASS. However, the primarily weak relationships between CLASS scores and teacher- and classroom-level variables raise questions about what classroom-level factors might significantly contribute to variability in CLASS scores. For example, the composition of students in the classroom (e.g., students with behavior difficulties, age-range of students) may influence teaching quality in the elementary grades such that the methods implemented to manage classroom behavior and support student learning may change considerably depending on these variables. Moreover, general professional development and classroom resources may not be as closely related to the constructs measured by CLASS as teacher-related variables that have been previously linked to the observation scale, such as socio-emotional functioning (La Paro et al., 2009; Pakarinen et al., 2010). Previous research has indicated that teachers’ experience with stress, burnout, and depression contributes significantly to aspects of teaching quality (Kyriacou, 2001; La Paro et al., 2009; Pakarinen, Kiuru, et al., 2010).
Limitations

The aforementioned results must be considered within the context of several limitations. First, the sample used in this study was not drawn from a nationally representative population, which limits the generalizability of the results. The uniqueness of the current sample, however, also could be regarded as a strength, as these data provide information about the validity and reliability of CLASS K-3 in regions and classrooms with a higher prevalence of rural poverty and a larger percentage of minority students. Further, the sample solely included kindergarten classrooms, and CLASS K-3 is intended to be used in kindergarten through third grade classrooms. An additional limitation to the sample was the relatively small number of teachers who were observed on more than one occasion, which consequently resulted in a small subsample for the stability analysis.

Second, although averaging scores is recommended in the CLASS manual, data variability is potentially reduced when scores are aggregated across observations. In addition, the limited range of the Negative Climate dimension affected both the validity and reliability analyses. The truncated distribution of Negative Climate scores required log linear transformation to normalize the data; it also yielded intraclass correlations that could not be interpreted (i.e., no correlation) due to the lack of variability in scores on this dimension.

Third, the methodology used to collect data had some limitations. For example, the professional development questionnaire was administered to teachers at the start of each school year. Thus, teachers did not have much opportunity to experience professional development before these data were collected. In addition, the CLASS K-3 observations were conducted during the first few months of the academic year, which limited the amount of time teachers had to become acquainted with the academic and socio-emotional needs of their students.
Finally, the current study solely focused on select forms of reliability and validity evidence for CLASS K-3 scores in kindergarten classrooms. While test-retest reliability and structural validity are important to consider, examination of additional types of score validity and reliability evidence is essential to justify use of scores from a measurement system such as CLASS K-3.

**Directions for Future Research**

To address these limitations and extend the current findings, future studies should include a nationally representative sample of classrooms along with a larger subsample observed at multiple time points within and across school years. The short-term stability of CLASS scores is particularly important to explore, as these data would provide information about variations in day-to-day teaching practices. Moreover, an analysis of the structural invariance of CLASS K-3 across samples drawn from different grade-levels, socioeconomic status, race, and community type (urban, suburban, rural) will provide further information regarding the nature of CLASS K-3 domains and dimensions across varying subpopulations.

Continued exploration of variables potentially contributing to variance in CLASS scores also is needed. Specifically, further research on professional development data and the accompanying classroom observations may be more informative if data are collected at the middle to end of the school year when teachers can provide a comprehensive account of the career development opportunities available to them annually. In addition, mid- to end-of-year data collection would provide teachers with additional time to adapt to the potentially unique socio-emotional and instructional needs of a new cohort of students. Further research examining the stability of observations conducted in shorter increments across a school year may enable users to better determine the point at which CLASS scores (across domains) become most stable,
and these data could be valuable indicators of the most opportune time to provide feedback during the academic year. Future research also should attempt to investigate the stability of scores with larger samples of classrooms over longer (e.g., 2-3 years) and shorter periods of time (several times within 1 year) to examine the variability of CLASS within and across school years. The potential effects of student-level variables, teacher socio-emotional functioning, and rater characteristics on the stability and validity of CLASS K-3 scores also warrant future investigation. Finally, researchers should explore the reliability and validity of CLASS K-3 scores in Grades 1-3 to determine if the psychometric properties of the scale change depending on grade level.

**Implications**

Early childhood education and its contribution to children’s cognitive, social, and emotional development has come to the forefront of educational research and policy. Information obtained from an observation system such as CLASS K-3 could provide valuable and constructive feedback to practitioners and researchers regarding effective didactic practices. The data from the CLASS observation systems are already being utilized throughout the U.S. for the purpose of enhancing early childhood education. As such, it is critical that scores yielded from this scale demonstrate strong psychometric properties.

Structural validity analyses in the current study revealed the presence of somewhat elevated error, as demonstrated by fit indices, in the original CLASS model. The non-negligible error in the original model has implications for the validity of score interpretation, as the fit of the original model raises questions about the organizational structure of CLASS, most specifically the placement of the Behavior Management dimension on the Classroom Management domain. The 3-factor and 10-dimension structure best fit the data after several
modifications. Although most were minor (i.e.,); a more substantive modification was the placement of a pathway from Emotional Support to Behavior Management. Thus, the findings suggest that the Behavior Management dimension may be a better reflection of Emotional Support than Classroom Management in the primary grades. The strong relationship between Emotional Support and Behavior Management has implications for teacher training in the early grades, as the implementation of socio-emotional training for students (e.g., teaching children to express feelings effectively) may be most effective for managing and preventing problem behaviors in the primary grades.

While the structural validity of a scale is critical, it is also of particular importance to consider the effectiveness of a scale in providing feedback and bringing about change. Previous research has linked the use of CLASS as a professional development tool to higher reading scores in students (Hamre et al., 2010). Further research studying the relationship between CLASS as a means of facilitating teacher professional development and student achievement will provide valuable insight regarding the validity and ultimately support the use of this scale. In addition, examination of the extent to which the factor structure of CLASS changes depending on the grade-level being observed would be beneficial, as this information may better elucidate the impact of students’ age/grade on classroom management practices. At present, findings from CLASS should not be relied upon as the only indicator of instructional quality. Scores should be interpreted carefully and in conjunction with other measures of teaching quality (e.g., less formal observations, parent/student feedback, student academic data) while continued studies of construct, predictive, and concurrent validity are conducted.

The stability, or in some instances lack thereof, of the dimensions and domains of CLASS K-3 has implications for the frequency of observations and the appropriateness of
interpreting scores “globally.” Based on the results of this study, the stability of Emotional Support and Classroom Organization scores lends some support for these domains as global snapshots of quality. However, the less stable nature of Instructional Support indicates that this domain might be affected by other factors, such as the varying composition of a classroom from one year to the next. Providing feedback to teachers regarding their levels of Instructional Support may help them to be mindful of the need for modifying instruction across their classes, and enable them to be more adept at perceiving and addressing those needs.

The timing of classroom observations is another variable that could be impacting the consistency of CLASS scores. Ratings at the start of the school year may be less stable because teachers are still adjusting to the new group of students in their classroom. Research assessing the stability of CLASS in secondary grades indicated that early-career teachers displayed a rapid increase in Classroom Organization scores when they were observed over time, while Instructional Support scores developed at a gradual rate (Malmberg et al., 2010). Similarly in the current study, observational data were collected within the first few months of the school year, and Classroom Organization was found to have higher stability than Instructional Support. As such, it may be that all three CLASS domains become more stable at the middle to end of the school, as teachers have had more time to adjust to the needs of their students. As noted previously, the low interrater agreement for Instructional Support has been documented in recent research (Sandilos & DiPerna, 2011). Thus, when providing training on the CLASS K-3 scale, it may be useful to provide observers with additional time and training to understand the way in which Instructional Support dimensions are demonstrated in classrooms. Practicing coding in classrooms with students who have a variety of academic needs before using the scale for research and practice would be beneficial for training purposes.
Because CLASS is intended to be a global indicator of the quality of classroom instruction (Pianta et al., 2008), the stability of scores on this measure is critical. However, if scores are not consistent over time, and teachers must vary their didactic techniques (as measured by CLASS) in order to effectively differentiate instruction to their students, then regarding a single observation as a “global snapshot” of instructional quality may not be appropriate. Instead, scores from multiple observations may be better able to capture the differentiation of instruction over time, and users may need to conduct more frequent observations in order to be confident in the overall judgment of quality. In addition, some dimensions/domains of CLASS could be more sensitive to changes in instruction than others. Previous CLASS research found that teachers with higher Instructional Support scores rated their students as having lower levels of cognitive and behavioral regulation (Rimm-Kaufman et al., 2009), indicating that more significant academic and socio-emotional needs may require higher instructional support levels provided by the teacher. If teaching practices change depending on students’ needs from year to year, or even from day to day, then CLASS users must be especially careful in their interpretation of the scores, as broad generalizations based on a limited number of observations would be inappropriate for drawing conclusions about overall teaching effectiveness.

The limitations to both structural validity and stability of scores on CLASS K-3 have potential implications for the use of the scale as a measure of accountability or a professional development tool (e.g., MyTeachingPartner; Hamre et al., 2010). Emphasis on accountability in education has increased over the past decade, and this has caused stakeholders to look for measures that can be used for high stakes decisions, such as formal evaluation of teacher performance. It is important to note that CLASS was not developed for this purpose, and does
not have sufficient evidence to support its use in this manner. Current psychometric evidence does, however, provide support for continued use of the CLASS to guide intervention, instruction, and professional development. As noted in the literature review, the dimensions and domains of CLASS have substantial theoretical support. However, the somewhat elevated error in the CLASS model when it is used with pre-kindergarten and kindergarten classroom samples indicates that some aspects of the model may be less stable than others. Furthermore, some scores yielded from the measure, in particular those of the Instructional Support domain, do not appear to be highly stable over time. Thus, if CLASS scores are used for career development purposes, it is important for users to be cognizant of the potential presence of error in the CLASS K-3 constructs, as well as the possible impact of the timing of the observation or changes in classroom composition on CLASS scores.

To most accurately assess characteristics of classroom environments, interpretation of CLASS dimensions and domains may need to be more explicitly considered in the context of student- and teacher-level variables (e.g., grade-level, class composition). Further inquiry into the structural validity and stability of CLASS scores with diverse populations, with various grades, and over shorter periods of time will help to clarify the need (or not) for such contextual differentiation. Continued research in this area should be conducted to help inform professional development models based on CLASS scores.

Conclusion

Understanding key characteristics of high quality teaching is an important consideration in early childhood education. With the increased demand for the education system’s ability to generate effective accountability measures, there is growing need for the development of valid measures of instructional quality. Creating instruments that accurately evaluate instructional
practices is one way to pinpoint effective teaching techniques and related outcomes. CLASS K-3 is one observation system that has been developed to assess the quality of instruction in the primary grades and to provide a framework for researchers and educators to systematically observe critical classroom and instructional variables (Pianta et al., 2008).

CLASS K-3 already has been widely adopted as a tool for informing research and practice in education; however, previous empirical literature has produced limited (and somewhat mixed) results about the reliability and validity of scores derived from the scale. The current study attempted to further examine psychometric evidence for CLASS K-3 through the evaluation of structural validity of the scale with a unique sample, the assessment of the stability of the scores over a 1-year period, and an examination of the relationship between CLASS scores, professional development opportunities, and classroom resources. The best-fitting 3-factor structure in the current study had several minor modifications and one substantive modification from the authors’ published model (Pianta et al., 2008), and some fit indices only met less-stringent criteria (Browne & Cudeck, 1993; Schermelleh-Engel et al., 2003). Stability analyses indicated that Emotional Support and Classroom Organization domain scores demonstrated more consistency over time than those of Instructional Support. Based on regression analyses, professional development and classroom resources played a negligible role in predicting CLASS K-3 domain scores for the current sample of teachers.

Limitations to the current study include characteristics of the sample, scores derived from CLASS K-3 (aggregated domains, truncated Negative Climate scores), and the timing of data collection procedures (professional development questionnaire). To address these limitations, future research should continue to explore the internal structure of CLASS across a variety of classrooms populations and the consistency of CLASS scores over a shorter period of time.
Investigators may want to consider additional classroom-, teacher-, and student-level variables that may impact CLASS K-3 ratings. Research should also extend the analysis of the reliability and validity of CLASS K-3 scores to Grades 1-3.

The results of structural validity analyses indicated that the original 3-factor model of CLASS K-3 fell somewhat short of the specified goodness-of-fit criteria when tested with the current sample. However, the results provided some additional support for the presence of 10 dimensions loading on three factors (domains). The results also provide support for long-term stability of certain CLASS scores and potential insight as to why some scores may be more stable than others. More research is needed regarding teacher-, student-, and classroom-level variables that may contribute variance in CLASS K-3 scores, as these data would provide additional quantitative information regarding environmental factors affecting the quality of early childhood education.

Overall, the empirical review of literature indicated that there is a great deal of theoretical support for the constructs within CLASS K-3. However, consistent with previous research, the results of the current study indicate that there are some limitations to the psychometric evidence for scores from the scale (e.g. slight variations in factor structure across studies, stability of scores over time). As such, consideration of these limitations should be taken into account when interpreting and using the scores from CLASS K-3 in research and practice. Provided future research yields evidence consistent with the current and previous studies, CLASS K-3 can be used to assess critical classroom processes, inform professional development for teachers, and, ultimately, improve teaching and learning in primary classrooms.
References


doi:10.1111/0161-4681.00057


doi:10.3102/01623737024002081


doi: 10.1016/j.ecresq.2007.05.002


doi:10.1080/10705519909540118


doi:10.1080/0144341900100304

Appendix A

Early Childhood Measures Excluded from Literature Review

Scale identified as appropriate for use in one state (i.e., Pennsylvania)

1. Caregiver Observation Form and Scale

Format is not direct observation

1. Program Administration Scale (also not classroom focus)
2. Emlen Scales
3. Assessment of Classroom Environment
4. Teacher Evaluation Scale

Does not have classroom- or teacher-level focus

1. School Age Care Environment Rating Scale (after-school care)
2. Family Child Care Environment Rating Scale (home)
3. Assessment Profile for Family and Child Care Homes (home)
4. The Child Care Assessment Tool for Relatives (home)
5. The Child Care Home Inventories (home)
6. Child/Home Early Language and Literacy (home; also has specific subject focus)
7. Observation Record of Caregiving Environment (home; also not appropriate for kindergarten)
8. Child Development Program Evaluation Scale (entire pre-k program)
9. Ready School Assessment (entire elementary school)
10. School Observation Measure (entire elementary, secondary, or high school)
11. The Emergent Academic Snapshot (individual child)
12. Child-Caregiver Observation System (individual child)
13. Early Childhood Observation System (individual child)
14. Individualizing Student Instruction (ISI) Classroom Observation and Coding System (individual child)
15. Classroom Observation System – Revised (gifted students)
16. Differentiated Classroom Observation Scales (gifted students)
17. Sheltered Instruction Observation Protocol (ELL students)

Scale is not appropriate for kindergarten classroom

1. Infant and Toddler Environment Rating Scale
2. Teacher Behavior Rating Scale
3. Teacher Interaction and Language Scale
## Appendix B

### CLASS K-3 Protocol

<table>
<thead>
<tr>
<th>CONTENT</th>
<th>FORMAT</th>
<th>CLASS K-3 Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### OBSERVATION SHEET

**Teacher:** GOI  
**Observer:** 222  
**Start time:** 6:30 a.m.  
**End time:** 6:50 a.m.  
**Number of adults:** 3  
**Number of children:** 5  

<table>
<thead>
<tr>
<th>CONTENT</th>
<th>Notes</th>
<th>FORMAT</th>
<th>Notes</th>
<th>CLASS K-3 Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. **Positive Climate (PC)**  
   - Positive Relationships  
   - Positive Affect  
   - Positive Communication  
   - Respect  
   - Notes: Teacher sitting on floor close to students  
   - Notes: Good. Fast. (Score 6)

2. **Negative Climate (NC)**  
   - Negative Affect  
   - Punitive Control  
   - Sarcasm/Disrespect  
   - Severe Negativity  
   - Notes: None observed  
   - Notes: Good. Fast. (Score 6)

3. **Teacher Sensitivity (TS)**  
   - Awareness  
   - Responsiveness  
   - Addressing Problems  
   - Student Comfort  
   - Notes: Teacher directed lesson, but gave tasks a role  
   - Notes: Good. Fast. (Score 6)

4. **Regard for Student Perspectives (RSP)**  
   - Flexibility and Student Focus  
   - Support for Autonomy and Leadership  
   - Student Expression  
   - Restriction of Movement  
   - Notes: Teacher directed lesson, but gave tasks a role  
   - Notes: Good. Fast. (Score 6)

5. **Behavior Management (BM)**  
   - Clear Behavior Expectations  
   - Proactive  
   - Restriction of Misbehavior  
   - Student Behavior  
   - Notes: Teacher directed lesson, but gave tasks a role  
   - Notes: Good. Fast. (Score 6)

6. **Productivity (PO)**  
   - Maximizing Learning Time  
   - Routines  
   - Transitions  
   - Preparation  
   - Notes: Teacher directed lesson, but gave tasks a role  
   - Notes: Good. Fast. (Score 6)

7. **Instructional Learning Formats (ILF)**  
   - Effective Facilitation  
   - Variety of Modalities and Materials  
   - Student Interest  
   - Clarity of Learning Objectives  
   - Notes: Teacher directed lesson, but gave tasks a role  
   - Notes: Good. Fast. (Score 6)

8. **Concept Development (CD)**  
   - Analysis and Reasoning  
   - Creating  
   - Integration  
   - Connections to the Real World  
   - Notes: Teacher directed lesson, but gave tasks a role  
   - Notes: Good. Fast. (Score 6)

9. **Quality of Feedback (QF)**  
   - Scaffolding  
   - Feedback Loops  
   - Promoting Thought Processes  
   - Providing Information  
   - Encouragement and Affirmation  
   - Notes: Teacher directed lesson, but gave tasks a role  
   - Notes: Good. Fast. (Score 6)

10. **Language Modeling (LM)**  
    - Frequent Conversation  
    - Open-Ended Questions  
    - Repetition and Extension  
    - Self- and Parallel Talk  
    - Advanced Language  
    - Notes: Teacher directed lesson, but gave tasks a role  
    - Notes: Good. Fast. (Score 6)
Appendix C

Current Development and Training Questionnaire

E. CURRENT DEVELOPMENT AND TRAINING

How often have you participated in the following school-related activities since the beginning of the academic year?

CHOOSE ONE NUMBER ON EACH LINE.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Never</th>
<th>Once a month or less</th>
<th>Two or three times a month</th>
<th>Once or twice a week</th>
<th>Three or four times a week</th>
<th>Daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Meeting with other teachers to discuss lesson planning</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2 Meeting with other teachers to discuss Curriculum development</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Next, please describe the availability and your use of the following professional development activities to provide ongoing, direct support teaching reading this academic year.

Choose only one response.

<table>
<thead>
<tr>
<th>Activity</th>
<th>I received this type of assistance this year</th>
<th>Available, but I did not receive</th>
<th>Not available at my school</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Coaching or mentoring by reading coach in programs, materials, or strategies</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4 Coaching or mentoring from fellow teacher</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5 Peer study group or collegial circle for group study</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6 Demonstrations in my classroom</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7 Observations of other teachers</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8 Interpretation of assessment data</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9 Using assessment data to determine topics that require additional instruction or practice</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

During the current school year, including the summer before this school year, in how many of each of the following types of professional development activities in reading have you participated?

Please count each activity only once. If you have participated in these activities, also indicate the total number of hours you spent in these activities.

10 Professional Development Workshop on Reading
   Number activities (If '0' skip to next item) ________
   Total Hours ________

11 Professional Development Workshop on Mathematics
   Number activities (If '0' skip to next item) ________
   Total Hours ________

12 Professional Development Workshop on Social and Emotional Development
   Number activities (If '0' skip to next item) ________
   Total Hours ________
# Appendix D

## Classroom Resources Scale - Variable List

<table>
<thead>
<tr>
<th>Classroom Resources Scale</th>
<th>Present</th>
<th>Quality</th>
<th>Quantity</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0 = No</td>
<td>0 = Low</td>
<td>1 = 1</td>
<td>0 = No</td>
</tr>
<tr>
<td></td>
<td>1 = Yes</td>
<td>1 = Med</td>
<td>2 = 2</td>
<td>1 = Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 = High</td>
<td>3 = 3+</td>
<td></td>
</tr>
</tbody>
</table>

### WORK SPACES
- Individual Desks
- Cluster Desks
- Floor Space for Whole Group
- Small Group Table Space

### Centers
- Writing
- Reading
- Mixed Literacy
- Dramatic Play
- Science
- Math

### LITERACY MATERIALS
- Fiction and Non-Fiction
  - Trade Books
  - Big Books
  - Leveled Readers
  - Basal Readers
  - Children’s Magazines
- Word Wall
- Literacy Chart/Poster
- Literacy Manipulatives
- Listening Center
- Alphabet Banner
  - Letters Only
  - Letters and Objects/Sounds
  - Print Tags on Objects

### CHILDREN’S WORK
- Art Displayed
- Writing Displayed
- Worksheets Displayed

### ORGANIZATION
- Behavior Incentives
- Class Rules
- Daily Schedule

### OTHER MATERIALS
- Television
- Computer(s)

*Black shading indicates that no score is computed for variable*
VITA
Lia Elaine Sandilos, M.Ed.
les207@psu.edu
215-603-9608

EDUCATION
The Pennsylvania State University
Graduate Student, Doctoral Candidate 9/07 – 8/12
Currently completing graduate studies in the
School Psychology Ph.D. Program (5th year)
Received M.Ed. (Spring 2010); Ph.D. (exp. August 2012)

PROFESSIONAL CERTIFICATION
Pennsylvania
Certified School Psychologist 6/11 - present

PUBLICATION
Assessment Scoring System Pre-K (CLASS Pre-K). The Journal of Early
Childhood and Infant Psychology, 7, 65-85.

EMPLOYMENT EXPERIENCES
Counseling or Referral Assistance (CORA) Service, Inc.
School Psychology Doctoral Intern 8/11 – 6/12
Currently conducting psychoeducational evaluations for non-public
and charter schools in the Philadelphia area

The Pennsylvania State University, CEDAR Mobile Clinic
Graduate Student Clinician 3/11 - 6/11
Conducted psychoeducational evaluations for local school
districts in need of extra school psychological services

The Pennsylvania State University, The Prevention Research Center
Graduate Research Assistant, Family Life Project 8/09 – 5/11
Collect data and conducted statistical analyses for longitudinal research project

PRACTICUM EXPERIENCE
State College Area School District
School Psychology Practicum Student 9/09 – 5/10
Conducted Response to Intervention (RTI) reading groups,
administered AIMSweb progress monitoring assessments,
completed psycho-educational evaluations of students in preschool,
elementary, and secondary grades

The Pennsylvania State University CEDAR School Psychology Clinic
Student Clinician and Supervisor 6/08 – 5/11
Conducted and supervised psycho-educational evaluations to
determine appropriate psychological and educational services