Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills

Andrew J. Mashburn, Robert C. Pianta,
Bridget K. Hamre, and Jason T. Downer
University of Virginia

Oscar A. Barbarin, Donna Bryant,
Margaret Burchinal, and Diane M. Early
University of North Carolina at Chapel Hill

Carollee Howes
University of California at Los Angeles

This study examined development of academic, language, and social skills among 4-year-olds in publicly supported prekindergarten (pre-K) programs in relation to 3 methods of measuring pre-K quality, which are as follows: (a) adherence to 9 standards of quality related to program infrastructure and design, (b) observations of the overall quality of classroom environments, and (c) observations of teachers' emotional and instructional interactions with children in classrooms. Participants were 2,439 children enrolled in 671 pre-K classrooms in 11 states. Adjusting for prior skill levels, child and family characteristics, program characteristics, and state, teachers' instructional interactions predicted academic and language skills and teachers' emotional interactions predicted teacher-reported social skills. Findings suggest that policies, program development, and professional development efforts that improve teacher–child interactions can facilitate children's school readiness.

Over the past four decades, the federal government and most states invested in providing public preschool programs for 3- and 4-year-old children. This trend began in 1965 with the establishment of Head Start, and by 2004–2005, 38 states offered one or more state-funded programs that served approximately 800,000 children across the nation and cost more than $2.8 billion (Barnett, Hustedt, Robin, & Schulman, 2005). Coinciding with the creation of public preschool programs was a fourfold increase in the prevalence of 4-year-olds who attended preschool over the past 40 years, with attendance rising from 17% in 1965 to 66% in 2002 (U.S. Census Bureau, 2002). The combination of increased enrollment, expansion of publicly funded preschool programs, and recognition of the unique role of early education experiences in the establishment of education success has led to a current state in which children's entry into the system of formal education, for all intents and purposes, starts for the majority of children in the United States at age 4 (Pianta, 2005).

Policies determine many features of preschool programs, such as who is eligible to attend, minimum qualifications of teachers, additional services that are available to children and their families, the amount and types of teacher professional development, and approaches to monitoring program and classroom quality (Pianta, Cox, & Snow, 2007). In the present study, using a data set that includes 11 states that serve approximately 80% of children in the United States who attend state prekindergarten (pre-K) programs for 4-year-olds, we examine children's academic, language, and social development during pre-K in relation to three commonly used methods of assessing program quality, which are as follows: features of program design and infrastructure, observations of overall classroom quality, and observations of teachers' interactions with children. The results of this analysis can provide direction for policies aimed toward program development efforts (Meisels, 2007) and to the recent emergence of quality rating systems (Mitchell, 2007) as a mechanism for program improvement and accountability.

Correspondence concerning this article should be addressed to Andrew J. Mashburn, Center for Advanced Study of Teaching and Learning, University of Virginia, 350 Old Ivy Way Suite 100, Charlottesville, VA 22903. Electronic mail may be sent to amashburn@virginia.edu.

High-Quality Pre-K

Numerous research studies document short- and long-term benefits of attending preschool, which has led to the creation and expansion of public programs nationwide (e.g., Barnett, 1993; Gormley, Gayer, Phillips, & Dawson, 2005; Lazar, Darlington, Murray, Royce, & Snipper, 1982; Magnuson, Ruhm, & Waldfogel, 2004; Reynolds, 2000; Reynolds, Temple, Robertson, & Mann, 2001; Schweinhart et al., 2005). Once programs are created, policymakers and program administrators must then establish policies that regulate the design and structure of these programs, so that they maximize benefits for children who attend. Nearly every state regulation pertaining to pre-K programs emphasizes the importance of providing high-quality services. Yet, despite all the attention directed toward high-quality pre-K, there is no singular definition for or method of evaluating pre-K quality (Lamb, 1998; Phillips & Howes, 1987; Vandell & Wolfe, 2000).

Theoretical definitions of preschool quality generally reflect two broad areas: features of program infrastructure or design and aspects of the classroom environments to which children are directly exposed as a result of enrollment (Lamb, 1998; Phillips & Howes, 1987; Vandell & Wolfe, 2000). Features of program infrastructure and design include program features targeted by regulation or financing, such as the nature and level of teacher training and experience, adoption of certain curricula, class size, child-to-teacher ratio, and whether or not the program offers additional services to children and their families. These features are viewed as foundational to the program and requisite components of providing effective educational services to young children (National Association for the Education of Young Children [NAEYC], 2005). Program quality has also been conceptualized and assessed in terms of children's direct experiences while they are enrolled in programs, such as the ways teachers organize routines, implement activities and lessons, and make interesting materials available to children, and the qualities of teachers' interactions with children (Hamre & Pianta, 2007; Harms, Clifford, & Cryer, 1998).

In research and program development (Meisels, 2007), including the more specific policy tool of quality rating systems (Mitchell, 2007), three approaches for assessing various aspects of pre-K quality have been commonly used. In many contexts, particularly state policy and professional organizations, pre-K quality is primarily defined in terms of whether programs adhere to recommended minimum standards related to programs' infrastructure

and design (such as class size or teacher qualifications), and the number of minimum standards to which a program adheres is considered a benchmark measure of program quality (Barnett et al., 2005). In contexts in which program improvement, program effects, and professional development are critical, classroom observations tend to be the metric of quality most often used. In some instances, the focus of observations is a composite indicator, reflecting physical safety, appropriate materials, and teacher–student interaction (Harms et al., 1998), whereas in other instances, the focus is solely on aspects of emotional and instructional interactions between teachers and children within classrooms (Pianta, La Paro, & Hamre, 2007).

Each metric of pre-K quality may serve different purposes in comprehensive approaches to program design, development, and improvement (Meisels, 2007). In the present study, we assess the extent to which each of the three forms of program quality assessment alluded to previously mentioned accounts for variation in development of skills among children attending state-funded pre-K programs. Importantly, we examine these features' predictive value in pre-K programs that had been operating for a period of several years and had achieved a certain level of "maturity" with respect to operations; thus, the results may not be as relevant for understanding the relative role of different forms of program quality in new or rapidly expanding programs. Next, we describe each of the three forms of quality metrics in more detail.

Minimum Standards for Program Infrastructure and Design

Professional organizations concerned with the well-being and appropriate education of young children offer recommendations for minimum standards of quality, most of which address features of program design and infrastructure. For example, the American Public Health Association and the American Academy of Pediatrics (APHA & AAP; 1992), the NAEYC (2005), and the National Institute for Early Education Research (NIEER; Barnett, Hustedt, Robin, & Schulman, 2004; Barnett et al., 2005) each advance a set of recommended standards for preschool programs, which have informed policymakers' and program administrators' decisions about how to design high-quality programs. The NAEYC (2005) recommends a comprehensive set of standards for early childcare programs that include descriptions of classroom characteristics that children directly experience, such as positive relationships in classrooms, a wide repertoire of teaching practices, developmentally

appropriate assessment practices, and learning environments rich with physical resources. Recommendations also include minimum standards related to teacher preparation, curricula, class size, and child-to-teacher ratio. Specifically, attaining accreditation from NAEYC requires that all teachers participate in professional development training, assistant teachers have at least a high school diploma or have passed the General Educational Development (GED) test, and programs meet increasing standards for teachers' levels of education that are being phased in between 2006 and 2020 (NAEYC, 2005). In addition, for classes serving 3-year-olds, the maximum class size is 18 children and the maximum child-to-teacher ratio is 9:1, and for classes serving 4-year-olds, the maximum class size is 20 children and the maximum child-to-teacher ratio is 10:1 (NAEYC, 2005). The view within the profession is that such standards contribute to better experiences and outcomes for children (Shonkoff & Phillips, 2000) and more organized, definable programs, and these metrics are often used for designing and implementing new or rapidly expanding programs intended to serve many children.

There is some evidence that recommendations offered by professional organizations such as those of NAEYC and the APHA and the AAP, when applied to the full range of childcare and early education programs, are associated with improved developmental outcomes for children. For example, in the National Institute of Child Health and Human Development (NICHD) Study of Early Child Care and Youth Development (NICHD Early Child Care Research Network [ECCRN], 1999), children at the ages of 6, 15, 24, and 36 months, who were enrolled in child care centers that met more standards recommended by the APHA and AAP (1992) regarding child-to-staff ratio, group size, caregiver training, and caregiver level of education, performed better on cognitive, language, and social competence measures compared to those enrolled in classes that met fewer of these standards. In a similar study, Howes (1990) reported evidence that a composite measure of quality in pre-K that comprised minimum standards for child-to-staff ratio, group size, caregiver training, and physical space was positively associated with children's adjustment in kindergarten. Thus, features of program design and infrastructure that are promoted by professional organizations can be important components of a regulatory system aimed at providing classroom capacities that contribute to improvements in children's learning and social adjustment.

A recent addition to these recommended standards comes from the NIEER, whose mission includes providing policymakers with information that promotes good education for 3- and 4-year-olds (Barnett et al., 2005). The NIEER-published *The State of Preschool Yearbooks* for 2003, 2004, and 2005 provide an overall summary of the status of state pre-K initiatives regarding accessibility, funding, and quality (Barnett et al., 2004). Program quality is operationally defined as whether programs adhere to the following 10 benchmarks that are considered by NIEER to be minimum standards needed for educationally effective preschool programs (Barnett et al., 2004): (a) teachers with a bachelor's degree; (b) teachers who have received specialized training in early childhood education (ECE) such as licensure/endorsement in the pre-K area or a degree or credential in early childhood, such as a child development associate's (CDA) degree; (c) teachers complete at least 15 hr per year of in-service training; (d) assistant teachers with a CDA or equivalent; (e) a comprehensive curriculum that covers domains of language/literacy, math, science, socioemotional skills, cognitive development, health, physical development, and social studies; (f) a maximum class size that is less than or equal to 20 children; (g) a child-to-teacher ratio of 10:1 or better; (h) at least one meal served each day; (i) vision, hearing, health screening/referral for children; and (j) at least one family support service that may include parent conferences, home visits, parenting support or training, referral to social services, and information relating to nutrition. In the 2005 state preschool yearbook published by NIEER (Barnett et al., 2005), the authors combined Standards 9 and 10 into "required screening/referral and support services," and added a new benchmark—whether the state monitored program quality. All these benchmarks are supported by a mixture of evidence and professional consensus, and they have been helpful policy shaping tools for state legislatures as they decide how to expand or construct a high-quality pre-K educational system.

In each edition of the NIEER state preschool yearbook, each state received a summary score indicating the number of benchmarks met or exceeded by their program(s), and the results were used to rank the quality of each state's pre-K initiative as indicated by the number of benchmarks each state's program(s) achieved (Barnett et al., 2004). In 2004 and 2005, Arkansas was the only state to meet or exceed each of the 10 benchmarks of quality, and in both years, Pennsylvania was rated as the state providing the lowest quality programs, according to these features of infrastructure and design. The authors concluded that most states lack adequate quality standards for their children and that states need to improve policies that enact higher quality standards (Barnett et al., 2004, 2005).

Observations of Overall Quality

The Early Childhood Environment Rating Scale (ECERS; Harms & Clifford, 1980) and its revised version (Early Childhood Environment Rating Scale – Revised [ECERS – R]; Harms et al., 1998) are the most commonly used comprehensive observational measures of quality of preschool classroom environments and have served as the standard measure in the field of early education for more than 25 years. The ECERS – R includes 36 items that measure the following dimensions of the classroom environment: space and furnishings, routines, language reasoning, activities, interactions, and program structure. A single, overall measure of the quality of an early childhood environment is computed as the average of 36 items, and scores range from 1 to 7, with 1 indicating *inadequate* quality, 3 indicating *minimal* quality, 5 indicating *good* quality, and 7 indicating *excellent* quality. Results from the ECERS and ECERS – R have been extensively used in research that has examined associations between preschool quality and children's development, and dozens of investigations have demonstrated an association between higher scores on ECERS observations and children's developmental outcomes (e.g., Byrant, Burchinal, Lau, & Sparling, 1994; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001). Results from the ECERS and ECERS – R have also been used to monitor the quality of program(s) and provide guidance for improving quality, and there is ample evidence suggesting that an assortment of program investments, technical assistance, and professional development efforts can be used to improve such scores (Frank Porter Graham Child Development Institute, 2005).

Observations of Teacher – Child Interactions

The Classroom Assessment Scoring System (CLASS; Pianta et al., 2007) is an observational measure of the quality of several dimensions of teacher – child interaction in classrooms. The theoretical framework for the CLASS (Hamre & Pianta, 2007) posits that the interactions that take place among teachers and students on a daily basis are the primary mechanisms through which children learn, and the CLASS observation system assesses different dimensions of these interactions within classrooms. These dimensions reflect social features of interactions (e.g., the extent to which teachers are sensitive and responsive to children's needs and cues) and instructional aspects of interactions (e.g., the extent to which teachers' behaviors promote concept development or scaffold children's performance of skills). Each dimension included in the CLASS is rated along a 1 – 7 scale, with 1 or 2 indicating *low quality*; 3, 4, or 5 indicating *mid-range of quality*; and 6 or 7 indicating *high quality*, and results provide a measure of two factors of the quality of classroom interactions—Emotional Support and Instructional Support (La Paro, Pianta, & Stuhlman, 2004).

Research Question

Preschool quality is a multifaceted, multilevel construct that includes a variety of program and classroom features, and a large body of literature exists in which correlates and consequences of these features have been evaluated, often separately in different studies that may focus on one or a few aspects of quality (e.g., Baydar & Brooks-Gunn, 1991; Bryant, Peisner-Feinberg, & Clifford, 1993; Bryant et al., 1994; Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Burchinal, Roberts, Nabors, & Bryant, 1996; Caughy, DiPietro, & Strobino, 1994; Dunn, 1993; Early et al., 2007; Gormley et al., 2005; Hagekull & Bohlin, 1995; Howes, 1997; Howes et al., 2008; NICHD ECCRN, 2002a, 2002b; Peisner-Feinberg & Burchinal, 1997; Phillips, Howes, & Whitebook, 1992). As a result of the different methods of conceptualizing and measuring quality across studies, there is a mixture of evidence, particularly concerning the magnitude of associations between specific features of quality and children's development (NICHD ECCRN & Duncan, 2003). This can be a concern not only for parties interested in development and improvement of existing programs but also for advocates and policymakers interested in program expansion and design.

In the present study, drawing from the largest study of pre-K programs conducted to date, we examined the extent to which children's development during pre-K was associated with each of three different approaches for evaluating quality of pre-K programs. In this large sample of operationally mature, state-funded pre-K programs, we examined associations between children's development of academic, language, and social skills during pre-K and (a) whether pre-K programs met nine standards of quality proposed by the NIEER (Barnett et al., 2004, 2005), (b) observations of the overall quality of the pre-K environment measured by the ECERS – R, and (c) observations of the quality of emotional and instructional interactions measured by the CLASS. The results may provide a basis for planning next steps for the further development and improvement for existing programs and a glimpse of future work and initiatives for emerging and new programs in this fast-growing sector of publicly supported education.

Method

Participants

Participants were 2,439 children enrolled in 671 pre-K classrooms in 11 states participating in two studies, which for the purposes of this article, used identical assessments: the National Center for Early Development and Learning's (NCEDL) Multi-State Study of Pre-Kindergarten (Multi-State Study) and the NCEDL–NIEER State-Wide Early Education Programs Study (SWEEP Study). The purpose of these studies was to describe pre-K programs in states with large state-funded programs that had been in operation for several years, and the 11 states included in these studies served approximately 80% of children in the United States who attended state pre-K programs. States were chosen specifically because their programs were operationally stable and mature, having been in existence for a number of years, and because they served a large number of the 4-year-old population in their states. This sampling strategy was used because emerging and new programs often suffer from an assortment of start-up challenges that could result in underestimating the potential positive effects of pre-K on children's outcomes (Clifford et al., 2005).

The Multi-State Study involved a stratified sampling of 40 pre-K sites within each of six states during 2001–2002. The SWEEP Study involved a stratified random sample of 100 state-funded pre-K programs within each of five additional states during 2003–2004. In both studies, within each pre-K site, one classroom was randomly selected to participate. In each participating classroom, teachers sent packets home with all children in their classes containing (a) a consent form describing the study, (b) a family contact sheet, and (c) a short demographic questionnaire. Parents returned these packets to the teacher, and on the first day of data collection, data collectors determined which of the children were eligible to participate. The average rate of consent for children in the Multi-State Study and SWEEP Study was 61% and 55%, respectively.

Children eligible for participation were those who (a) had parental consent, (b) met the age criteria for kindergarten eligibility the following year, (c) did not have an individualized education plan, and (d) spoke English or Spanish well enough to understand simple instructions. From that group, four children were randomly selected to participate, and whenever possible, two boys and two girls were selected from each classroom. Classes averaged 17 children, and this sample represents approximately 25% of the population of children who attended each class. The original study sample included 2,938 children enrolled in 704 classrooms; however, 499 children and 33 classrooms were excluded from the study for the following two reasons: (a) children participated in Spanish language version of these assessments at the end of pre-K, which did not measure constructs that were equivalent to those measured in the English version, or (b) children did not participate in a direct assessment of academic and language skills at the end of pre-K. Of the 2,439 children included in the study, teachers' ratings of social skills at the end of spring were missing for 132 children, and analyses examining associations between quality and children's development of social skills during pre-K included 2,307 children from 633 classrooms. Tables 1 and 2 present descriptive characteristics of children and classrooms participating in this study.

Table 1

Child and Family Characteristics, and Academic, Language, and Social Outcomes (n = 2,439)

	n	%	Missing	M	SD
Child characteristics					
Gender					
Boy	1,194	49			
Girl	1,245	51			
Race					
African American	516	21			
Latino	415	17			
White	1,127	46			
Other race	373	15			
Family characteristics					
Family income					
Poor	1,157	47			
Not poor	1,282	53			
Mother's education (years)	2,439			12.9	2.33
Assessment scores					
PPVT—pretest	2,127		312	94.2	15.0
PPVT—posttest	2,433		6	96.3	14.3
OWLS—pretest	2,113		326	91.6	13.1
OWLS—posttest	2,436		3	93.6	13.0
Rhyming—pretest	2,120		319	2.26	3.23
Rhyming—posttest	2,438		1	3.65	4.02
Applied problems—pretest	2,107		332	98.5	13.6
Applied problems—posttest	2,435		4	99.1	12.9
Letter naming—pretest	2,125		314	8.69	9.03
Letter naming—posttest	2,439		0	13.9	9.42
Social competence—pretest	2,029	278		3.52	0.76
Social competence—posttest	2,307		0	3.66	0.77
Problem behavior—pretest	2,028	279		1.49	0.52
Problem behavior—posttest	2,307		0	1.49	0.54

Note. OWLS = Oral and Written Language Scale; PPVT = Peabody Picture Vocabulary Test.

Table 2
Measures of Pre-K Quality (n = 671)

	n	Missing	% yes	M	SD
Program infrastructure and design					
Teacher has BA or higher degree	671	0	70		
Teacher has training in ECE/CD	671	0	60		
Teacher's aide has CDA or BA degree	567	104	20		
Class has 20 or fewer students	671	0	82		
Program uses comprehensive curriculum	546	125	57		
Child-to-teacher ratio is 10:1 or better	671	0	87		
Program serves meals	581	90	78		
Program provides family support services	578	93	81		
Program offers health services	574	97	47		
9-item NIEER index	398	274		5.89	1.50
Quality of the overall environment					
Overall quality (ECERS – R)	671	0		3.85	0.82
Quality of teacher – child interactions					
Emotional support (CLASS)	671	0		5.57	0.68
Instructional support (CLASS)	671	0		2.08	0.83

Note. CD = child development; CDA = child development associate; CLASS = Classroom Assessment Scoring System; ECE = early childhood education; ECERS – R = Early Childhood Environment Rating Scale – Revised; NIEER = National Institute for Early Education Research; pre-K = prekindergarten.

Measures

Features of program infrastructure and design: NIEER quality. The following nine characteristics of pre-K programs recommended by NIEER were measured by either teacher report or program administrator report on paper-and-pencil surveys: (a) lead teacher has a bachelor's degree; (b) lead teacher has received specialized training in ECE such as licensure/endorsement in the pre-K area or a degree or credential in early childhood, such as a CDA; (c) assistant teacher has a CDA or a bachelor's degree; (d) the class size is less than or equal to 20; (e) a comprehensive curriculum is used in the class, which was defined as whether a single curriculum was used that addressed all developmental domains (programs that used multiple curricula, no curriculum, locally developed curriculum, or a curriculum that did not address all developmental domains were coded as not comprehensive); (f) the child-to-teacher ratio is 10:1 or better; (g) at least one meal is served each day; (h) the program offers vision, hearing, health screening/referral for children; and (i) at least one family support service is provided, which may include parent confer-

ences, home visits, parenting support or training, referral to social services, and information relating to nutrition. Two quality indicators suggested by NIEER in 2004 and 2005 yearbooks—teachers participating in at least 15 hr of professional development activities and whether the state used systems of program monitoring—were not measured using the same criteria proposed by NIEER or were not assessed altogether and therefore were not included in this study. For each pre-K program, an index ranging from 0 to 9 was computed that indicates the number of benchmarks that each program achieved. Descriptive information for these nine features of program infrastructure and design are presented in Table 2. Regarding teacher characteristics, 70% of lead teachers had a bachelor's degree or more, 60% of lead teachers had specialized training in ECE, and 20% of assistant teachers had either a CDA or a bachelor's degree. Regarding classroom characteristics, 82% of classrooms had 20 or fewer students enrolled, 87% of classrooms met the minimum standard for child-to-teacher ratio (10:1), and 57% of classrooms used a comprehensive curriculum. Regarding program characteristics, 78% of programs served at least one meal per day, 81% provided family support services, and 47% offered health services. For the 0–9 index, the mean number of standards of program infrastructure and design programs met was 5.89, with a standard deviation of 1.5.

Quality of the overall environment. Overall quality of the environment of each participating pre-K classroom was rated by an observer using the ECERS – R (Harms et al., 1998). Each observer was trained using videos, live visits to classrooms, and guidance from an expert coder whose codes were presumed valid. The ECERS – R provides an overall measure of quality that is the mean of 36 items that measure various dimensions including space and furnishings, personal care routines, language reasoning, activities, interactions, and program structure, and scores range from 1 to 7, with 1 indicating *inadequate* quality, 3 indicating *minimal* quality, 5 indicating *good* quality, and 7 indicating *excellent* quality. For the Multi-State Study, the ECERS – R score included in the analysis is the mean score from two observations conducted in the fall and spring. There was not an overall mean difference in ECERS – R scores from fall (M = 3.77) and spring (M = 3.76). For the SWEEP Study, the ECERS – R score was derived from a single classroom observation that occurred in the spring.

Prior to data collection, each observer's reliability on the ECERS – R was tested during a live visit to a practice classroom with a staff member who was an expert in this measure. The mean weighted kappa

with the expert trainer was 0.69 ($SD = 0.07$, $n = 44$) on their final test. On average, 84% of data collector responses were exact matches or within one scale point of the expert trainers' responses. These levels of agreement are consistent with the use of the ECERS–R in other studies (e.g., Burchinal, Cryer, & Clifford, 2002; Burchinal et al., 2000; Peisner-Feinberg et al., 1999, 2001).

Quality of teacher–child interactions. The CLASS (Pianta et al., 2007) provides a measure of the quality of two global domains of teacher–child interaction within pre-K classes—instructional and emotional support. A trained observer rated the pre-K classroom and teacher on nine specific dimensions of interactions roughly every 30 min during an observation day, and observation days lasted from the time children arrived until they started nap (in full-day programs) or left for the day (in half-day programs). A classroom's score for each of the nine dimensions was computed as the average rating for that dimensions across the observation day.

Seven scales from the CLASS were used in this study. Positive climate reflects the enthusiasm, enjoyment, and respect displayed during interactions between the teacher and the children and among children. Negative climate is the degree to which the classroom has a negative emotional and social tone (displays of anger, aggression, and/or harshness), which is reverse coded for classroom averages. Teacher sensitivity is the extent to which teachers provide comfort, reassurance, and encouragement. Overcontrol reflects the extent to which classroom activities are rigidly structured or regimented, which is reverse coded for classroom averages. Behavior management concerns the teacher's ability to use effective methods to prevent and redirect children's misbehavior. Concept development considers the strategies teachers employ to promote children's higher order thinking skills and creativity through problem solving, integration, and instructional discussions. Finally, quality of feedback concerns the quality of verbal evaluation provided to children about their work, comments, and ideas.

For the SWEEP Study, the CLASS scores were derived from a classroom observation during one school day during spring. For the Multi-State Study, the CLASS scores for each construct are mean ratings from two observations conducted in the fall and spring. For the CLASS emotional quality scores, the difference between the mean spring score ($M = 5.33$) and the mean fall score ($M = 5.22$) was different from zero ($p = .02$). There was a larger difference on CLASS instructional quality scores between fall ($M = 2.47$) and spring ($M = 1.94$). Our decision to combine the

fall and spring CLASS scores for the Multi-State Study was to provide a more robust estimate of the quality that children directly experience during the year. To examine the impact of this decision on the results, we conducted these analyses using only the spring CLASS scores for the Multi-State Study and SWEEP Study, and the results were consistent with what is reported using the combined fall and spring scores.

Each dimension included in the CLASS is rated along a 1–7 scale, with 1 or 2 indicating *low quality*; 3, 4, or 5 indicating *mid-range of quality*; and 6 or 7 indicating *high quality*. A factor analysis of the CLASS yielded two factors (La Paro et al., 2004). The first factor is termed Emotional Support, and it is computed as a composite of Positive Climate, Negative Climate (reverse scored), Teacher Sensitivity, Overcontrol (reverse scored), and Behavior Management. The second factor is termed Instructional Support, and it is computed as a composite of Concept Development and Quality of Feedback. Prior to data collection, the observers' reliabilities were tested on the CLASS using videotapes of preschool classrooms and live visits to classrooms with one of the measures' authors. Data collectors' mean weighted kappa was 0.66 ($SD = 0.04$, $n = 43$) on their final test. On average, 86% of data collector responses were exactly the same or within one scale point of the expert's responses. This level of agreement was equal to or higher, on average, than that obtained in studies using these scales in kindergarten (Pianta et al., 2002) and first grade (NICHD ECCRN, 2002b) in which the scales were also related significantly to children's social and academic functioning. The internal consistencies for the emotional and instructional support factors within this study sample were 0.86 and 0.78, respectively. Table 2 presents means and standard deviations for the ECERS–R and CLASS scores included in this study.

Academic and language skills. Children in the study participated in direct assessments of academic and language skills at the beginning and end of the pre-K year. The assessment battery included assessments of children's receptive language, expressive language, rhyming, applied problem solving, and letter naming. The Peabody Picture Vocabulary Test–third edition was used to measure children's receptive vocabulary skills (Dunn & Dunn, 1997). During the administration of this test, a child is shown a card with four pictures, the child is read a word that corresponds to one of the pictures, and the child points to the picture that best represents the word. Raw scores were converted into standardized scores ($M = 100$, $SD = 15$) that reflect each child's performance relative to the expected performance of children in the population

who are the same age. The Peabody Picture Vocabulary Test – Revised demonstrates acceptable levels of test – retest reliability and split-half reliability and is strongly correlated with other measures of receptive language, achievement, and intelligence (Chow & McBride-Chang, 2003; Dunn & Dunn, 1997).

The Oral Expression Scale from the Oral and Written Language Scale (OWLS) was used to assess children's understanding and use of spoken language (Carrow-Woolfolk, 1995). During the administration of the test, the examiner reads a verbal stimulus out loud while the child looks at a card containing one or more pictures. Children respond orally to the stimulus by completing a sentence, answering a question or creating a new sentence or sentences. Raw scores were converted to standard scores ($M = 100$, $SD = 15$), and according to the measure's author, the test – retest reliability is 0.86 for children 4 – 5 years of age. The measure's author also reports correlations between the OWLS and other tests of achievement that range from 0.44 to 0.89.

The Woodcock – Johnson – III Test of Achievement, Sound Awareness, Rhyming Subtest was also used to assess children's rhyme awareness, which is a sub-category of phonological awareness (Woodcock, McGrew, & Mather, 2001). The maximum possible raw score on this scale is 17, and because the test was not norm referenced, raw scores were not converted to standardized scores, and scores ranged from 0 to 17. The reliability coefficients for the 4- to 5-year-old age-group range from .71 to .85 according to the measures' authors.

The Woodcock – Johnson – III Test of Achievement, Applied Problems subtest was used to measure children's applied problem-solving skills, including basic math skills such as counting, numeracy, comparisons, and word problems (Woodcock et al., 2001). Items include basic mathematical operations such as addition and subtraction to solve problems and basic skills such as telling time and reading a thermometer. For this subtest, raw scores were converted to standardized scores ($M = 100$, $SD = 15$) that reflect each child's performance relative to the expected performance of children in the population who are the same age. This subtest has demonstrated high internal consistency reliability, and it is positively correlated with other measures of academic achievement (Woodcock et al., 2001). The final assessment was a letter naming test that involved presenting children with 26 letters, and children were asked to name as many of the letters as possible (Bryant, Barbarin, & Aytch, 2001). Possible scores range from 0 to 26.

For children who spoke Spanish as their primary language at home, an English language screening test was administered (Pre-LAS; Duncan & De Avila, 1998), and children who did not pass this test in the fall and spring participated in assessments of receptive language, applied problem solving, and letter naming using Spanish versions of the assessments. The constructs measured by Spanish and English versions of these assessments were not equivalent, and as a result, these two versions of the assessments could not be analyzed simultaneously. Thus, the primary analyses included 2,439 children who completed English language versions of these assessments during the spring. A follow-up analysis reported in the Results section was conducted for 283 children who completed Spanish versions of the assessments during fall and spring.

Social skills. In fall and spring, teachers completed the Teacher – Child Rating Scale (TCRS; Hightower et al., 1986), a behavioral rating scale that assesses children's social competence and problem behaviors, Examples of social competence items include "participates in class discussions," "completes work," and "well-liked by classmates," and teachers used a 5-point scale (1 = *not at all*, 2 = *a little*, 3 = *moderately well*, 4 = *well*, and 5 = *very well*) to indicate how well the statements described the child. The social competence scale was computed as the mean of 20 items and had a Cronbach's alpha of .95. Examples of problem behavior items include "disruptive in class," "anxious," and "difficulty following directions." Teachers used a 5-point scale (1 = *not a problem*, 2 = *mild*, 3 = *moderate*, 4 = *serious*, 5 = *very serious problem*) to indicate how well the statements described the child. The problem behavior scale was computed as the mean of 18 items and had a Cronbach's alpha of .91. An evaluation of the normative and parametric characteristics of the TCRS is reported by Weissberg et al. (1987).

Control variables. The following child and family characteristics were included in this study as control variables: pretest scores, sex, race, mother's education, and whether the family is poor. In addition, we hypothesized effects associated with systematic differences across states in children's development; thus, 10 dummy variables were included in the analysis to account for differences in development across 11 participating states.

Analysis

This study involved a nested design that included approximately 4 children participating within each classroom, and hierarchical linear modeling (Raudenbush & Bryk, 2002) provided the conceptual framework for specifying two-level models that examined the associations between three forms of

pre-K quality (NIEER benchmarks, overall quality of the classroom environment, and teacher–child interactions) and individual-level child outcomes (posttest scores, after controlling for pretest scores, child and family characteristics, and state). Proc Mixed in SAS was used to specify the models derived from the following equations (Singer, 1998). In the first level of the two-level model (Equation 1), a spring assessment score (Y) for a child (i) who is in classroom (j) is a function of the mean posttest score for children in this class (β_{00}) after adjusting for pretest scores (β_{01}) and characteristics of children and families (β_{0n}), and the error term associated with this estimated mean (r_{ij}).

$$Y_{ij} = \beta_{00} + \beta_{01}(\text{pretest}) \\ + \beta_{0n}(\text{child and family}) + r_{ij}. \tag{1}$$

Equation 2 specifies in the second-level model that the adjusted mean posttest score for children in each classroom (β_{00}) is a function of the grand mean posttest score (γ_{00}), pre-K quality (γ_{01}), the state in which the pre-K program was located (γ_{02}), and the error term associated with this estimated mean (u_{0j}).

$$\beta_{00} = \gamma_{00} + \gamma_{01}(\text{quality}) + \gamma_{02}(\text{state}) + u_{0j}. \tag{2}$$

Model 1 included only the child, family, and state covariates. Four separate models were then analyzed that substituted different measures of quality in order to examine their independent associations with children's estimated academic, language, and social skills at the end of pre-K. Model 2 included all nine indicators of quality recommended by NIEER, and Model 3 included the 9-item NIEER quality index. Model 4 included the ECERS–R total score, and Model 5 included the two domains of observed teacher–child interaction derived from the CLASS—emotional and instructional support.

There were missing data for child and family characteristics and the measures of quality (Tables 1 and 2). Missing data were estimated using multiple imputation procedures in SAS, which created five complete data files. The multilevel analyses were conducted for each of the five imputed data files, and coefficients and standard errors resulting from each analysis were averaged to provide estimates of the associations between children's estimated developmental outcomes at the end of pre-K, and pretest scores, child and family characteristics, and the three forms of quality.

Results

Table 3 presents a matrix displaying bivariate correlations between each measure of program quality, including all nine indicators recommended by NIEER, the 9-item NIEER index, the overall quality of preschool classroom environments (ECERS–R), and the quality of teachers' emotional and instructional interactions (CLASS). Most correlations were significantly different from zero in part due to the large sample size. For the

Table 3
Bivariate Correlations Between Measures of Pre-K Quality

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Teacher has BA or higher degree	—												
2. Teacher has training in ECE/CD	−.32**	—											
3. Teacher's aide has CDA or BA degree	−.05	.03	—										
4. Class has 20 or fewer students	−.10**	.02	.02	—									
5. Program uses comprehensive curriculum	.18**	.05	.11*	.13**	—								
6. Child-to-teacher ratio is 10:1 or better	−.23**	.09*	.08	.41**	.10*	—							
7. Program serves meals	−.19**	.13**	.11*	−.04	.29**	−.06	—						
8. Program provides family support services	−.05	.09*	.08	−.01	.11*	.03	.20**	—					
9. Program offers health services	−.19**	.06	.15**	.11**	.14**	.14**	.32**	.27**	—				
10. 9-item NIEER index	−.05	.32**	.43**	.39**	.53**	.35**	.50**	.47**	.60**	—			
11. Overall quality (ECERS–R)	−.26**	.05	.07	.19**	.05	.32**	−.03	.13**	.12**	.11*	—		
12. Emotional support (CLASS)	−.02	−.04	−.04	.04	−.08	.08*	−.17**	−.05	−.09*	−.11*	.54**	—	
13. Instructional support (CLASS)	−.13**	.06	−.03	.17**	.02	.23**	.08*	.14**	.10*	.17**	.43**	.41**	—

Note. CD = child development; CDA = child development associate; CLASS = Classroom Assessment Scoring System; ECE = early childhood education; ECERS–R = Early Childhood Environment Rating Scale–Revised; NIEER = National Institute for Early Education Research; pre-K = prekindergarten.
*$p \leq .05$. **$p \leq .01$.

nine indicators recommended by NIEER included in Model 2, only one correlation identified two variables that shared more than 20% of the variance with each other ($r = .41$ for class size and child-to-teacher ratio). In addition, the observed measures of emotional and instructional interactions included in Model 5 had a bivariate correlation of $r = .41$.

Table 4 presents unstandardized coefficients and standard errors that indicate the magnitude of associations between posttest scores in language and academic skills during pre-K and pretest scores, child and family characteristics, and the various measures of quality. The first block of predictors in the analyses were control variables including pretest scores, characteristics of children and families, and the state in which programs were located (Model 1). Model 2 included the child, family, and state control variables and all nine indicators recommended by NIEER.

None of these nine indicators were associated positively with children's development during pre-K; however, children enrolled in pre-K classes with fewer than 20 students scored significantly lower on letter naming, and children enrolled in classes that served meals scored significantly lower on the measure of receptive language. Model 3 included the child, family, and state control variables and replaced the nine separate indicators with the NIEER quality index. The nine-item index was not significantly associated with children's academic and language skills at the end of pre-K.

Model 4 examined associations between a comprehensive measure of classroom quality (ECERS–R) and children's academic and language skills at the end of pre-K, after controlling for pretest scores, child and family characteristics, and state. Results indicate that the total ECERS–R score was associated

Table 4

Associations Between Quality and Children's Development of Language and Academic Skills During Pre-K (n = 2,439)

	PPVT		OWLS		Rhyming		Applied problems		Letter naming	
	B	SE	B	SE	B	SE	B	SE	B	SE
Model 1: Child characteristics										
Pretest	.66***	0.01	.72***	0.01	.77***	0.02	.62***	0.01	.79***	0.01
Boy	.28	0.36	−.49	0.34	−.27*	0.12	−.82*	0.35	−.91***	0.23
Black vs. White	−1.92**	0.70	−.19	0.69	−.19	0.24	−3.06***	0.76	.31	0.51
Latino vs. White	−3.75***	0.79	−1.19	0.80	−.75**	0.28	.95	0.90	−.66	0.57
Other race vs. White	.92	0.75	−.44	0.69	.18	0.27	1.06	0.79	.13	0.53
Mother's education	.26**	0.09	.24**	0.08	.16***	0.03	.41***	0.09	.22***	0.06
Family is poor	−1.39**	0.44	−1.16**	0.40	−.38*	0.15	−.93*	0.44	−.01	0.29
Model 2: Infrastructure and design										
Teacher has BA degree	.65	0.52	.24	0.50	.03	0.18	.98	0.58	.36	0.38
Teacher's field ECE/CD	.19	0.39	−.51	0.38	−.08	0.14	−.35	0.44	.05	0.29
Teacher's aide has CDA degree	.42	0.55	.51	0.46	−.07	0.19	−.44	0.52	−.02	0.37
Class size is 20	.16	0.56	−.63	0.58	.08	0.20	−.82	0.63	−1.23	0.42**
Comprehensive curriculum	−.27	0.46	−.09	0.42	−.26	0.17	.60	0.52	.44	0.35
Child-to-teacher ratio is 10:1	−.12	0.70	−.25	0.67	.12	0.25	−.64	0.77	−1.01	0.53
Program serves meals	−1.27*	0.58	.05	0.52	.13	0.21	−.40	0.61	−.24	0.43
Program has health services	.25	0.42	.02	0.48	.13	0.17	.20	0.54	.52	0.31
Program provides family services	−.08	0.52	−.73	0.50	.12	0.18	−.84	0.61	−.35	0.37
Model 3: NIEER index										
9-item index	−.04	0.14	−.15	0.13	.00	0.05	−.16	0.16	−.22	0.12
Model 4: Overall quality										
ECERS–R total score	.49	0.28	.79**	0.27	.13	0.10	.05	0.32	.27	0.22
Model 5: Teacher–child interactions										
CLASS emotional support	.23	0.32	.21	0.32	−.08	0.12	.40	0.36	.12	0.25
CLASS instructional support	.69*	0.30	1.07***	0.26	.33**	0.10	.66*	0.33	.43*	0.21

Note. Child characteristics in Model 1 are included as covariates in each of the subsequent models. In each model, state is entered as a Level 2 covariate. Values are unstandardized regression coefficients, with standard errors provided. CD = child development; CDA = child development associate; CLASS = Classroom Assessment Scoring System; ECE = Early Childhood Education; ECERS–R = Early Childhood Environment Rating Scale–Revised; NIEER = National Institute for Early Education Research; OWLS = Oral and Written Language Scale; pre-K = prekindergarten; PPVT = Peabody Picture Vocabulary Test.
*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

positively with expressive language skills. In the final model (Model 5), the total score from the ECERS–R was replaced with the two domains of teacher–child interaction measured by the CLASS. Quality of emotional interactions was not associated with development of academic and language skills during pre-K; however, quality of instructional interactions was positively associated with all five measures of academic and language development.

Table 5 presents unstandardized coefficients and standard errors that indicate the magnitude of the

Table 5

Associations Between Quality and Children's Development of Social Skills During Pre-K (n = 2,307)

	Social competence		Problem behaviors	
	B	SE	B	SE
Model 1: Child characteristics				
Pretest	.61***	0.02	.63***	0.02
Boy	−.13***	0.02	.10***	0.02
Black vs. White	−.07	0.05	.08*	0.04
Latino vs. White	.02	0.06	−.06	0.04
Other race vs. White	.02	0.05	−.00	0.04
Mother's education	.01	0.01	−.02***	0.00
Family is poor	−.06	0.03	.02	0.02
Model 2: Infrastructure and design				
Teacher has BA degree	.09*	0.04	−.02	0.03
Teacher's field ECE/CD	−.00	0.04	−.00	0.03
Teacher's aide has CDA degree	−.00	0.05	−.00	0.03
Class size is 20	.04	0.05	.01	0.04
Comprehensive curriculum	.01	0.04	−.01	0.03
Child-to-teacher ratio is 10:1	.01	0.07	.01	0.04
Program serves meals	.00	0.05	−.00	0.04
Program has health services	.07	0.03	−.03	0.02
Program provides family Services	.05	0.05	−.03	0.03
Model 3: NIEER index				
9-item index	.02	0.02	−.00	0.01
Model 4: Overall quality				
ECERS–R total score	.02	0.03	−.02	0.02
Model 5: Teacher–child interactions				
CLASS emotional support	.06*	0.03	−.05**	0.02
CLASS instructional support	.02	0.03	.01	0.02

Note. Child characteristics in Model 1 are included as covariates in each of the subsequent models. In each model, state is entered as a Level 2 covariate. Values are unstandardized regression coefficients, with standard errors provided. CD = child development; CDA = child development associate; CLASS = Classroom Assessment Scoring System; ECE = early childhood education; ECERS–R = Early Childhood Environment Rating Scale–Revised; NIEER = National Institute for Early Education Research; pre-K = prekindergarten.
*p ≤ .05. **p ≤ .01. ***p ≤ .001.

associations between social skills at the end of pre-K (social competence and problem behaviors) and social skills at pre-K entry, child and family characteristics, and the measures of quality. In Model 2, one of nine indicators recommended by NIEER—teachers with a bachelor's degree—was positively associated with children's development of social competence during pre-K; however, it is unclear whether teachers' level of education directly influenced children's social competence because it was teachers who provided ratings of children. Results from Model 3 indicate that the nine-item NIEER index was not related to development of social competence or problem behaviors. In addition, associations between children's development of social skills and observed overall quality (ECERS–R) and quality of instructional support (CLASS) were nonsignificant. Observed quality of emotional interactions measured by the CLASS was positively related to children's development of social competence and negatively related to children's development of problem behaviors during pre-K.

An identical set of analyses were conducted among 283 children who completed Spanish versions of receptive language, applied problem solving, and letter naming and whose social skills were rated by their teachers. Results indicate that the direction of the associations between the metrics of quality and development of language, academic, and social skills among Spanish-speaking children was similar to the effects found among children assessed in English. However, the magnitudes of these effects were not significantly different from zero, which is the result of the smaller sample size and the resulting larger standard errors.

Discussion

Assessments of three aspects of the quality of pre-K programs were examined in relation to predicting child outcomes: features of program design and infrastructure such as teacher qualifications, class size, and ratio; a comprehensive observation of overall quality of the classroom environment; and the nature and quality of teachers' emotional and instructional interactions with children in pre-K classrooms. Findings indicate that despite their relevance to discussions of program development and quality, none of the minimum standards recommended by NIEER, or the nine-item NIEER quality index, were consistently associated with measures of academic, language, and social development during pre-K, among a large sample of 4-year-old children who attended state-funded programs. For observations of the overall quality of pre-K environments as measured by the

ECERS–R, the only link to child outcomes was the positive association with children's development of expressive language skills. The measure of pre-K quality that was most consistently and strongly associated with children's development was dimensions of teacher–child interactions that children directly experienced in classrooms.

Specifically, using the CLASS as a standardized measure of emotional and instructional supports available for children in their interactions with teachers, higher quality instructional interactions were positively associated with all five measures of academic or language skills, and higher quality emotional interactions were associated with teachers' ratings of higher social competence and lower problem behaviors. These findings are consistent with theoretically driven expectations that the mechanisms through which pre-K directly impacts children's learning involve proximal processes that children experience in settings, for example, the quality of teacher–child interactions (Bronfenbrenner & Morris, 2006). These results demonstrating that the "value-added" of classroom settings is in part attributable to qualities of teacher–child interactions, found here for pre-K, are also consistent with findings for first-grade classrooms in which instructional and emotional supports reduce achievement gaps (Hamre & Pianta, 2005) and in relation to the quality of interactions in childcare settings that contribute to later achievement (NICHD ECCRN, 2002, 2005). For a variety of reasons, it is not surprising that young children in pre-K classrooms learn more when teachers interact with them in stimulating and emotionally supportive ways.

What was surprising, however, was the lack of significant associations with child outcomes for the other features of program quality that in other studies have been shown to be predictive of learning outcomes, particularly the indicators of observed overall classroom quality (Byrant et al., 1994; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001). In this study, the ECERS–R was positively associated only with children's development of expressive language, perhaps due to the emphasis of the measure in assessing the richness of language resources in classrooms. It may be that the lack of associations between the ECERS–R and other outcomes is the result of collapsing the multitude of features of preschool environments assessed by the ECERS–R (e.g., health and safety, space and furnishings, language reasoning) into a unidimensional comprehensive index that lacks some degree of precision in relation to gauging mechanisms responsible for development (Cassidy et al., 2005). A recent study

that decomposed the ECERS–R into two distinct factors—language/interactions and physical environment—indicates that the factor reflecting language inputs and interactions is indeed somewhat more predictive of children's development than the factor reflecting the physical environment (Howes et al., 2008). It should also be pointed out that in the past decade, the ECERS–R has been a guide for program development, which has resulted in investments in the features of the environment in early childhood settings that are assessed using the ECERS–R (Frank Porter Graham Child Development Institute, 2005). It may very well be that this process of program development, with the ECERS–R as a target, has indeed produced overall higher levels of program quality that reduces variation in quality between classrooms, thereby attenuating prediction of the ECERS–R to children's development.

The lack of significant associations between the NIEER benchmarks and children's development is also somewhat unexpected because it seems only reasonable that better trained teachers, smaller class sizes, and better curricula should result in greater developmental outcomes for children, and there are some indications from previous studies that indeed this is the case (Howes, 1990; NICHD ECCRN, 1999). However in the present instance, there are a number of potential explanations for the null findings for the associations between these features of pre-K program infrastructure and design and children's development. First, these results may be related to the types of pre-K programs that were included in this study. Pre-K classrooms were selected from 11 states with well-developed, large-scale programs, which have been the targets of public funding and regulations designed to deliver high-quality programs to children. As a result, there was little variation for many of the NIEER indicators because programs were designed to reflect this feature. If this study was conducted within pre-K programs that received little or no public funding, were held to less stringent standards of quality, and had a wider range of characteristics across programs, these features of programs' infrastructure and design may indeed be associated with developmental outcomes of children who attend. Second, the lack of significant results may be related to the manner in which the NIEER benchmarks of quality are measured. Each benchmark of quality is measured as a dichotomous variable that indicates whether a program did or did not meet each criterion, which reduces the actual range of teacher, classroom, and program characteristics. For example, the number of children enrolled in pre-K classes ranged from 3 to 40, with 85% of classes having

14–24 children enrolled. However, the NIEER benchmark for class size truncates this variability to a 0 or 1 corresponding with whether the program had more than 20 children or 20 or fewer children. Thus, the rather narrow range of program quality using the benchmarking type of metric promoted by NIEER clearly could have attenuated associations with children's development. Further, theory suggests that elements of program design and infrastructure, such as requiring teachers to have a bachelor's degree and a degree in ECE, are not directly associated with student outcomes, which received empirical support in another recent large-scale study of pre-K programs (Early et al., 2007). The bioecological model of development by Bronfenbrenner and Morris (2006) proposes that the pathways through which the exosystem (comprising features of program infrastructure and design) leads to children's development are not direct, as was examined in this study. Instead, classroom and program characteristics are proposed to be ecological conditions that moderate the impacts of classroom interactions on children's development; thus, small class sizes and child-to-teacher ratios may create conditions in classrooms in which high-quality social and instructional interactions have a stronger positive effect on children's development. In addition, teacher characteristics may not have a direct impact on children's development either; they may directly influence emotional and instructional interactions within classrooms, thereby indirectly influencing children's development. Support for this indirect path was demonstrated in the NICHD Study of Early Child Care and Youth Development, such that higher levels of teacher education were related to increased sensitivity in child care, and in turn better child outcomes (NICHD ECCRN, 2002a). Thus, teachers with higher credentials may indeed influence children's outcomes, to the extent that these qualifications lead to higher quality emotional and instructional interactions that children experience in classrooms.

Besides these potential moderating and indirect effects of features of program infrastructure and design on children's development, there may be other reasons to believe that certain features of program design may be beneficial in ways that are not based on their associations with children's development of academic, language, or social skills but rather for their value in institutionalizing the program in a larger social, cultural, or political context. For example, it has been suggested (and to some extent demonstrated) that requiring teachers to have a BA degree "professionalizes" the field of ECE (e.g., Bellm & Whitebook, 2003), and teachers may become better integrated into the pre-K-3 education system (Bogard & Takanishi, 2005). In addition, some design features such as whether programs serve meals, offer health screening and referrals, and provide resources for parents are components, by definition, of comprehensive early childhood programs. Although in this study, these program features were not related to improved academic, language, and social outcomes, these are important program components, particularly for the well-being of economically disadvantaged children and their families who otherwise may not have access to these types of supports. It is also important to note that although results indicate that program infrastructure and design features were not directly associated with development of children, this study did not assess whether *changes* in policies that improve these features are associated with children's outcomes. For example, state policies that improve or decrease these features of programs from 1 year to the next may result in measurable changes to both the quality of children's experiences in classrooms and children's outcomes.

Implications—Improving Quality Through Policies and Regulations

Results indicate that in state-funded pre-K programs serving 4-year olds, requiring teachers to have a college education or degrees in ECE and mandating small class sizes and child-to-teacher ratios may not be sufficient to ensure that children are learning in classrooms. Rather, these results confirm that for young children, learning occurs via interactions, and high-quality emotional and instructional interactions are the mechanisms through which pre-K programs transmit academic, language, and social competencies to children (Hamre & Pianta, 2007; Howes et al., 2008; Pianta, 2003). Thus, we argue that program policies and regulations aimed at improving the effectiveness of children's exposure to pre-K should focus more directly on improving interactions that children experience in classrooms. There are two avenues through which policies can directly improve the quality of emotional and instructional interactions in classrooms—teacher professional development and program monitoring systems.

In-service professional development. In-service professional development interventions are a potentially effective method to improve the quality of interactions in pre-K classes. Professional development practices have traditionally involved direct training of teachers through workshops (Birman, Desimone, Porter, & Garet, 2000), which tend to be ineffective because the content is vague, irrelevant,

or disconnected from classroom context, follow-up is limited, and methods involve passive learning techniques (Haymore-Sandholtz, 2002). In fact, teachers report that among a list of professional development activities, district-sponsored workshops are the least enjoyable and least useful to their everyday practices (Smylie, 1989).

No Child Left Behind describes a new vision for high-quality in-service professional development for teachers that is intensive, sustained, and classroom focused, and as a result, there has been a recent movement to design professional development activities that are active, collaborative, embedded within a classroom context, and part of school culture (Abdal-Haqq, 1995; Darling-Hammond & McLaughlin, 1995; Lieberman, 1995; Putnam & Borko, 1997; Richardson, 2003). For example, through mentoring, experienced teachers can focus their support and guidance on their less experienced colleagues' classroom interactions and practices (Fideler & Haselkorn, 1999). In addition, professional development techniques in which teachers receive consistent, nonevaluative feedback about their interactions with children have the potential to address the need for improving the quality of teacher–child interactions, particularly when these efforts are informed by a focus on the specific dimensions of interactions shown to produce gains for children (Pianta, 2005). Results from an evaluation study of such a teacher professional development program have found promise using this approach—pre-K teachers showed significant improvements in the quality of their instructional and emotional interactions during the course of the school year compared to teachers who did not participate in consultation directed toward improving classroom interactions (Pianta, Mashburn, Downer, Hamre, & Justice, 2007).

Program monitoring. Program monitoring is another approach to promote high-quality emotional and instructional interactions that children experience within pre-K programs and can include observing and rating features of early childhood classroom settings that children directly experience. Observational data may then be used in feedback to centers about their areas of strengths and weaknesses, as well as linked to resources to implement changes. There are a number of examples of using observations of the overall preschool environment to monitor and improve programs. For example, the Los Angeles Universal Preschool program (Los Angeles Universal Preschool, Quality Standards, 2005) uses a 5-star system for rating the quality of preschool programs, and one dimension of quality that is included in the rating system is based on hours of observing classroom process quality using the ECERS–R. Along the

1–7 ECERS–R rating scale, programs that achieve a mean score of 4 achieve a 3-star rating, programs that achieve a mean score of 5 achieve a 4-star rating, and programs that receive a 6 achieve a 5-star rating. Only programs that receive 3 or more stars are eligible to receive funding from the Los Angeles Universal Preschool program. In addition, programs that fall below three stars may receive support from a ''quality support coach'' who works directly with the preschool programs and offers support and guidance for improving quality, and teachers are eligible to participate in further education and training opportunities. The effectiveness of future program monitoring initiatives that use observations of preschool environments hinges upon providing guidance and resources that improve those features of classrooms that are most strongly associated with children's academic, language, and social development. Thus, the CLASS (Pianta et al., 2007) is an important tool for providing feedback and resources that improve emotional and instructional interactions, which in turn, promote children's development.

Study Limitations

Our analysis of the associations between quality in state-funded pre-K programs and children's development of school-related competencies found consistent evidence that it is the quality of classroom interactions that is directly associated with children's acquisition of academic, language, and social skills. However, there are four limitations of this study that limit the strength of our conclusions. First, the study did not measure two of the indicators of quality that were included in the NIEER checklist—whether teachers participated in 15 or more hours of professional development training and whether there was a system of program monitoring in place. As discussed previously, professional development and program monitoring have the potential to directly improve the quality of interactions children experience, and in turn, children's development; however, due to limitations of the data, we were not able to test whether programs that adhered to these minimum standards resulted in improved developmental outcomes in academic, language, or social skills for children who attended.

A second limitation concerns the generalizability of findings to other types of pre-K programs. As described earlier in the Discussion section, programs in this study were selected from 11 states with well-developed, large-scale programs that serve 4-year-olds, which have been the targets of public funding and regulations designed to deliver high-quality

programs to children. However, many other current forms of pre-K—private programs, programs that serve children younger than 4 years, or programs in states that are earlier in the process of developing their programs—have not received an equal investment of public resources or the same intense focus on program regulations, and across these programs, there is wide variability in features of the programs' infrastructure and design. As a result, findings about the associations between quality and development may be different in these settings.

A third limitation concerns the extent to which the children participating in the study represent the population of children served by these public pre-K programs. As described in the Method section, the rate of consent within each participating class was 61% in the Multi-State Study and 55% in the SWEEP Study. As a result, children who were selected to participate in the study may differ from the children whose parents did not provide consent for their child to participate. It is impossible to know the ways in which the study children differ from nonstudy children because demographic data were not available from nonconsenting parents; however, it may be the case that nonconsenting parents have children who experience greater risks at home (higher levels of poverty, less likely to speak English). The extent to which this is true may result in an underestimation of the effects of each measure of quality on children's development during pre-K because there is evidence indicating that higher quality pre-K has a stronger positive influence on children who experience social and economic risks (Baydar & Brooks-Gunn, 1991; Bryant et al., 1994; Bryant et al., 1993; Burchinal et al., 2000; Caughy et al., 1994; Hagekull & Bohlin, 1995; Peisner-Feinberg & Burchinal, 1997).

A final and fourth limitation of this study concerns the relatively small magnitudes of these effects on children's development of academic, language, and social skills, even when significant. For example, on the three standardized academic and language assessments that have a mean of 100 and standard deviation of 15, a 1-point increase in instructional quality was associated with posttest scores that were between 0.66 and 1.06 points higher, which were significantly different from zero as a result of the large sample size included in the study. There are some important considerations when interpreting these relatively small effects of emotional and instructional interactions on children's developmental outcomes. First, the instructional quality rating scale had low mean and narrow distribution ($M = 2.08$, $SD = 0.83$), such that most classrooms were characterized by relatively low levels of instructional quality in part

due to the operational definition of instructional quality that requires evidence of instructional interactions by teachers that includes feedback and effective elaboration of concepts. Assuming that the linear trend between instructional quality and estimated posttest scores continues along the entire rating scale, large differences in instructional quality would have incrementally greater effects on children's outcomes. To the extent that professional development and program monitoring interventions can improve instructional interactions in classrooms so that they achieve quality standards at the high end of this rating scale, the magnitude of these effects of instructional quality on children's development of academic and language skills may, in fact, be quite large. A second consideration when interpreting these relatively small effects of process quality on children's development is that small differences in children's competencies attributable to quality could have a proportionately greater impact on the population over time. This is particularly true among young children, for whom small increases in competencies early in children's academic careers may provide an early boost in their patterns of development that result in improvements in children's long-term academic trajectories (Hamre & Pianta, 2005).

Conclusions

Much of the policy-related discourse about pre-K program quality concerns minimum standards related to infrastructure and design features of pre-K programs. These features of programs may produce certain benefits such as improving children's health and family support, aligning pre-K programs with later grades, influencing the quality of social and instructional interactions children experience within classes, and establishing a design space wherein high-quality emotional and instructional interactions are more likely to lead to better outcomes for children. Thus, decisions to require more stringent standards for teachers related to their field of study or level of education may indeed lead to better outcomes for children, if these qualifications lead to higher quality emotional and instructional interactions that children experience in classrooms. In addition, decisions to reduce class sizes and child-to-teacher ratios may create conditions in classrooms in which high-quality social and instructional interactions have a stronger positive effect on children's development.

To the extent that the criterion for deciding where to invest resources is the features of state pre-K programs that are directly associated with children's development of academic, language, and social

competencies, then policies and practices should focus on promoting high-quality instructional and emotional interactions that children directly experience within classrooms. Direct efforts such as mentoring, consultee-centered consultation, and program monitoring and feedback, when informed by a focus on the types of emotional and instructional interactions in classrooms that are theoretically and empirically associated with children's development, may be the most effective and direct avenues to improve the outcomes of children who attend state pre-K programs.

References

Abdal-Haqq, I. (1995). *Making time for teacher professional development* (Digest 95–4). Washington, DC: ERIC Clearinghouse on Teaching and Teacher Education.

American Public Health Association & the American Academy of Pediatrics. (1992). *Caring for our children: National health and safety performance standards: Standards for out of home child care programs*. Washington, DC: American Public Health Association.

Barnett, W. S. (1993). Benefit-cost analysis of preschool education: Findings from a 25-year follow-up. *American Journal of Orthopsychiatry, 63*, 500–508.

Barnett, W. S., Hustedt, J. T., Robin, K. B., & Schulman, K. L. (2004). *The state of preschool: 2004 state preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.

Barnett, W. S., Hustedt, J. T., Robin, K. B., & Schulman, K. L. (2005). *The state of preschool: 2005 state preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research, Rutgers University.

Baydar, N., & Brooks-Gunn, J. (1991). Effects of maternal employment and child-care arrangements on preschoolers' cognitive and behavioral outcomes: Evidence from the children of the National Longitudinal Survey of Youth. *Developmental Psychology, 27*, 932–945.

Bellm, D., & Whitebook, M. (2003). *Universal preschool in California: An overview of workforce issues*. Berkeley, CA: Center for Childcare Employment.

Birman, B. F., Desimone, L., Porter, A. C., & Garet, M. S. (2000). Designing professional development that works. *Educational Leadership, 57*(8), 1–8.

Bogard, K., & Takanishi, R. (2005). PK-3: An aligned and coordinated approach to education for children 3 to 8 years old. *SRCD Social Policy Reports, 19*(3), 1–23.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner (Ed.) *Handbook of child development: Vol. 1. Theoretical models of human development* (6th ed., pp. 793–828). Hoboken, NJ: Wiley.

Bryant, D., Barbarin, O., & Aytch, L. S. (2001). *Naming letters* (English and Spanish versions). Unpublished measure, University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Center.

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

Bryant, D., Peisner-Feinberg, E., & Clifford, D. (1993). *Evaluation of public preschool programs in North Carolina: Final report*. Chapel Hill, NC: Frank Porter Graham Child Development Center.

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

Burchinal, M. R., Peisner-Feinberg, E., Bryant, D. M., & Clifford, R. (2000). Children's social and cognitive development and child care quality: Testing for differential associations related to poverty, gender, or ethnicity. *Journal of Applied Developmental Sciences, 4*, 149–165.

Burchinal, M. R., Roberts, J. E., Nabors, L. A., & Bryant, D. M. (1996). Quality of center child care and infant cognitive and language development. *Child Development, 67*, 606–620.

Carrow-Woolfolk, E. (1995). *Oral and Written Language Scales*. Circle Pines, MN: American Guidance Service.

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

Caughy, M., DiPietro, J., & Strobino, D. (1994). Day-care participation as a protective factor in the cognitive development of low-income children. *Child Development, 65*, 457–471.

Chow, B. W., & McBride-Chang, C. (2003). Promoting language and literacy development through parent-child reading in Hong Kong preschoolers. *Early Education and Development, 14*, 233–248.

Clifford, R. M., Barbarin, O., Chang, F., Early, D. M., Bryant, D., Howes, C., et al. (2005). What is pre-kindergarten? Characteristics of public prekindergarten programs. *Applied Developmental Science, 9*, 126–143.

Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan, 76*, 597–604.

Duncan, S., & De Avila, E. (1998). *Pre-LAS 2000*. Monterrey, CA: CTB/McGraw-Hill.

Dunn, L. (1993). Ratio and group size in day care programs. *Child Youth Care Forum, 22*, 193–226.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test–Revised* (3rd ed.). Circle Pines, MN: American Guidance Services.

Early, D., Maxwell, K., Burchinal, M., Alva, S., Bender, R., et al. (2007). Teacher education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development, 78*, 558–580.

Fideler, E., & Haselkorn, D. (1999). *Learning the ropes: Urban teacher induction programs and practices in the United States*. Belmont, MA: Recruiting New Teachers.

Frank Porter Graham Child Development Institute. (2005). *Environment rating scales: A directory of projects that use the*

rating scales. Retrieved November 11, 2005, from http://www.fpg.unc.edu/~ecers/

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology, 41*, 872–884.

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

Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first grade classroom make a difference for children at risk of school failure? *Child Development, 76*, 949–967.

Hamre, B. K., & Pianta, R. C. (2007). Learning opportunities in preschool and early elementary classrooms. In R. C. Pianta, M. J. Cox, & K. Snow (Eds.), *School readiness, early learning and the transition to kindergarten*. (pp. 49–84). Baltimore: Brookes.

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

Harms, T., Clifford, R. M., & Cryer, D. (1998). *The early childhood environment rating scale* (Rev. ed.). New York: Teachers College Press.

Haymore-Sandholtz, J. (2002). Inservice training or professional development: Contrasting opportunities in a school/university partnership. *Teaching and Teacher Education, 18*, 815–830.

Hightower, A. D., Work, W. C., Cowen, E. L., Lotczewski, B. S., Spinnell, A. P., Guare, J. C., et al. (1986). The Teacher-Rating Scale: A brief objective measure of elementary children's school problem behaviors and competencies. *School Psychology Review, 15*, 393–409.

Howes, C. (1990). Can the age of entry into child care and the quality of child care predict adjustment in kindergarten? *Developmental Psychology, 26*, 292–303.

Howes, C. (1997). Children's experiences in center-based child care as a function of caregiver background and adult:child ratio. *Merrill-Palmer Quarterly, 43*, 404–425.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford R. M., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly, 23*, 27–50.

Lamb, M. (1998). Nonparental child care: Context, quality, correlates, and consequences. In W. Damon (Series Ed.) & I. E. Spiegel & K. A. Renniger (Vol. Eds.) *Handbook of child psychology: Child psychology in practice* (5th ed., pp. 73–134). New York: Wiley.

La Paro, L. M., Pianta, R. C., & Stuhlman, M. (2004). Classroom Assessment Scoring System (CLASS). Findings from the pre-K year. *Elementary School Journal, 105*, 409–426.

Lazar, I., Darlington, R., Murray, H., Royce, J., & Snipper, A. (1982). Lasting effects of early education: A report from the consortium for longitudinal studies. *Monograph of the Society for Research in Child Development 47*(2–3, Serial No. 195).

Lieberman, A. (1995). Practices that support teacher development. *Phi Delta Kappan, 76*, 591–596.

Los Angeles Universal Preschool, Quality Standards. (2005). *Abbreviated LAUP 5-Star Quality Rating Scale for Centers*. Retrieved November 11, 2005, from http://www.laup.net/downloads/CenterStandards.star.pdf

Magnuson, K. A., Ruhm, C. J., & Waldfogel, J. (2004). *Does prekindergarten improve school preparation and performance?* NBER Working Paper W10452. Available at http://ssm.com/abstract=538442.

Meisels, S. J. (2007). Accountability in early childhood: No easy answers. In R. C. Pianta, M. J. Cox, & K. Snow (Eds.), *School readiness, early learning and the transition to kindergarten*. (pp. 31–47). Baltimore: Brookes.

Mitchell, A. (2007, February). *Financing state supported early education programs*. Symposium conducted at the leadership meeting of the National Center for Research in Early Childhood Education, Washington, DC.

National Association for the Education of Young Children. (2005). *Position statements of NAEYC*. Retrieved November 11, 2005, from http://www.naeyc.org/about/positions.asp

National Institute of Child Health and Human Development, Early Child Care Research Network. (1999). Child outcomes when child care center classes meet recommended standards for quality. *American Journal of Public Health, 89*, 1072–1077.

National Institute of Child Health and Human Development, Early Child Care Research Network. (2002a). Child-care structure-process-outcome: Direct and indirect effects of child-care quality on young children's development. *American Psychological Society, 13*, 199–206.

National Institute of Child Health and Human Development, Early Child Care Research Network. (2002b). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *Elementary School Journal, 102*, 367–387.

National Institute of Child Health and Human Development, Early Child Care Research Network. (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *Elementary School Journal, 105*, 305–323.

National Institute of Child Health and Human Development, Early Child Care Research Network, & Duncan, G. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development, 74*, 1454–1475.

Peisner-Feinberg, E., & Burchinal, M. (1997). Relations between preschool children's child-care experiences and concurrent development: The cost, quality, and outcomes study. *Merrill-Palmer Quarterly, 43*, 451–477.

Peisner-Feinberg, E., Burchinal, M., Clifford, R., Culkin, M., Howes, C., Kagan, S., et al. (1999). *The children of the cost, quality, and outcomes study go to school: Executive summary*. Chapel Hill: University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Center.

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

Pianta, R. C. (2003). *Standardized observations from pre-K to 3rd grade: A mechanism for improving access to high quality classroom experiences and practices during the P-3 years.* Working paper, Foundation for Child Development.

Pianta, R. C. (2005). A new elementary school for American children. *SRCD Social Policy Report*, *19*(3), 4–5.

Pianta, R. C., Cox, M. J., & Snow, K. (2007). *School readiness, early learning and the transition to kindergarten.* Baltimore: Brookes.

Pianta, R. C., La Paro, K. M., Hamre, B. K (2007). *Classroom Assessment Scoring System—CLASS.* Baltimore: Brookes.

Pianta, R., Mashburn, A., Downer, J., Hamre, B., & Justice L. (2007). *Effects of Web-mediated professional development resources on teacher-child interactions in pre-kindergarten classes.* Unpublished manuscript.

Phillips, D., & Howes, C. (1987). Indicators of quality in child care: Review of the research. In D. Phillips (Ed.), *Quality in child care: What does the research tell us?* (pp. 1–19). Washington, DC: National Association for the Education of Young Children.

Phillips, D., Howes, C., & Whitebook, M. (1992). The social policy context of child care: Effects on quality. *American Journal of Community Psychology, 20*, 25–51.

Putnam, R. T., & Borko, H. (1997). Teacher learning: Implications of new views of cognition. In B. J. Biddle, T. L. Good, & I. F. Goodson (Eds.), *The international handbook of teachers and teaching* (Vol. 2, pp. 1223–1296). Dordrecht, the Netherlands: Kluwer.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Reynolds, A. J. (2000). *Success in early intervention: The Chicago Child-Parent Centers.* Lincoln: University of Nebraska Press.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement. *Journal of the American Medical Association*, *285*, 2339–2346.

Richardson, V. (2003). The dilemmas of professional development. *Phi Delta Kappan*, *84*, 401–406.

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool study through age 40* (Monographs of the High/Scope Educational Research Foundation, *14*). Ypsilanti, MI: High/Scope Press.

Shonkoff, J., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods.* Washington, DC: National Academy of Sciences.

Singer, J. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *24*, 323–355.

Smylie, M. A. (1989). Teachers' view of the effectiveness of sources of learning to teach. *Elementary School Journal*, *89*, 543–558.

U.S. Census Bureau. (2002). *Census of population survey* Available at http://www.census.gov

Vandell, D. L., & Wolfe, B. (2000). *Child care quality: Does it matter and does it need to be improved?* Institute for Research on Child Poverty, University of Wisconsin–Madison.

Weissberg, R. P., Cowen, E. L., Loyyczewski, B. S., Boike, M. F., Orara, N. A., Stalonas, P., et al. (1987). Teacher ratings of children's problem and competence behaviors: Normative and parametric characteristics. *American Journal of Community Psychology, 15*, 387–401.

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson III.* Itasca, IL: Riverside.