Early Childhood Research Quarterly 25 (2010) 166–176

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs

Margaret Burchinal[a,*], Nathan Vandergrift[b], Robert Pianta[c], Andrew Mashburn[c]

[a] FPG Child Development Institute, CB #8185, 521 S. Greensboro, University of North Carolina, Chapel Hill, NC 27599-8185, United States
[b] Duke University, United States
[c] University of Virginia, United States

ARTICLE INFO

Article history:
Received 27 June 2008
Received in revised form
29 September 2009
Accepted 16 October 2009

Keywords:
Pre-kindergarten
Quality-outcome associations
Threshold analysis
Low-income children

ABSTRACT

Over the past five decades, the federal government and most states have invested heavily in providing publicly-funded child care and early education opportunities for 3- and 4-year-old children from low-income families. Policy makers and parents want to identify *the level* or *threshold* in quality of teacher–child interaction and intentional instruction related to better child outcomes to most efficiently use child care to improve school readiness. Academic and social outcomes for children from low-income families were predicted from measures of teacher–child interactions and instructional quality in a spline regression analysis of data from an 11-state pre-kindergarten evaluation. Findings suggested that the quality of teacher–child interactions was a stronger predictor of higher social competence and lower levels of behavior problems in higher than in lower quality classrooms. Further, findings suggested that quality of instruction was related to language, read and math skills more strongly in higher quality than in lower quality classrooms. These findings suggest that high-quality classrooms may be necessary to improve social and academic outcomes in pre-kindergarten programs for low-income children.

1. Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs

Over the past two decades, most states have invested heavily in providing publicly-funded child care and early education opportunities for 3- and 4-year-old children. States invested in public pre-kindergarten programs to promote academic success of children, especially children from low-income families. These programs grew rapidly so that by 2007, 38 states offered one or more state-funded programs that served approximately 1.1 million children across the nation and cost over 2.8 billion dollars (Barnett, Hustedt, Friedman, Boyd, & Ainsworth, 2008). This increase in publicly-funded child care has been accompanied by a greater focus on the quality of that care based on extensive evidence linking child care quality and children's academic and social development, especially for low-income children (c.f., Lamb, 1998; Vandell, 2004). While most pre-kindergarten legislation, policy, or program design requirements require programs be of high quality, there is considerable interest in refining those guidelines through determining whether there are thresholds above or below which the association between child care quality and child outcomes is stronger.

Numerous experimental and observational research studies document short-term and long-term benefits of attending pre-school (Gormley, Gayer, Phillips, & Dawson, 2005; Lazar, Darlington, Murray, Royce, & Snipper, 1982; Magnuson,

* Corresponding author. Tel.: +1 919 966 5059; fax: +1 919 962 5771.
E-mail address: burchinal@unc.edu (M. Burchinal).

0885-2006/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ecresq.2009.10.004

Meyers, Ruhm, & Waldfogel, 2004; Reynolds, Temple, Robertson, & Mann, 2001; Schweinhart et al., 2005). Early intervention projects such as Perry Preschool/High Scope (Schweinhart et al., 2005; Nores, Barnett, Belfield, & Schweinhart, 2005) and the Abecedarian Project (Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002) demonstrated that sustained high-quality child care experiences can improve cognitive and social outcomes for children from low-income families. These improved outcomes translated into lifelong savings in terms of reduced criminal activities and increased education and occupational opportunities with estimated cost-benefits associated with child care programs ranging from $2.50 per $1 spent on Abecedarian Project children to $12.06 per $1 spent on Perry Preschool children (Nores et al., 2005).

These early intervention programs and other child care research (Lamb, 1998) led to the funding of large federal child care programs like Head Start and state programs like public pre-kindergartens in an attempt to improve school readiness skills of low-income children. Evaluations of public pre-kindergarten programs (Gormley et al., 2005; Howes et al., 2008; Magnuson et al., 2004; Mashburn et al., 2008) also indicated that low-income children in these programs made substantial gains in language, academic, and social skills. Several studies indicated that these gains tended to be larger when programs were of higher quality (Howes et al., 2008; Mashburn et al., 2008). Studies of both low-income and middle-income children suggest that high-quality community child care experiences promote short and long-term cognitive, academic, and social outcomes (NICHD ECCRN, 1999, 2000, 2005; Peisner-Feinberg et al., 2001), but children from low-income families or families with other social risk factors showed larger gains in some studies of both community care (Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Burchinal, Roberts, Zeisel, Hennon, & Hooper, 2006; Peisner-Feinberg et al., 2001; Votruba-Drzal, Coley, & Chase-Lansdale, 2004) and pre-kindergartens (Gormley et al., 2005). Accordingly, most state regulation pertaining to pre-k programs emphasizes the importance of providing high-quality services. Yet, despite all of the evidence that the frequent positive and responsive interactions between teachers and children (Peisner-Feinberg et al., 2001; Lamb, 1998; NICHD ECCRN, 2005; Vandell, 2004) and intentional instruction (Mashburn et al., 2008) are key components of quality, little empirical work is available to support the kind of cut-offs or thresholds for levels of quality that can be of interest to policy makers or consumers.

Policy makers and parents want to identify *the level* or *threshold* in quality of teacher–child interaction related to better child outcomes to use in public policy regulations (Tout et al., 2009). Most of the literature has examined linear associations, yielding findings that higher quality is better and lower quality is worse (Vandell, 2004), but identification of thresholds in the association between quality and child outcomes has been a goal of researchers and policy makers for several reasons. A primary goal has been to identify levels in the association between quality and child outcomes at which the linear association begins to asymptote or level off, above or below which there is little evidence of increases in learning associated with increased in quality. A threshold that indicated that the quality-outcome association levels off above a given level of quality would suggest that policies should focus on improving quality up to that threshold level, but improving quality above that point may not be necessary for improving child outcomes. Policy to address this goal would invest in lower or average quality classrooms while leaving classrooms with quality scores above the threshold alone. In contrast, it is possible a threshold could define the minimum level at which a positive association between quality and outcomes is observed. In this scenario, there may be no detected relation between quality and outcome gains until quality reached a certain point on the scale; in other words, learning did not take place until classrooms demonstrated a minimal level and after that minimum, gains in learning increased as quality increased. This form of threshold effect would suggest that it is especially important to ensure that children experience at least the minimum level of quality child care in order for those experiences to be related to improved child outcomes. It would point perhaps to not allowing vouchers to pay for care that was below the threshold, while also incentivizing teachers above the threshold to continue to improve. One can see how the examination of thresholds may have considerable implications for the efficient and effective expenditure of funds, not only in relation to providing access to quality of a certain minimal level, but also in the targeting of funds to improve quality.

Several studies have tried to identify thresholds. The NICHD Study of Early Child Care failed to detect thresholds in several papers examining cognitive and social outcomes (NICHD ECCRN & Duncan, 2003; NICHD ECCRN, 2006). These studies added both linear and quadratic quality terms, but failed to detect curvature in the regression of child outcomes on quality. Subsequent analyses by McCartney, Dearing, Beck, and Bub, 2007 found that exposure to high-quality child care quality appeared to reduce the achievement gap between families with more and less income on school readiness. Using a median split, they found that income was a much stronger predictor of outcomes among children in no child care or lower quality child care. However, none of the previous work has systematically tested for threshold effects using the full range of the quality indicators and analytic methods that allow for different quality slopes in specific ranges of the quality scale.

The goal of the current study is to analyze the data for children from low-income families from an 11-state evaluation of public pre-kindergarten programs. We intentionally chose to focus on children from low-income families because it is this group of children that are largely the target of policy-related decisions in terms of program access and improvement, and for the most part, pre-kindergarten programs have been established to address their developmental and educational needs. The current examination of pre-k quality and children's learning gains follows from several reports (Howes et al., 2008; Mashburn et al., 2008) from the same project that demonstrated children showed nontrivial gains when they attended public pre-kindergartens and those gains were modestly larger when teachers provided higher quality instruction and more positive and responsive interactions with the students in the classroom. Those analyses focused only on linear associations between gains over time and observed child care quality. In this study we test for thresholds in the association between quality and child outcomes.

Table 1
Descriptive statistics pre-imputation.

Variable	Fall			Spring		
	N	*M*	*SD*	*N*	*M*	*SD*
PPVT-III	1140	89.41	13.72	1197	91.98	13.40
OWLS	1130	88.41	12.16	1195	90.39	12.08
WJ-III Applied problems	1124	397.69	19.64	1196	408.83	18.39
WJ-III Letter-word	731	322.36	22.98	774	337.13	24.92
Hightower competence	1385	3.40	0.76	1402	3.57	0.79
Hightower behavior	1386	1.56	0.55	1400	1.54	0.57
Maternal education	1588	11.77	1.90			
Proportion male	1605	0.49	0.50			
Proportion Latino	1583	0.36	0.48			
Proportion Black	1583	0.21	0.41			
Proportion Native	1583	0.01	0.09			
Proportion Asian	1583	0.03	0.16			
Proportion White	1583	0.29	0.45			

2. Methods

2.1. Participants

Participants were the 1129 children from low-income families enrolled in 671 pre-k classrooms in 11 states who participated in two studies. These two studies were designed to be combined together for analysis, using the same research team and 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) (see Howes et al., 2008 for details on both studies). 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 US who attend state pre-k programs at the time of the studies. The Multi-State Study involved a stratified sampling of 40 pre-k sites within each of 6 states during the 2001–2002 school year. The SWEEP study involved a stratified random sample of 100 state-funded pre-k programs within each of 5 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: (1) a consent form describing the study; (2) a family contact sheet; and (3) 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 in the overall study were those who: (1) had parental consent; (2) met the age criteria for kindergarten eligibility the following year; (3) did not have and Individualized Education Plan; and (4) 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.

This analysis focused exclusively on low-income children. Low-income was defined as having a household income less than 150% of the federal poverty threshold for a household of that size. If family income was missing, we assumed children were from low-income families if they attended a program targeted for low-income children. The original study sample included 2983 children enrolled in 704 classrooms; however, 1378 children were excluded because they were not from low-income families. We focused solely on the low-income children because most federal and state programs were mandated to address concerns about school readiness among low-income children. The number of children eligible for inclusion was 1605. Table 1 presents descriptive characteristics of the children participating in this study.

2.2. Measures

2.2.1. Classroom quality

The nature and quality of teacher–child interactions were assessed for each participating pre-k teacher using the Classroom assessment scoring system (CLASS; Pianta, La Paro, & Hamre, 2004). The CLASS assesses two global dimensions of the quality of teacher–child interactions within pre-k classes, Instructional Quality and Emotional Support. In prior work, higher scores on these dimensions of teacher–child interactions predicted growth in pre-k children's achievement (Howes et al., 2008; Mashburn et al., 2008); gains in academic skills in kindergarten and first grade (Burchinal et al., 2008; Hamre & Pianta, 2005); social adjustment in early childhood and elementary school (NICHD ECCRN, 2002) and concurrent levels of student engagement (Downer, Rimm-Kaufman, & Pianta, 2007; La Paro, Pianta, & Stuhlman, 2004).

In these studies reported here, a trained observer rated the pre-k classroom and teacher on nine CLASS scales (scaled from 1 = low to 7 = high) 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

Table 2
Descriptive statistics for quality measures.

CLASS	*N*	*M*	*SD*	Quality subdivided into segments	
				Lower (range)	Higher (range)
Emotional Support	1578	5.49	0.68	53% (1–4.99)	47% (5–7)
Instructional Quality	1578	2.04	0.85	87% (1–3.24)	13% (3.25–7)

Note: The cut-off score was 5 for Emotional Support and was 3.25 for Instructional Quality.

of the nine scales was computed as the average rating for that scale across the observation day. The observers rated each classroom in both the fall and the spring.

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 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. Over-control 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.

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 of process quality (La Paro et al., 2004). The first factor is termed Emotional Support, and is computed as the mean of the Positive Climate, Negative Climate (reverse scored), Teacher Sensitivity, Over-control (reverse scored), and Behavior Management ratings. The second factor is termed Instructional Quality, and it is computed as the mean of the Concept Development and Quality of Feedback ratings. The internal consistencies for the emotional and instructional quality scales were within this study sample were 0.86 and 0.78, respectively. The scales were moderately to highly consistent over time, $r = 0.59$ for the CLASS Emotional Support and $r = 0.32$ for CLASS Instructional Quality, so across time means were computed for use in this study. The two scales were moderately correlated, $r = 0.41$. Table 2 presents means and standard deviations for the CLASS scores included in this study.

Prior to data collection, the observers' reliabilities were tested on the CLASS using video tapes of pre-school classrooms and live visits to classrooms with one of the measures' authors. Across the two samples, 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, La Paro, Payne, Cox, & Bradley, 2002) and first grade (NICHD ECCRN, 2002) in which the scales were also related significantly to children's social and academic functioning.

2.2.2. 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-kindergarten 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—3rd edition (PPVT-III) 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 PPVT-III 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.

Academic achievement was measured using two scales from the Third Edition of the Woodcock-Johnson Psycho-Educational Battery (WJ) (Woodcock, McGrew, & Mather, 2001). Applied Problems (AP) was administered in both NCEDL and SWEEP studies and Letter-Word Identification (LWID) in SWEEP only. The AP test measures early math reasoning and problem-solving abilities. It requires the child to analyze and solve math problems while performing relatively simple calculations. The LWID test measures pre-reading and reading skills. It requires children to identify letters appearing in large

type and to pronounce words correctly. Both measures are widely used as measures of academic achievement and have good psychometric properties (Woodcock et al., 2001).

2.2.3. 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 five-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 0.95. Examples of problem behavior items include: "disruptive in class," "anxious," and "difficulty following directions." Teachers used a five-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 0.91. An evaluation of the normative and parametric characteristics of the TCRS is reported by Hightower et al. (1986).

2.2.4. Program type

All classrooms were selected because they were part of pre-kindergarten programs, but not all programs were located in public school buildings and some classrooms were also Head Start classes. The Head Start classrooms had to meet all state regulations for their pre-kindergarten programs and for the Head Start program. Two dummy variables were created that indicated whether the classroom was part of a Head Start program or located in public school, respectively. These were not mutually exclusive categories. For example, if the local school system was also the local Head Start agency, then a pre-kindergarten class could be in a public school building and also be a Head Start classroom.

2.2.5. Control variables

The following child and family characteristics were included in this study as control variables: pre-test scores, sex, race, and mother's education. The mother was asked to indicate the race(s) of the child using census categories. These were collapsed categories of Latino, African-American, Native American, Asian American, and European American, and children were allowed to be in multiple categories if so indicated by the mother. The mother's education was recorded in terms of the number of years of education associated with her final degree (e.g., high school degree is 12 years, associate degree is 14 years, baccalaureate is 16 years). In addition, we hypothesized effects associated with systematic differences across states in children's development; thus, ten dummy variables were included in the analysis to account for differences in rates of development across classrooms from the 11 participating states.

2.3. Analysis strategy

The general analytic strategy involved estimating a separate linear regression of child outcomes onto child care quality for low- or average-quality programs and for high- quality programs. We used a spline technique to estimate piecewise linear models (Greene, 2000) to predict four academic achievement outcomes (WJ Letter-Word, WJ Applied Problems, PPVT Receptive Language, OWLS Expressive Language) and two behavior measures (Hightower Social Competence and Behavior Problems) from the CLASS Instructional Quality and CLASS Emotional Support composites. A repeated-measures hierarchical linear model was fit to account for nesting of children within classrooms. Therefore, there was one residual term that has taken into account the clustering within classroom and an independent residual term that represented error in the individual children's scores. The model estimated separate "splines" or linear regressions for programs considered to be lower quality and higher quality, thereby estimating one slope for describing the association between quality and outcomes for programs in the higher quality range and another slope for programs in the lower quality range.

The general linear mixed analysis modeled the ith child's spring outcome score for the jth classroom, $Y_{ij\text{Spring}}$, adjusting for the fall score for that child, $Y_{ij\text{Fall}}$, and covariates. There were separate slopes estimated for each range of quality on the quality measure for the jth classroom, Quality_j, using High_j as a dummy variable. High_j has a value of 1 when the jth classroom quality is at or above the high-quality cut-off score. As recommended by Greene (2000), Quality_{ij} was centered at the first value in each range. The model is shown below,

$$Y_{ij\text{Spring}} = \delta_j + B_1 Y_{ij\text{Fall}} + B_2\,\text{Covariates}_{ij} + \varepsilon_{ij}$$

$$\delta_{ij} = B_0 + B_3\,\text{Quality}_j + B_4\,\text{High-Quality Class}_j \times \text{Quality}_j + B_5\,\text{State}_j + e_{ij}$$

Where $Y_{ij\text{Spring}}$ is the ith child's spring outcome score for the jth classroom

δ_j is the estimated intercept for the jth classroom; B_0 is the overall intercept; B_1 & B_2 are the coefficients for the fall outcome score and child/family covariates, respectively; B_3 defines the slope for quality among classrooms in the lower quality range; B_4 is the increment to the slope for quality among classrooms in the high-quality range; B_5 is the increment to the intercept associated with the 11 states.

The ranges used to define lower and higher quality, shown in Table 1, were selected based on the overall sample distribution for the CLASS. We examined several different ranges for defining high-quality care. For Emotional Support, we examined 5–7 because the CLASS developers used the same model as the Early Childhood Environmental Rating System (ECERS; Harms, Clifford, & Cryer, 1998) to define scores above 5 as high quality and 6–7 because that roughly defined the upper quartile. As shown in Table 2, the level of Instructional Quality tended to be quite low, so it was not possible to use the recommended guidelines to define high quality (e.g., less than 1% of the classrooms achieved a score of 5 or higher). Instead, we examined 2.5–7 because that roughly defined the upper quartile, 3–7 because it corresponded to definitions set by the CLASS developers and ECERS to define quality that is medium to high, and 3.25–7 because that defined the upper 15% of the distribution (i.e., the smallest group for which we felt we could reliably estimate quality slopes).

The analyses involved fitting piecewise regression models – also called spline models. These models estimate separate regression lines within different ranges of the data. For example, the Abecedarian project used a piecewise regression to determine whether the rate of change over age was different during early childhood when children received an intervention or during the follow-up period following the intervention (see Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001).

We fit a series of models. First we allowed separate intercepts and slopes in each range of quality. This approach estimated a regression line for classroom quality for the low/medium-quality classrooms and a separate regression line for classroom quality for the higher quality range. Findings did not suggest that separate intercepts provided a reasonable fit, so we allowed for separate slopes only.

The analysis models include the quality indicators (Emotional Support or Instructional Quality) and a dummy variable indicating whether the classroom was in the higher quality range as the primary predictor of interest and the fall score on that outcome measure and state, maternal education, gender, and the child's ethnicity as covariates. Whether the classroom was located in a public school or was part of a Head Start agency was also included. Classroom was indicated a fixed-effect clustering variable to account for the nesting of children in classrooms. A dummy variable indicated whether the classroom was in the high-quality or the low-medium-quality range, and this dummy variable was crossed with the quality indicator to create a quality × high interaction. We estimated effect sizes for quality indicator for each of the two quality ranges by multiplying the unstandardized coefficient by the standard deviation of the quality indicator and dividing by the standard deviation of the outcome (see NICHD ECCRN & Duncan, 2003 for details). This effect size indicated the change in the outcome in standard deviation units when quality increased by a standard deviation.

To account for missing data, multiple imputation using SAS vs 9.1 PROC MI was used to create 20 imputed data sets. Using the multiple imputation approach (Rubin, 1976, 1987; Schafer, 1997; Schafer & Graham, 2002; Widaman, 2006), missing data are estimated interatively based on assuming data were missing at random. The variables in the imputation were data collection state, maternal education, poverty status, gender, ethnicity, ECERS Language Scale, ECERS Interaction Scale, CLASS Emotional Support, CLASS Instructional Quality, teacher education, adult–child ratio, and fall and spring scale scores on the PPVT-III, OWLS, WJ Letter-Word Scale, WJ Applied Problems Scale, Hightower Competence Scale, and Hightower Behavior Problems Scale. We did not use imputed values for the missing Letter-Word Identification Scale (LWIS) scores for the NCEDL sample because that measure was not collected in that study. The procedure involves estimating predicted values for each variable from regression analyses in which that variable is predicted from all other variables in the multiple imputation. Those predicted values are then used when other variables are estimated. This process continues iteratively until the changes in the predicted values are not detectable. This missing-at-random assumption implies that the other variables in the imputation model contain sufficient information to allow reasonable estimation of missing data. All analyses are then run on each of the resulting data sets that include imputed values, and results are combined across the analyses in a manner than retains both within- and between-datasets variability in computing standard errors and test statistics.

3. Results

Several preliminary analyses were conducted prior to the primary analysis. First, we fit a series of spline regression models using the 3 selected cut-off scores. The same pattern of results obtained using all three cut-offs for both Emotional Support and Instructional Quality. We chose to use 5 as the cut-off for Emotional Support for two reasons. There tended to be slightly larger differences between the quality slopes in the lower and higher quality classrooms with that cut-off than with the other two cut-offs. Perhaps more importantly, the cut-off of 5 corresponds to the recommended cut-off for high quality on the ECERS and both the CLASS and ECERS have items with similar 7 point rating scales. In contrast, we chose the 3.25–7 to define good Instructional Quality because 3.25 was the highest value on the scale in which we had sufficient numbers of classrooms to reliably estimate a slope in the higher quality range and the same findings obtained across the three definitions of higher Instructional Quality.

Second, we asked whether the children in higher or lower quality programs differed. The demographic characteristics of the children and program characteristics in the lower and higher quality classrooms were compared. Proportionately more higher quality classrooms based on Emotional Support were located in schools than not (83% vs 71%, $\chi^2(1, N = 1583) = 13.9$, $p < 0.01$), but fewer were in a Head Start program than not (65% vs 82%, $\chi^2(1, N = 1583) = 44.7$, $p < 0.01$). In addition, proportionately more higher quality classrooms according to Instructional Quality were in schools than not (10% v 14%, $\chi^2(1, N = 1583) = 14.7$, $p < 0.01$). Compared with white children, black children were disproportionately likely to be in lower quality classrooms based on Emotional Support (63% vs 83%, $\chi^2(1, N = 1583) = 70.6$, $p < 0.01$) or based on Instructional Quality (6%

Table 3
Predicting social and academic outcomes from observed quality and testing whether associations are stronger when quality is high.

Outcomes	Main effect of quality t (1129)	Low/medium vs. high difference t (1129)	Low/medium-quality B (se)		High-quality B (se)	
CLASS emotional climate						
Behavior problems	2.30*	2.81**	0.03	(0.01)**	−0.10	(0.04)*
Social competence	2.17*	2.32*	−0.01	(0.01)	0.09	(0.04)*
Expressive language	0.97	0.89	−0.28	(0.15)	0.33	(0.61)
Receptive language	0.36	0.67	−0.21	(0.16)	0.29	(0.68)
Reading[a]	1.20	0.97	0.07	(0.47)	1.70	(1.46)
Math	1.71	1.83	0.02	(0.24)	1.70	(0.97)
CLASS instructional climate						
Behavior problems	1.80	0.17	−0.04	(0.02)*	−0.04	(0.06)
Social competence	1.44	1.28	0.03	(0.03)	0.12	(0.08)
Expressive language	3.51**	2.26*	1.34	(0.40)**	3.97	(1.37)*
Receptive language	1.55	1.71	0.52	(0.41)	2.15	(1.16)
Reading[a]	1.34	2.31*	0.70	(0.84)	5.54	(2.00)*
Math	3.36**	2.90*	2.52	(0.84)**	8.30	(2.46)**

Covariates include maternal education, gender, and race (African-American, Latino, White, or other) at the child level and whether the program was located in a school or was also a Head Start program at the classroom level.

[a] Only the data from the five states in which the WJIII Letter-Word Scale was administered are included in this analysis. The df-error for Reading is 762.

* $p < 0.05$.

** $p < 0.01$.

vs 18%, $\chi^2(1, N = 1583) = 11.9$, $p < 0.05$). To account, in part, for potential confounding, the program type and demographic variables were included in all analyses as covariates and the dependent variable was a change score. Using the change scores should reduce bias related to factors that have similar impacts on the outcomes at the pre- and post-tests (Winship & Morgan, 1999).

Third, we tested whether quality predicted outcomes differently for classrooms in Head Start programs or in public schools. The spline regression models added interactions between both the dummy variables for Head Start and being in a public school with both the quality and the interaction between being a higher quality classroom and quality. No evidence emerged that quality related to child outcomes differently overall or showed different patterns of associations in higher or lower quality classrooms in Head Start programs or public school for either Emotional Support or Instructional Quality. Those interaction terms were dropped from the model.

Finally, we conducted our primary analyses. The results are discussed below for the two measures of quality, and described in Table 3. The columns in Table 3, from left to right, show the t-statistic testing whether that quality measure was associated with outcome gains across both groups of classrooms (lower and higher quality), the t-statistic testing whether the association between quality and outcome was different for lower and higher quality classrooms, and the estimated slopes describing those associations in each of the two quality groups.

3.1. Emotional Support

The spline regressions tested whether Emotional Support predicted the six outcomes overall and whether the magnitude of prediction was stronger in higher quality classrooms (i.e., classrooms with a score of 5 or higher) than in lower quality classrooms. Results, shown in Table 3, indicated that Emotional Support was a more positive predictor of social competence and negative predictor of behavior problems in classes in the high range on Emotional Support than in classes in the low/medium range. Effect sizes were plotted in Fig. 1. Emotional Support was more strongly positively related to teacher ratings of social competence in high-quality ($B = 0.09$, $d = 0.08$) than in moderate-to-low-quality classrooms ($B = -0.01$, $d = 0.01$). Similarly, Emotional Support was more strongly negatively related to teacher ratings of behavior problems in high-quality

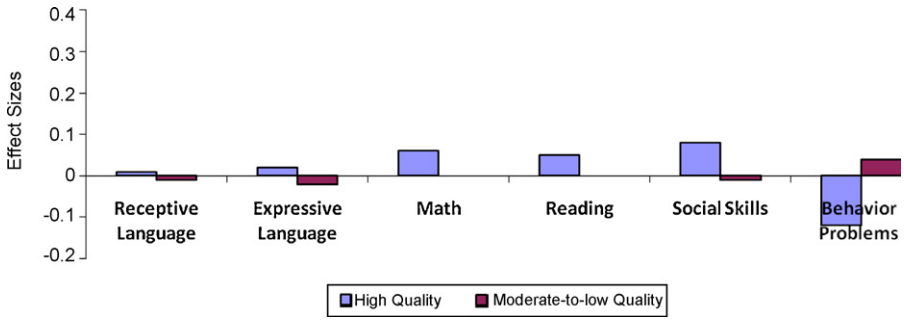

Fig. 1. Emotional Support effect sizes for high- and moderate-to-low-quality classrooms.

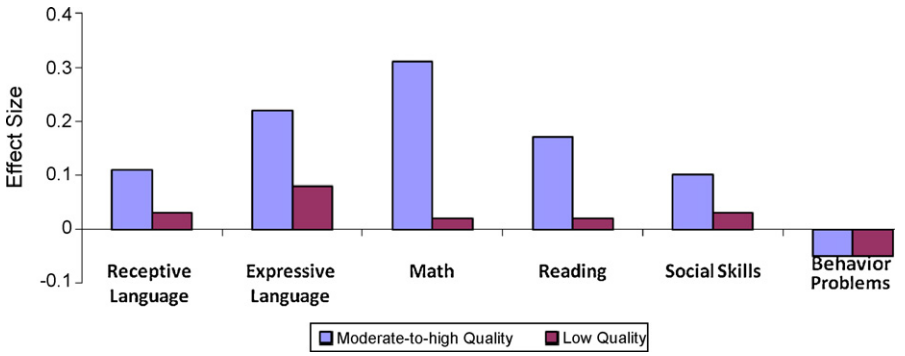

Fig. 2. Instructional Quality effect sizes for moderate-to-high- and low-quality classrooms.

($B = -0.10$, $d = -0.12$) than in moderate-to-low-quality classrooms ($B = 0.03$, $d = 0.04$). Emotional Support predicted these two social outcomes in the hypothesized direction only in the high-quality classrooms; it was a nonsignificant predictor of social competence and a positive predictor of behavior problems in moderate-to-low-quality classrooms.

3.2. Instructional Quality

Instructional Quality was related to three of the four academic outcomes in these spline regressions. As shown in Table 3 and Fig. 2, children who experienced higher quality Instructional Quality (i.e., classrooms with a score of 3.25 or higher) tended to score higher on expressive language and math overall, but the magnitude of that association was stronger in higher quality classrooms for reading, math, and expressive language. Instructional Quality predicted outcomes more strongly in higher quality than lower quality programs on expressive language (Higher: $B = 3.97$, $d = 0.23$; Lower $B = 1.34$, $d = 0.08$), reading (Higher: $B = 5.54$, $d = 0.17$; Lower $B = 0.70$, $d = 0.02$), and math (Higher: $B = 8.30$, $d = 0.34$; Lower $B = 2.52$, $d = 0.02$).

Thus, findings suggested a linear association between quality across the full range of quality and some child outcomes, but an association between quality and other outcomes only in high-quality classrooms. The quality of the teacher–child relationships predicted higher levels of social skills and lower levels of behavior problems only in classrooms characterized by more responsive and sensitive caregiving. The quality of instruction was predictive of higher levels of expressive language development for all classrooms, but the magnitude of this association was stronger in the higher quality classrooms. In contrast, quality of instruction predicted reading and math skills only in classrooms described as showing moderate-to-good instructional practices.

4. Discussion

The present study analyzed data from a large 11-state study that has reported associations between the quality of teacher–child interactions in publicly-funded pre-kindergarten programs and children's gains in academic and social performance across the pre-k year (Howes et al., 2008; Mashburn et al., 2008). The analysis was targeted around two central issues facing policy makers at the state and federal levels when deciding how to distribute scarce resources: to what extent does pre-k quality matter for children coming from poor families, and if quality matters, is there a threshold, or level, at which the effect is more or less pronounced? In the present study we combined these questions by studying threshold effects for the quality of teacher–child interactions on child outcomes for children from poor families only. The results demonstrate some support for the concept of thresholds: when the quality of teacher–child interactions met conventional definitions of good quality, higher quality teacher–child interactions predicted higher levels of social skills and lower levels of behavior problems. In lower quality classrooms, quality of teacher–child interactions did not predict social skills and predicted slightly higher – not lower – levels of behavior problems. The quality of instruction predicted language skills in general, but was a much stronger predictor of language, reading, and math skills when teachers provided moderate- to good-quality instruction. Quality of instruction predicted reading and math only in higher quality classrooms and was a much stronger predictor of language skills in higher than lower quality classrooms. The results are provocative in relation to further refinement of the association between quality and outcome gains, and have implications both for policies related to targeting the needs of children and teachers, and to further refinement of measurement systems for assessing quality.

High-quality teacher–child interactions and instruction were assessed in this study using the CLASS. A high-quality classroom according to the CLASS (La Paro et al., 2004). would involve a positive emotional tone in teachers' interactions with children and deliberate engaging instruction. Teachers frequently engage with children in interactions that reflect positive affect and comfort with one another. Teachers would be seen actively monitoring children's behavior, looking for cues for distress or confusion and they would quickly respond in ways that would enable the child to return to learning. Teachers' behaviors would be predictable, would provide children cues for how to behave; and children would be persistently offered and engaged in activities that are interesting to them. Teachers would use each moment as an opportunity to stretch

children's learning and thinking; extending conceptual understanding and thinking and teachers' feedback to children would not just focus on "correctness" but rather would elicit more complex performance of a skill through "loops" of feedback and response. Finally teachers would be actively engaged in conversations with children, eliciting their expressions, thoughts, and ideas, and at the same time shaping their use of language and vocabulary.

Identification of thresholds in the association between quality and child outcomes has been a goal of researchers and policy makers largely because they want to use resources as efficiently as possible (Barnett et al., 2008). Our large evaluation was, in part, designed to address this question because of its importance to policy makers. Policy makers were interested in determining whether there was a lower level of quality that produced relatively similar child outcomes as higher levels of quality (Blau, 2000). We found no evidence for indicating that a certain level of quality was sufficient for producing a certain level of gains as might be obtained at higher levels of quality (i.e., a level of quality beyond which further increases did not produce learning gains). In contrast, we found some evidence of minimum levels of quality, at or below which would not return learning gains or provided much lower levels of gains to exposed children. Importantly, the actual levels of these gains are, in this study, relative to both the distribution of quality (which was the primary basis for distinguishing high-quality programs from other programs) and the scale used to observe teacher–child interactions. We detected minimum levels at which a positive association between quality and outcomes was observed, for reading and for social behavior. There was no or smaller detected relation between quality and outcome gains (for emotional support and child outcomes, and for Instructional Quality and reading outcomes) until quality reached a certain point on the scale and after that minimum, gains in learning increased as quality increased, and did not level off.

These analyses indicated that there is not an asymptotic level of quality above which increases in quality are no longer associated with increases in child outcomes. Some researchers and policy makers have hoped that they could identify "good enough" levels of quality so that they could aim to fund programs that met minimal quality criteria within that good-enough level (Blau, 2000). These analyses suggested the opposite – instead of suggesting that child outcomes were not improved after quality reached some upper asymptotic level, they indicated that some child outcomes were not improved when quality fell below some lower asymptotic levels. Indeed, the magnitude of association between quality and outcomes at the higher levels of quality were larger than at the lower levels of quality for Instructional Support and academic outcomes and for Emotional Support and social outcomes, suggesting there was the potential for greater cost-benefits in implementing improvements at the higher ranges of quality.

We believe that, if our findings are causal, the goals of pre-kindergarten programs may only be achieved if programs ensure high-quality teacher–child interactions and at least moderate-quality instruction. These findings indicate that social outcomes were more strongly influenced by the quality of teacher–child interactions, but only when teachers are actively and positively engaged with children as indicated when caregivers are rated in the 5–7 range on the CLASS Emotional Support Scale. It is likely that young children need a caregiver who actively and positively engages them in order for these teachers to be able to teach or modify these social skills (Hamre and Pianta, 2005). Similarly, these findings indicate that children acquire academic skills only when the minimal standards represented by our cut-off point of above a 3.25 on the CLASS Instructional Quality Dimension are met, and that higher quality instruction produces more academic gains. It is likely that below that point, there is too little explicit instruction or guided child-centered teaching for academic learning to occur.

This form of threshold effect suggests two implications. First, if child outcomes for low-income children are improved through child care experiences in high-quality programs, then it is especially important to ensure that low-income children experience at least the minimum level of quality child care. Such a finding might translate into policies that restrict vouchers to pay for care that met that threshold or only funding programs that met the minimum level of quality. Relatedly, policy might also provide incentives for teachers to obtain professional development to improve the quality of their classrooms in order to access state support. Second, the results also imply that supports for teachers above the threshold are important so they continue to improve the quality of their interactions; recall there was no asymptotic association and children's gains continued to increase even as quality topped out on the scale.

There are several limitations. First, all analyses were tests of association not causation. We adjusted for the children's academic and social skills at entry to pre-kindergarten, but this analysis strategy does not allow us to make causation inferences. Further research should replicate these findings with independent data sets and, hopefully, use data from successful quality improvement clinical trials to examine whether these threshold effects obtain in studies with more rigorous designs. Second, the effect sizes tended to be small even in the high-quality classroom. They ranged from $0.08 < d < 0.34$. These are as large or larger than the effect sizes reported in most large-scale observational studies of child care (see Burchinal et al., 2009) and we believe they could have practical significance. Very few studies of fully implemented programs such as the mature state pre-kindergarten programs we examined produce moderate-to-large effect sizes. However, others would argue they do not meet that standard of $d > 0.30$ that defines educationally meaningful differences (Slavin, 1989). Third, our policy recommendations assume the primary purpose of subsidized preschool care is to improve child outcomes. Encouraging parental employment is another purpose, and, while we believe we should strive toward both goals simultaneously, it is clear that many leaders in the field worry about restricting access if costs increase associated with adding quality restrictions (Blau, 2000). Fourth, high-quality instruction was not typically observed in this study, so conclusions about the level of quality for instruction in pre-kindergarten may be underestimated in practice. It may be difficult for these programs to ensure that instructional quality is sufficiently high to obtain the apparent gains in child outcomes associated with higher quality care. Fifth, only about half of the children approached for recruitment were successfully enrolled. Most cases involved failure to obtain a returned parental permission rather than an active declination. Nevertheless, it is likely that parents with less

income or education may be been less likely to consent (NICHD ECCRN, 1997), so findings may be biased even within this low-income sample.

In conclusion, findings from this study suggest that high-quality classrooms may be necessary to optimally improve social skills, reduce behavior problems, and promote reading, math, and language skills. Unlike the expected threshold levels suggesting there was a "good-enough" level, these findings suggest that children may not obtain social and academic benefits from pre-kindergarten experiences unless the teacher maintains high-quality teacher–child interactions and at least moderate- to high-quality instruction.

Acknowledgements

The authors gratefully acknowledge the National Institute for Early Education Research (NIEER), The Pew Charitable Trusts, and the Foundation for Child Development for their support of the SWEEP Study, and the U.S. Department of Education for its support of the Multi-State Study of Pre-Kindergarten and the current Early Childhood Center. However, the contents do not necessarily represent the positions or policies of the funding agencies, and endorsement by these agencies should not be assumed. The authors are grateful for the help of the many children, parents, teachers, administrators, and field staff who participated in these studies.

References

Barnett, W. S., Hustedt, J. T., Friedman, A. H., Boyd, J. S., & Ainsworth, P. (2008). *The state of preschool, 2007*. (downloaded June 20, 2008). http://nieer.org/yearbook/pdf/yearbook.pdf

Blau, D. M. (2000). The production of quality in child-care centers: Another look. *Applied Developmental Science*, *4*(3), 136–147.

Burchinal, M., Roberts, J. E., Zeisel, S. A., Hennon, E. A., & Hooper, S. (2006). Risk and resiliency: Protective factors in early elementary school years. *Parenting: Science and Practice*, *6*, 79–113.

Burchinal, M. R., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher–child interactions and instruction. *Applied Development Science*, *12*, 140–153.

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. *Applied Developmental Science*, *4*, 149–165.

Burchinal, P., Kainz, K., Cai, K., Tout, K., Zaslow, M., Martinez-Beck, I., et al. (2009, May). *Early care and education quality and child outcomes*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, US DHHS, and Child Trends.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M. R., & Ramey, C. (2001). The development of cognitive and academic abilities: Growth curves from an early intervention educational experiment. *Developmental Psychology*, *37*(2), 231–242.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian project. *Applied Developmental Science*, *6*(1), 42–57.

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

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*(2), 233–248.

Downer, J., Rimm-Kaufman, S., & Pianta, R. (2007). How do classroom conditions and children's risk for school problems contribute to children's engagement in learning? *School Psychology Review*, *36*(3), 413–433.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test* (3rd Edition). Circle Pines, MN: American Guidance Service.

Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology*, *41*(6), 872–884.

Greene, W. H. (2000). *Econometric analysis* (4th ed.). Upper Saddle River, NJ: Prentice-Hall.

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*(5), 949–967.

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

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

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

La Paro, K., Pianta, R., & Stuhlman, M. (2004). Classroom assessment scoring system (CLASS): Findings from the pre-k year. *Elementary School Journal*, *104*(5), 409–426.

Lamb, M. (1998). *Nonparental child care: Context, quality, correlates, and consequences* (5th ed.). In I. Sigel, A. Renninger, & W. Damon (Eds.), *Handbook of child psychology: Child psychology in practice* New York: Wiley., pp. 73–133.

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

Magnuson, K. A., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal*, *41*(1), 115–157.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., et al. (2008). Measures of classroom quality in prekindergarten and childen's development of academic, language, and social skills. *Child Development*, *79*(3), 732–749.

McCartney, K., Dearing, E., Beck, A., & Bub, K. (2007). New findings from secondary data analysis – Results from the NICHD Study of Early Child Care and Youth Development. *Journal of Applied Developmental Psychology*, *28*, 411–426.

NICHD Early Child Care Research Network. (1997). Child care in the first year of life. *Merrill-Palmer Quarterly*, *43*, 340–360.

NICHD Early Child Care Research Network. (1999). Child outcomes when child-care center classes meet recommended standards for quality. *American Journal of Public Health*, *89*(7), 1072–1077.

NICHD Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development*, *71*(4), 960–980.

NICHD Early Child Care Research Network. (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *The Elementary School Journal*, *102*(5), 367–387.

NICHD Early Child Care Research Network. (2005). Early child care and children's development in the primary grades: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *42*(3), 537–570.

NICHD Early Child Care Research Network. (2006). Child care effect sizes for the NICHD Study of Early Child Care and Youth Development. *American Psychologist*, *61*(2), 99–116.

NICHD Early Child Care Research Network, & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, *74*(5), 1454–1475.

Nores, M., Barnett, W. S., Belfield, C. R., & Schweinhart, L. J. (2005). Updating the economic impacts of the High/Scope Perry Preschool program. *Educational Evaluation and Policy Analysis*, *27*(3), 245–262.
Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of preschool child care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, *72*, 1534–1553.
Pianta, R. C., La Paro, K., Payne, C., Cox, M., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *Elementary School Journal*, *102*(3), 225–238.
Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2004). *Classroom assessment scoring system [CLASS] manual: Pre-K*. Baltimore, MD: Brookes Publishing.
Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: A 15-year follow-up of low-income children in public schools. *Journal of the American Medical Association*, *285*, 2339–2346.
Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.
Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). Issues for the next decade of quality rating and improvement systems. *Issue Brief*, *14*(3) (Retrieved from http://www.childtrends.org/Files//Child_Trends-2009_5_19_RB_QualityRating.pdf)
Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.
Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177.
Slavin, R. (1989). Class size and student achievement: Small effects of small classes. *Educational Psychologist*, *24*, 99–110.
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*. Ypsilanti, MI: High/Scope Educational Research Foundation.
Vandell, D. (2004). Early child care: The known and the unknown. *Merrill-Palmer Quarterly*, *50*, 387–414.
Votruba-Drzal, E., Coley, R. L., & Chase-Lansdale, P. L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development*, *75*, 296–312.
Widaman, K. F. (2006). Missing data: What to do with or without them. In W. F. Overton, W. A. Collins, K. McCartney, M. R. Burchinal, & K. L. Bud (Eds.), *Monographs of the Society for Research in Child Development* (pp. 42–64).
Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*, 659–706.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III: Tests of achievement*. Itasca, IL: Riverside Publishing.