What's Happening

October 2016

Patterns of classroom quality in Head Start and center-based early childhood education programs

Clare W. Irwin
John P. Madura
David Bamat
Education Development Center, Inc.

Paul A. McDermott
University of Pennsylvania

In collaboration with the Early Childhood Education Research Alliance

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences
U.S. Department of Education

REL NORTHEAST & ISLANDS
Regional Educational Laboratory
At Education Development Center, Inc.

U.S. Department of Education
John B. King, Jr., *Secretary*

Institute of Education Sciences
Ruth Neild, *Deputy Director for Policy and Research*
Delegated Duties of the Director

National Center for Education Evaluation and Regional Assistance
Joy Lesnick, *Acting Commissioner*
Amy Johnson, *Action Editor*
Elizabeth Eisner, *Project Officer*

REL 2017–199

The National Center for Education Evaluation and Regional Assistance (NCEE) conducts unbiased large-scale evaluations of education programs and practices supported by federal funds; provides research-based technical assistance to educators and policymakers; and supports the synthesis and the widespread dissemination of the results of research and evaluation throughout the United States.

October 2016

This report was prepared for the Institute of Education Sciences (IES) under Contract ED-IES-12-C-0009 by Regional Educational Laboratory Northeast & Islands administered by Education Development Center, Inc. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Irwin, C. W., Madura, J. P., Bamat, D., & McDermott, P. A. (2016). *Patterns of classroom quality in Head Start and center-based early childhood education programs* (REL 2017–199). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

This report is available on the Regional Educational Laboratory website at http://ies.ed.gov/ncee/edlabs.

Summary

Numerous early childhood studies have found a positive relationship between participation in high-quality early childhood education classrooms and favorable academic outcomes, language development, and social skills (for example, Burchinal, Kainz, & Cai, 2011; Burchinal et al., 2009; Keys et al., 2013; National Institute of Child Health and Human Development Early Child Care Research Network, 2000; Peisner-Feinberg et al., 2001; and Weiland & Yoshikawa, 2013). But study results may be contingent on how quality is measured (Mashburn et al., 2008). Research also suggests that different aspects of early childhood education classroom quality may have different effects on academic outcomes, language development, and social skills—for example, positive interactions between teachers and children may have a greater influence than the number of books available in the classroom. In addition, studies show that higher quality teacher–child interactions and more effective use of curricula are most closely related to positive outcomes (Burchinal et al., 2008; Yoshikawa et al., 2013).

Measuring classroom quality and ensuring high-quality learning experiences for young children are interests of the Early Childhood Education Research Alliance, a research alliance of Regional Educational Laboratory Northeast & Islands. This study, conducted in collaboration with the alliance, addresses these interests by examining multiple measures of classroom quality simultaneously. Many measures of early childhood classroom quality have been examined, but previous research has not explored whether multiple measures of diverse aspects of classroom quality can be used to classify early childhood classrooms into classroom quality groups.

The main purpose of this study was to determine whether early childhood classrooms can be grouped based on their scores on multiple measures of quality. The study team used multiple measures of diverse aspects of classroom quality to determine classroom quality groups and examined the number of classroom quality groups that exist and the percentage of classrooms that fall within each classroom quality group. A second purpose of the study was to explore the extent to which each measure contributes to the identification of classroom quality groups. This study employs a methodological approach that has not previously been used to synthesize measures of classroom quality. This approach allowed the study team to provide an example of patterns of classroom quality in programs serving Head Start–eligible children across the country. In addition to informing practitioners about what quality looks like in these settings, the findings allow for the categorization of classrooms across multiple classroom quality dimensions. This is more useful for informing practitioners than the traditional way of measuring quality along a continuum—eliminating the need for users to come up with their own cutscores for these measures.

Latent class analysis was used with data from the spring 2003 Head Start Impact Study (Puma et al., 2012) to model patterns of classroom quality using such measures as the number of instructional activities and the nature of teacher–child interactions. Latent class analysis is an exploratory multivariate statistical approach that identifies a distinctive combination of related measures of classroom quality, including teacher–child interactions, instructional activities, and physical environment. Key findings include:

- Based on the 13 measures examined, classroom quality can be sorted into three distinct groups: good, fair, and poor.
- Classroom quality measures determined by independent observers distinguish classroom quality groups better than self-reported measures do.

Contents

Why this study?

Providing high-quality early childhood education to all children, especially children from low-income families, has recently gained national attention. In a 2014 Gallup Poll 70 percent of Americans reported that they support using federal funds to make high-quality preschool available to all children (McCann, 2014). In addition, grant competitions such as the Race to the Top—Early Learning Challenge and the President's Preschool for All and Preschool Development Grants initiatives have brought increased attention to measuring early childhood education quality to increase the chances that young children receive high-quality learning experiences. Many states aspiring to achieve universal high-quality preschool have implemented or are developing quality rating and improvement systems that aim to comprehensively assess and incentivize improvements in program quality.

The main purpose of this study was to determine whether early childhood classrooms can be grouped based on their scores on multiple measures of quality

Although higher quality early learning experiences are related to more favorable child outcomes (see, for example, Burchinal, Kainz, & Cai, 2011; Burchinal et al., 2008; National Institute of Child Health and Human Development Early Child Care Research Network, 2000; Peisner-Feinberg et al., 2001), different aspects of early childhood education classroom quality may have different effects on academic outcomes, language development, and social skills—for example, positive interactions between teachers and children may have a greater influence than the number of books available in the classroom. Some research suggests that high-quality teacher–child interactions and effective use of curricula are most closely related to positive outcomes (Yoshikawa et al., 2013).

Measuring classroom quality and ensuring high-quality learning experiences for young children is an interest of the Early Childhood Education Research Alliance, a research alliance of Regional Educational Laboratory Northeast & Islands whose members include employees of state agencies that oversee early childhood quality improvement initiatives. Alliance members wanted to better understand ways of measuring classroom quality and to identify which measures best differentiate between high- and low-quality classrooms (see box 1 for definitions of key terms). This study was conducted in collaboration with the alliance to address this interest by using multiple measures of diverse aspects of classroom quality to determine classroom quality groups.

What the study examined

The main purpose of this study was to determine whether early childhood classrooms can be grouped based on their scores on multiple measures of quality. The study team also examined the number of classroom quality groups that exist and the percentage of classrooms that fall within each classroom quality group. This study employs a methodological approach that has not previously been used to synthesize measures of classroom quality. This approach, known as latent class analysis, allowed the study team to uncover hidden patterns of classroom quality in data collected from programs serving Head Start–eligible children across the country. In addition to informing practitioners about what quality looks like in these settings, the findings allow for the categorization of classrooms across multiple classroom quality dimensions. This is more useful for informing practitioners than the traditional way of measuring quality along a single continuum—eliminating the need for users to come up with their own cut scores for these measures. A second purpose of the study was to explore the extent to which each measure contributes to the identification of classroom quality groups.

Box 1. Key terms

Center-based early childhood classroom. An early learning environment that takes place in a formal setting outside a child's or caregiver's home, usually provided by an incorporated organization of professional child care providers. For this study, center-based care providers are distinct from Head Start providers and are the centers that children who were eligible for Head Start attended but were randomly assigned to not receive Head Start. Center-based programs vary programmatically—for example, some provide full-day services while others provide part-day services, and some use an evidence-based curriculum while others do not use any curriculum at all; also, center-based programs vary in mission or purpose.

Classroom quality. The structure, processes, and interactions in the early childhood classroom setting that have been associated with positive child outcomes in the research literature.

Classroom quality group. A group of classrooms that display similar patterns of quality across 13 measures (box 2).

Head Start. Established in 1965 as one of President Lyndon B. Johnson's "war on poverty" initiatives, Head Start is a program administered by the U.S. Department of Health and Human Services that provides early childhood education, health, nutrition, and parent involvement services to low-income children from birth to age 5 and their families. Like the center-based programs, Head Start programs vary programmatically—for example, some provide full-day services while others provide part-day services, and the specific curriculum used varies.

Head Start Impact Study. A Congressionally mandated impact study conducted with a nationally representative sample of 4,667 3- and 4-year-old children who were eligible for Head Start and randomly assigned to either a participating Head Start program or a control group that did not have access to the participating Head Start programs but could enroll in other early childhood programs selected by their families. Data are collected through classroom observations, teacher reports, and surveys. Several publications have been produced from this study (for example, Bloom & Weiland, 2015; Puma et al., 2012).

Instructional activities. Classroom experiences prepared and carried out by teachers to foster the cognitive, academic, and social-emotional development of children in their care.

Latent class analysis. A multivariate statistical method for identifying groups that are empirically distinguishable based on different patterns of responses to observed measures. More specifically, it is a statistical analysis used to relate or assign a set of observations from a sample or population to unobservable (or latent) groups characterized by a pattern of conditional probabilities that indicate the chance that characteristics take on certain values.

Patterns of classroom quality. A distinctive combination of related measures of classroom quality, including teacher–child interactions, instructional activities, and physical environment.

Teacher–child interactions. The nature of the exchanges between the teacher and child in a child care setting, including the teacher's emotional tone, discipline style, and responsiveness, as well as other indicators of closeness or conflict between the teacher and child.

The study was guided by two research questions:

- How many patterns of quality are evident across classrooms, and what percentage of classrooms follow each pattern?
- Which measures of quality contribute to the identification of classroom quality groups?

The study used data for Head Start classrooms (n = 1,061) and center-based classrooms (n = 421) from the 2002/03 Head Start Impact Study (Puma et al., 2012), which includes a nationally representative sample of Head Start–eligible children, some of whom attended Head Start programs and others of whom were in the study's control group and attended alternative center-based programs (29 classrooms were removed because of missing data, leaving 1,453 classrooms in the analysis). On average, there were fewer than two classrooms per center or school. Family and home-based child care arrangements did not have the same data on classroom quality and were thus excluded from the study.

Thirteen measures were used to explore patterns of classroom quality, representing the most robust classroom quality measures that were available from the Head Start Impact Study

Practitioner experience suggests that classroom quality is a function of the complex interplay of the learning space and the people who inhabit it (Daniels & Shumow, 2003). For this reason, 13 measures were used to explore patterns of classroom quality (see box 2). Although the measures do not constitute an exhaustive list of the components of a broader quality construct that may lead to variations in quality across early childhood classrooms, they represent the most robust classroom quality measures that were available from the Head Start Impact Study. Latent class analysis was used to identify patterns of classroom quality across the 13 measures. Because the measures were not independently administered, the study team conducted analyses to assess the degree to which the 13 measures were intercorrelated. None of the measures was highly correlated with the others, as the correlations all fell below the standard .85 threshold (Clark & Watson, 1995). See appendix A for details on the study methodology.

Box 2. Thirteen measures used to identify patterns of early childhood classroom quality

The 13 measures used to identify patterns of classroom quality fall into two categories: self-reported measures and observation-based measures.

Self-reported measures

Two self-reported measures relate to student–teacher relationships based on the Student–Teacher Relationship Scale (Pianta, 2001). An examination of the intercorrelation between the measures resulted in a bivariate correlation of –.27 (p < .001).

- *Conflict.* This measure represents the degree to which a child care provider perceives detachment, disagreement, and mistrust between himself or herself and a particular child. Example item: "The child feels that I treat him/her unfairly." The measure contains eight items, which are rated on a five-point scale from definitely does not apply to definitely applies and were self-reported by teachers in spring 2003. Item scores are averaged to create the conflict score, which ranges from 1 to 5. For the current study, scores were averaged across all children in a classroom to establish a classroom score.

(continued)

Box 2. Thirteen measures used to identify patterns of early childhood classroom quality *(continued)*

- *Closeness.* This measure represents the degree to which a child care provider perceives attachment and trust between himself or herself and a particular child. Example item: "This child values his/her relationship with me." The measure contains seven items, which are rated on a five-point scale from definitely does not apply to definitely applies and were self-reported by teachers in spring 2003. Item scores are averaged to create the closeness score, which ranges from 1 to 5. For the current study, scores were averaged across all children in a classroom to establish a classroom score.

Three self-reported measures relate to instructional activities and are from the spring 2003 Head Start Impact Study teacher survey. An examination of the intercorrelation among the measures suggested low to moderate correlations.

- *Literacy activities.* This measure represents the number of literacy activities (such as practice writing letters and words, retelling or making up stories, and learning about rhyming words) offered in the classroom at least three times per week, as self-reported by teachers. Scores range from 0 to 12.
- *Math activities.* This measure represents the number of math activities (such as counting out loud, using music to understand math, and working with rulers and other measuring instruments) offered in the classroom at least three times per week, as self-reported by teachers. Scores range from 0 to 7.
- *Other activities.* This measure represents the number of other activities offered in the classroom at least three times per week (such as working on arts and crafts, playing sports or exercising, and having children assist with classroom chores), as self-reported by teachers. Scores range from 0 to 4.

Observation-based measures

Six observation-based measures relate to the classroom environment and use the Early Childhood Environment Rating Scale—Revised (Harms, Clifford, & Cryer, 1998). An examination of the intercorrelation among the measures suggested moderate correlations.

- *Space and furnishings.* This measure represents the quality of the classroom environment's space for routine care, play, learning, and privacy; furnishings for relaxation and comfort; and gross motor equipment and child-related displays. The measure contains eight items, which are rated on a seven-point scale from inadequate to excellent based on the judgment of an observer in spring 2003. Item scores are averaged to create the space and furnishings score, which ranges from 1 to 7.
- *Personal care routines.* This measure represents the quality of routines related to greeting or departing, meals or snacks, nap or rest, toileting or diapering, health practices, and safety practices. The measure has six items, which are rated on a seven-point scale from inadequate to excellent based on the judgment of an observer in spring 2003. Item scores are averaged to create the personal care routines score, which ranges from 1 to 7.
- *Language-reasoning.* This measure represents the quality of the classroom's books and pictures as well as the caregiver's encouragement of children to communicate, children's use of language to develop reasoning skills, and the informal use of language. The measure has four items, which are rated on a seven-point scale from inadequate to excellent based on the judgment of an observer in spring 2003. Item scores are averaged to create the language-reasoning score, which ranges from 1 to 7.

(continued)

Box 2. Thirteen measures used to identify patterns of early childhood classroom quality *(continued)*

- *Activities.* This measure represents the quality of fine motor, art, music/movement, nature/science, and math/number activities, as well as activities that involve blocks, dramatic play, sand/water, digital technology, and the promotion/acceptance of diversity. The measure has 10 items, which are rated on a seven-point scale from inadequate to excellent based on the judgment of an observer in spring 2003. Item scores are averaged to create the activities score, which ranges from 1 to 7.
- *Interactions.* This measure represents the quality of the caregiver's supervision of activities and discipline, as well as the teacher–child interactions and the interactions among children. The measure has five items, which are rated on a seven-point scale from inadequate to excellent based on the judgment of an observer in spring 2003. Item scores are averaged to create the interactions score, which ranges from 1 to 7.
- *Program structure.* This measure represents the quality of the classroom's schedule, free play, group time, and provisions for children with disabilities. The measure has four items, which are rated on a seven-point scale from inadequate to excellent based on the judgment of an observer in spring 2003. Item scores are averaged to create the program structures score, which ranges from 1 to 7.

Two independent observation–based measures relate to sensitivity in teacher interactions with children and use the Arnett Caregiver Interaction Scale (Arnett, 1989). An examination of the intercorrelation between the measures suggested moderate correlations.

- *Harshness.* This measure represents the degree to which a caregiver behaves harshly in his or her interactions with the children in their care. Example item: "Punishes the children without explanation." The measure has nine items, which are rated on a four-point scale from not at all true to very much true based on the judgment of an observer in spring 2003. Item scores are averaged to create the harshness score, which ranges from 1 to 4.
- *Sensitivity.* This measure represents the degree to which a caregiver behaves sensitively in his or her interactions with children in his or her care. Example item: "Listens attentively when children speak to him/her." The measure has 10 items, which are rated on a four-point scale from not at all true to very much true based on the judgment of an observer in spring 2003. Items scores are averaged to create the sensitivity score, which ranges from 1 to 4.

Note: See table A2 in appendix A for a full table of bivariate correlations among the 13 measures.

What the study found

Head Start and center-based early childhood classrooms can be grouped into three distinct classroom quality groups displaying different patterns of classroom quality.

There are three distinct classroom quality groups based on the 13 measures examined: good, fair, and poor quality

Three distinct classroom quality groups emerged from Head Start and center-based early childhood classrooms' scores on the 13 measures of quality examined (see box 2 for descriptions of the 13 measures). Four models were tested, ranging from two to five distinct groups, and the model with three distinct groups was found to be most appropriate for the 13 measures examined.

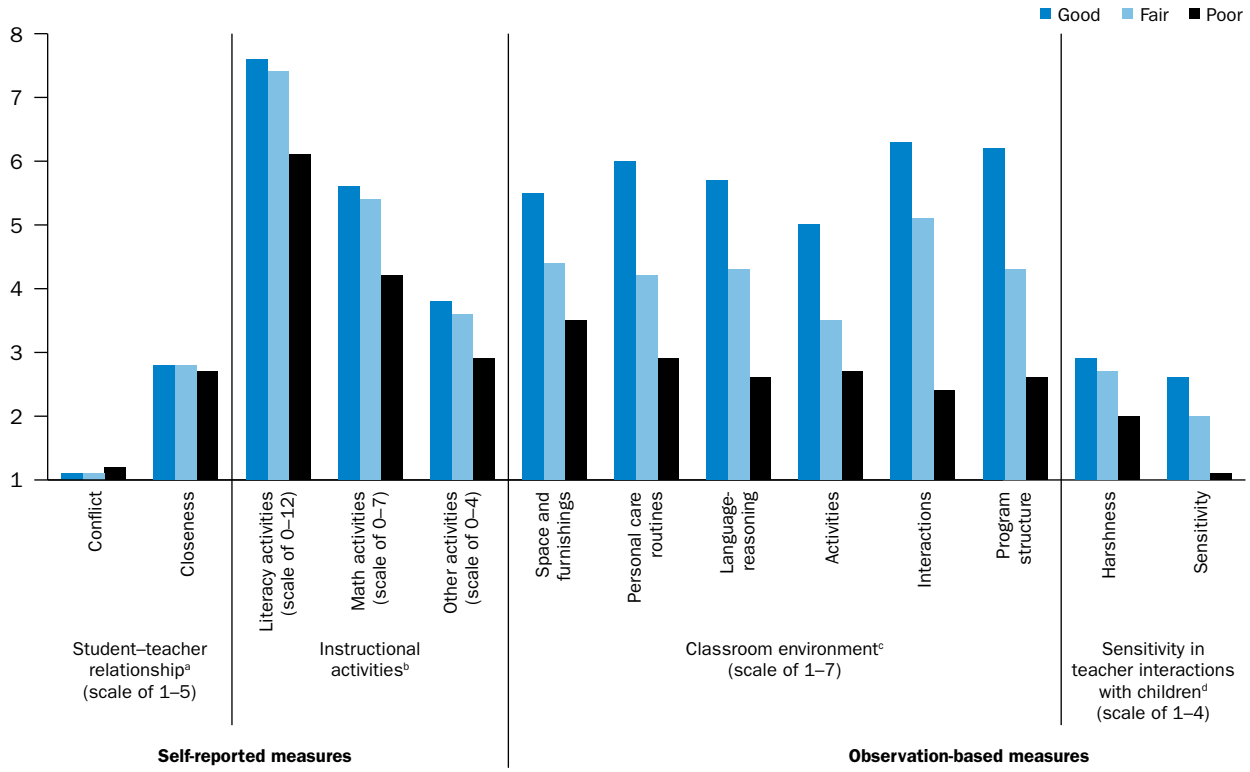
The three classroom quality groups can be described as good, fair, and poor. Classrooms with similar patterns were grouped, and group average scores on each measure were used to describe the three classroom quality groups:

- *Good quality.* This group is characterized by good to excellent classroom environment scores (means ranged from 5.0 to 6.3 on a seven-point scale) and above-average sensitivity in teacher interactions with children scores (mean of 2.6 on a four-point scale; figure 1 and table 1). Teachers in this group also reported average closeness with their children (mean of 2.8 on a five-point scale), low conflict with their children (mean of 1.1 on a four-point scale), and an average number of instructional activities at least three times per week (7.6 literacy activities, 5.6 math activities, and 3.8 other activities). Based on these features, this group could be described as a sensitive, environmentally rich classroom group. This group accounted for 62 percent of the classrooms examined in this study (n = 901).
- *Fair quality.* This group is characterized by minimal to good classroom environment scores (means ranged from 3.5 to 5.1 on a seven-point scale) as well as average sensitivity and harshness in teacher interactions with children scores (mean of 2.0 and 2.7 on a four-point scale, respectively). Teachers in this group also reported average closeness with their children (mean of 2.8 on a five-point scale). This group differs from the good-quality group in that this group has lower scores on average in both the classroom environment and sensitivity of teacher interactions with children than the good-quality group has. But this group is similar to the good-quality group in that teachers in this group reported low conflict with their children and an average number of instructional activities at least three times per week. This group accounted for 30 percent of the classrooms examined in this study (n = 436).
- *Poor quality.* This group is characterized by inadequate to minimal classroom environment scores (means ranged from 2.4 to 3.5 on a seven-point scale) and teachers displaying little sensitivity in their interactions with children (mean of 1.1 on a four-point scale). This group displays scores similar to those of the other groups on the closeness and conflict scales. Based on these features, this group could be described as an insensitive, somewhat harsh, environmentally poor classroom group. This group accounted for 8 percent of the classrooms examined in this study (n = 116).

Three distinct classroom quality groups emerged from Head Start and center-based early childhood classrooms' scores on the 13 measures of quality examined. Four models were tested, ranging from two to five distinct groups, and the model with three distinct groups was found to be most appropriate for the 13 measures examined

Classrooms can be grouped based on their performance across all 13 measures of classroom quality. But on any given measure an individual classroom may have a score that deviates from the average for its group. The standard deviation indicates how close the scores for all classrooms in a group are to the group's average. A larger standard deviation indicates that the scores are spread out over a wider range of values. For example, a classroom may be in the fair-quality group when examining multiple measures, but that classroom's score on language-reasoning may be as high as some of the classrooms in the good-quality group or as low as some of those in the low-quality group (figure 2). This overlap means that classrooms' performance on each measure vary both across classroom quality groups and within a given quality pattern group. See table A4 in appendix A for the standard deviation of all 13 measures for the full sample and each group.

Figure 1. The three distinct early childhood classroom quality groups each display a different average pattern of classroom quality, 2002/03



Note: See box 2 for a full description of the measures.

a. Uses the Student–Teacher Relationship Scale (Pianta, 2001).

b. From the spring 2003 Head Start Impact Study teacher survey.

c. Uses the Early Childhood Environment Rating Scale—Revised (Harms, Clifford, & Cryer, 1998).

d. Uses the Arnett Caregiver Interaction Scale (Arnett, 1989).

Source: Authors' analysis of 2002/03 data from the Head Start Impact Study.

Observation-based measures distinguish classroom quality patterns better than self-reported measures do

The eight measures that were based on independent observations contributed to the identification of classroom quality groups more than the five self-reported measures did. This can be seen by comparing the average scores on each measure across groups and by determining the extent to which the variation in scores is explained by the classroom quality groups (see table A4 in appendix A for the percentage of variation in the measures of classroom quality explained by group membership).

For the six classroom environment measures, which were scored by independent observers, average scores for the classrooms in the good-quality group ranged from 5.0 to 6.3 (on a seven-point scale), but scores for the fair-quality group ranged from 3.5 to 5.1 (see table 1) and scores for the poor-quality group ranged from 2.4 to 3.5. The three groups also had different average scores on the sensitivity in teacher interactions with children measure, which was also scored by an independent observer: 2.6 (on a four-point scale) for the good-quality group, 2.0 for the fair-quality group, and 1.1 for the poor-quality group.

Table 1. Average rating on each early childhood classroom quality measure, by group membership, 2002/03

Measure	Full sample (n = 1,309–1,349[a])	Good quality group (n = 787–819[a])	Fair quality group (n = 397–403[a])	Poor quality group (n = 95–102[a])
Self-reported measures				
Student–teacher relationship[b] (scale of 1–5)				
Conflict	1.1	1.1	1.1	1.2
Closeness	2.8	2.8	2.8	2.7
Instructional activities[c]				
Literacy activities (scale of 0–12)	7.4	7.6	7.4	6.1
Math activities (scale of 0–7)	5.4	5.6	5.4	4.2
Other activities (scale of 0–4)	3.6	3.8	3.6	2.9
Observation-based measures				
Classroom environment[d] (scale of 1–7)				
Space and furnishings	5.0	5.5	4.4	3.5
Personal care routines	5.2	6.0	4.2	2.9
Language-reasoning	5.0	5.7	4.3	2.6
Activities	4.4	5.0	3.5	2.7
Interactions	5.6	6.3	5.1	2.4
Program structure	5.4	6.2	4.3	2.6
Sensitivity in teacher interactions with children[e] (scale of 1–4)				
Harshness	2.7	2.9	2.7	2.0
Sensitivity	2.3	2.6	2.0	1.1

Note: See box 2 for a full description of the measures.

a. The latent class analysis model failed to assign classrooms to groups only when data were missing for all 13 measures of quality. As a result, the number of classrooms with values for each of the individual quality measures varies. The range of sample sizes for each group represents the smallest and largest number of classrooms with values over the individual measures of quality. The overall group membership remains the same as in the descriptions of the groups in the text: 901 for the good-quality group, 436 for the fair-quality group, and 116 for the poor-quality group.

b. Uses the Student–Teacher Relationship Scale (Pianta, 2001).

c. From the spring 2003 Head Start Impact Study teacher survey.

d. Uses the Early Childhood Environment Rating Scale—Revised (Harms, Clifford, & Cryer, 1998).

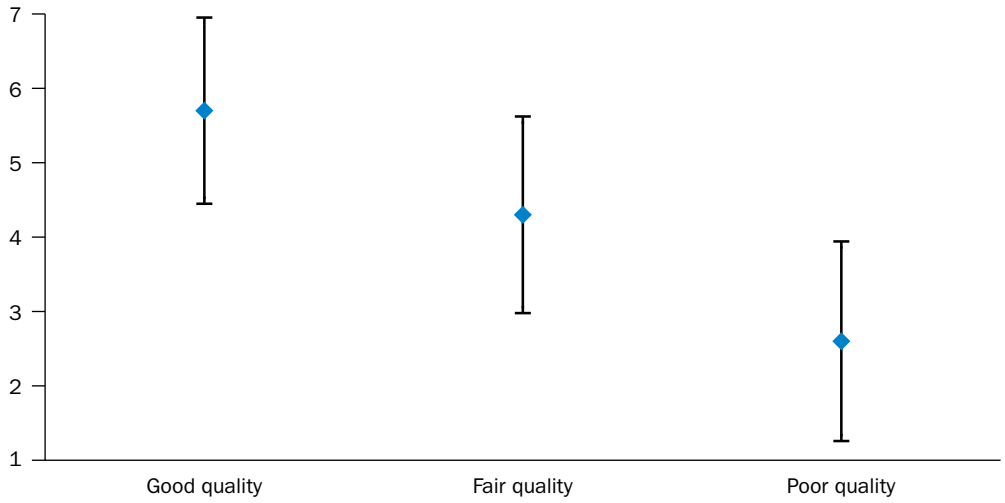
e. Uses the Arnett Caregiver Interaction Scale (Arnett, 1989).

Source: Authors' analysis of 2002/03 data from the Head Start Impact Study.

Average harshness scores differed for good- and poor-quality classrooms (2.9 versus 2.0 on a four-point scale), but the average scores for good- and fair-quality classrooms were closer (2.9 versus 2.7).

In contrast, average scores for the self-reported measures of conflict and closeness in the student–teacher relationships category did not differ substantively. The mean score that teachers reported for conflict in their relationship with their children was 1.1 (on a five-point scale) for the good- and fair-quality groups and 1.2 for the poor-quality group. The mean score that teachers reported for closeness in their relationship with children was 2.8 (on a five-point scale) for the good- and fair-quality groups and 2.7 for the poor-quality group). Although the average number of literacy, math, and other activities differs between the good- and poor-quality group, the scores vary so much within each group that they do not contribute to the identification of the classroom quality groups (see table 1 and table A4 in appendix A).

Figure 2. Example of the variability within each early childhood classroom quality group: Language-reasoning, 2002/03



Note: The language-reasoning measure is part of the Early Childhood Environment Rating Scale—Revised (Harms et al., 1998) and has a scale of 0–7. See box 2 for a full description of the measure. The diamond represents the mean, and the vertical bars represent the range of one standard deviation above and below the mean.

Source: Authors' analysis of 2002/03 data from the Head Start Impact Study.

Implications of the study findings

The study findings suggest that measures representing different elements of the classroom experience can identify classroom quality patterns, that classrooms can be grouped according to those patterns, and that the majority of classrooms can be described as good quality, while few can be described as poor quality

There are three main implications of the study findings.

It is possible to use multiple dimensions of the classroom experience to identify classroom quality patterns

The findings suggest that measures representing different elements of the classroom experience can identify classroom quality patterns and that classrooms can be grouped according to those patterns. The findings also suggest that the majority of classrooms can be described as good quality, while few can be described as poor quality. Practitioners and policymakers can use this information, along with the body of evidence currently available in the research literature, to inform how they define quality. In addition, an extension of this study could include a similar exploration using local data (in a collaboration between practitioners and researchers) and more updated measures of quality to see whether similar patterns unfold or whether there are associations with teacher and program characteristics or academic outcomes, language development, or social skills.

Identifying classroom quality patterns will likely require independent observers

The findings suggest that practitioners and policymakers may not be able to rely on self-reported measures of classroom quality. The self-reported measures in this study showed little difference in their means across the three classroom quality groups identified. In contrast, measures based on independent observation using the Early Childhood Environment Rating Scale—Revised and the Arnett Caregiver Interaction Scale provided results that better distinguished patterns of classroom quality. The self-reported measures contributed little statistical or descriptive information to the pattern identification and grouping

process. As with any self-reported measures, practitioners should consider their limitations in differentiating classroom quality. The self-report measures used in this study did not differentiate classrooms on quality as effectively as observation-based measures. If self-reported measures are desired or deemed necessary, practitioners may need to explore alternative self-reported measures that better differentiate classrooms on quality.

An individual classroom may not be perfectly characterized by its classroom quality group

Although it is possible to assign individual classrooms to quality groups with confidence, classroom scores will likely vary on the individual measures. Practitioners could use classroom quality groups as a way to identify professional development opportunities that may be suitable for a particular set of classrooms but may also need to consider the individual needs of specific classrooms by examining the results on each measure. There are many possible explanations (and models) for variation across individual classrooms within a quality group, but even this small example suggests that additional analyses could help shape policies surrounding classroom quality.

The self-reported measures in this study showed little difference in their means across the three classroom quality groups identified, suggesting that practitioners and policymakers may not be able to rely on self-reported measures of classroom quality

Limitations of the study

This study has three main limitations. First, the data used in the study are from the 2002/03 school year. In addition to improvements in the measurement of classroom quality since that time, there have been several national and numerous state-level preschool initiatives aimed at raising early childhood education quality (for example, the Race to the Top—Early Learning Challenge grant program, the Preschool Development Grants program, and state quality rating and improvement systems). Because of these changes, the results of the current study may not reflect current classroom quality, even in similar samples of Head Start–eligible children.

Second, the sample used in the study includes only classrooms serving Head Start–eligible children (such as children below the federal poverty line whose parents applied for Head Start services). The results of the study may not be relevant for programs serving other populations of children (such as children from middle- or upper-income families). In addition, the findings do not provide information about the early learning experiences of children attending home-based early education and child care arrangements.

Third, although latent class analysis identifies patterns of classroom quality, it does not account for observed within-group variation in the classroom quality measures. Modeling could be improved by expanding to a more general factor mixture model that includes a factor analytic component that specifically accounts for this observed heterogeneity. In this type of factor mixture model, if there are statistically significant variations of quality on measures for classrooms within a group, the variation could be modeled. The fit of this type of model would suggest that even within a particular classroom quality group, classrooms can have varying amounts of quality on a given measure. This knowledge could help practitioners understand the range of quality for given measures across groups. In addition, because of the high percentage of missing data on covariates related to teacher preparation and teacher supports, it was not possible to gather validity evidence for the three-group solution based on their relationship with these variables.

Appendix A. Data and methodology

This appendix describes the data and methodology used for the study.

Study sample

The study used data for Head Start (n = 1,061) and center-based (n = 421) classrooms from the 2002/03 Head Start Impact Study (Puma et al., 2012), which includes a nationally representative sample of Head Start–eligible children. Head Start Impact Study data were collected through classroom observations, teacher reports and surveys, and center director interviews.

To address classrooms being nested within centers, the degree to which multiple classrooms were sampled within a center was explored. There were 1,482 classrooms nested within 799 centers in the data file as a whole—or fewer than two classrooms clustered within a center, on average. Further inspection indicated that there are, on average, 1.8 classrooms per center in the study sample, that 72 percent of the centers had only one classroom, and that 85 percent of the centers had no more than two classrooms. As a result, it was assumed that clustering effects were negligible and that reporting intraclass correlations would be misleading because of the large number of clusters with fewer than two members.[1] Twenty-nine classrooms were missing data on all the measures used to generate the latent classes and were removed. A total of 1,453 classrooms were included in the analysis.

Measures

Thirteen measures were used to explore patterns of classroom quality. The measures were from the Student–Teacher Relationship Scale (Pianta, 2001); the number of literacy, math, and other activities provided in the classroom at least three times per week; the Early Childhood Environment Rating Scale—Revised (Harms et al., 1998); and the Arnett Caregiver Interaction Scale (Arnett, 1989; see box 2 of the main text and table A1 for more details). In addition, the study team used bivariate correlations to determine the degree to which the 13 measures of classroom quality were related to each other. Many of the measures were weakly correlated. The highest correlations were among the measures using the Early Childhood Environment Rating Scale—Revised but were all below the standard .85 threshold for multicollinearity and thus provided evidence of discriminant validity (Clark & Watson, 1995; also see table A2). As a result, the 13 measures were deemed independent and theoretically unrelated to each other and were used independently in the study analysis.

Data analysis

Latent class analysis. A series of latent class analysis models were estimated to explain the relationships among the 13 measures of classroom quality. Statistical analyses were conducted using Mplus Version 7 (Muthén & Muthén, 1998–2012). Latent class analysis identifies groups of cases (classrooms) with similar characteristics. Latent class analysis models differ from traditional types of cluster analyses in that latent class analysis models are probability-based classifications. In latent class analysis, cases are classified into groups based on membership probabilities estimated directly from the model.

Table A1. Measures used to identify early childhood classroom quality patterns, 2002/03

Measuresz	Description
Self-reported measures	
Instructional activities	The number of self-reported literacy, math, and other activities offered at least three times per week in each of these three domain areas, as reported by the lead teacher in the Head Start Impact Study data file.
Student–Teacher Relationship Scale	The short form of an instrument previously developed by Pianta (2001). It includes a conflict (α = .92), closeness (α = .86), and total positive relationship scale (α = 89). Continuous raw scores were provided in the data for each child (as rated by their teacher). Spring 2003 classroom mean scores from each of the closeness and conflict scales were used.
Observation-based measures	
Early Childhood Environment Rating Scale—Revised	Includes information on six subscales of classroom quality, including space and furnishings (α = .76), personal care routines (α = .72), language-reasoning (α = .83), activities (α = .88), interactions (α = .86), and program structure (α = .77), as well as a total environment scale (α = .92). Criterion-referenced scores ranging from 1 to 7 for each of the six subscales were provided by the Head Start Impact Study data file for spring 2003.
Arnett Caregiver Interaction Scale	Includes information on teacher sensitivity, responsiveness, encouragement of independence, and punitiveness and detachment. Item-level data were available in the Head Start Impact Study data file for spring 2003. The harshness (α = .83) and sensitivity (α = .93) scores were the only two measures with measures of internal consistency that exceeded .70, so these were the two Arnett measures included in the study. Average scores for each of the two measures were provided by the Head Start Impact Study data file for spring 2003.

Source: Authors' compilation.

Table A2. Measures used to find early childhood classroom quality patterns display weak to moderate bivariate correlations, 2002/03

Classroom quality measure	1	2	3	4	5	6	7	8	9	10	11	12
Student–Teacher Relationship Scale												
1 Conflict												
2 Closeness	–.27***											
Early Childhood Environment Rating Scale—Revised												
3 Space and furnishings	–.02	.00										
4 Personal care routines	–.03	.00	.59***									
5 Language-reasoning	–.01	.00	.49***	.54***								
6 Activities	–.01	–.01	.69***	.56***	.67***							
7 Interactions	–.03	.03	.48***	.60***	.70***	.53***						
8 Program structure	.03	–.01	.61***	.56***	.57***	.68***	.60***					
Arnett Caregiver Interaction Scale												
9 Harshness	–.02	.01	.40***	.44***	.48***	.42***	.66***	.48***				
10 Sensitivity	–.03	.04	.41***	.50***	.63***	.49***	.72***	.48***	.62***			
Instructional activities												
11 Literacy activities	–.10***	.18***	.05	.02	.05	.05	.01	–.01	.00	.03		
12 Math activities	–.10***	.12***	.08***	.05	.05	.09***	.02	.06**	.07**	.05	.62***	
13 Other activities	–.07**	.10***	.15***	.17***	.14***	.19***	.11***	.18***	.09***	.08***	.35***	.43***

** Significant at $p < .01$; *** significant at $p < .001$.

Source: Authors' analysis of 2002/03 data from the Head Start Impact Study.

Measures may be continuous, categorical, counts, or any combination thereof. Demographics and other variables can be used to describe the groups. In latent class analysis multivariate normality is not assumed; however, latent class analysis does assume that the observed measures are independent from each other. The consequence of this assumption is that the relationship among the observed measures is entirely accounted for by an underlying categorical variable (latent class variable). In other words, the selected observed measures are conditionally and locally independent given their latent class membership (Collins & Lanza, 2010).

Model-based approaches such as latent class analysis use estimated membership probabilities to classify cases into the appropriate group. The most popular model-based approach is known as mixture modeling, where each latent class represents an unobserved group (McLachlan & Basford, 1988; Vermunt & Magidson, 2002). Latent class analysis assumes a simple parametric model and uses observed data to estimate parameter values for it. As described above, the current study uses 13 measures to define classroom quality groups. The model parameters are the prevalence of cases in each pattern, and for each latent group the probability of being placed into one of the resultant latent groups.

The optimal class solution was first determined through a statistical analysis of the results. Entropy is one measure of classification error (or model fit), where measures of entropy that tend toward 1.0 indicate less classification error of classrooms into groups. Another measure of the fit of a latent class model to the data is the Bayesian information criterion, where the model with the lowest Bayesian information criterion is considered to have the best fit.

To further test the fit of a model, the Mplus software also computes the likelihood ratio test for each model of k classes compared with a model that has one fewer class (k–1 classes) and provides a p-value that can be used to determine whether there is a statistically significant improvement in fit for the inclusion of one more class. Specifically, the k–1 class model is tested as the null model against the k class model. The likelihood ratio test is performed incrementally where all (k) class solutions are tested. The preferred class solution is the one that shows significantly better fit than the solution with one fewer class and where the solution with one more class does not show improved fit—this solution is considered the simplest or most parsimonious class solution.

Two to five latent class solutions were tested using the procedures outlined above. Results of the statistical tests referenced are summarized in table A3.

Patterns of classroom quality. Results suggested that the two-, three-, four-, and five-class solutions all produced acceptable classification error rates in the proposed class solutions (entropy values were .90, .87, .88, and .85). The Bayesian information criterion decreased with each increasing complex model (values were 45,153.7, 43,676.6, 43,019.4, and 42,478.8) and provided little insight into the best fitting model. However, the adjusted likelihood ratio test provided the statistical evidence needed to determine the appropriate number of groups to explain scores on the classroom quality measures. For the Head Start and center-based early childhood classrooms the three-class model was determined to be the best-fitting model. The likelihood ratio test yielded a significant improvement over the two-class solution (adjusted likelihood ratio test = 4,093.4, $p < 0.001$), and the four-class solution was not significantly better than the three-class solution (adjusted likelihood ratio

Table A3. Model fit statistics for the tested latent class models of early childhood classroom quality groups suggest the three-class model is the best solution, 2002/03

Model	Number of free parameters	Bayesian information criterion	–2 log likelihood	Adjusted likelihood ratio test	*p*-value	Entropy
Two-class	40	45,153.7	–22,431.2	4,093.4	< .001	.90
Three-class	54	43,676.6	–21,641.7	1,563.7	< .001	.87
Four-class	68	43,019.4	–21,262.1	751.8	.121	.88
Five-class	82	42,478.8	–20,940.9	636.3	.067	.85

Note: The model represents the number of classroom quality groups tested. Number of free parameters represents the complexity of the model (sum of latent class prevalences and the estimated means). The Bayesian information criterion is a comparative fit index where Bayesian information criterion = –2*(log likelihood) + (number of parameters)*[ln(*n*+2/24)]. –2 log likelihood is used to determine optimal values of estimated coefficients. The adjusted likelihood ratio test compares the improvement in fit between neighboring class models (that is, comparing *k*–1 and the *k* class models) and provides a *p*-value that can be used to determine whether there is a statistically significant improvement in fit for the inclusion of one more class. Entropy is the measure of the degree of disorder or uncertainty in the model. *P*-values less than .05 represent significantly better model fit above the model with one less class. Based on this criteria, coupled with the lower Bayesian information criterion for the three-class model over the two-class model, the three-class model appears to provide the best fit with the study data.

Source: Authors' analysis of 2002/03 data from the Head Start Impact Study.

test = 751.8, $p = 0.121$); therefore, the three-class model was determined to model the data in the most parsimonious way.

In addition, while the four-class solution differentiated classrooms better than the three-class solution at the upper level, lowering the percentage of classrooms in this group from 62 percent to 47 percent, it further lowered the poor quality group from 8 percent to 5 percent of the sample (only 69 classrooms). Thus, in combination with the statistical reasons, there were theoretical and interpretive reasons for rejecting the four-class solution in favor of the more parsimonious three-class solution.

The means across the 13 observed measures of classroom quality differ among the classroom quality groups. The estimated means for each of the 13 measures used to characterize the latent classroom quality groups and the variability associated with each of the measures are presented in table A4. The variability in the classroom quality measures that is explained by the latent classroom quality group membership is presented in the last column of table A4—under the assumption that all of the variance in the responses is explained by class membership and the residual variances for each latent class is fixed at zero. The latent class model also assumes that the variability of the classroom quality indicators is invariant across all three groups. The results indicate that 0.6–65.2 percent of the observed variance in the measures can be explained by group membership in the three-class solution. The measures where little of the variance is explained by group membership (Student–Teacher Relationship Scale and instructional measures) are also the measures that provide poor separation among the three classes (see figure 1 in the main text).

Classrooms have a probability of membership in each of the three classroom quality groups, where a particular classroom may have a 35 percent chance of being assigned to the good-quality group at the same time that it has a 20 percent chance of being assigned to the poor-quality group and 45 percent chance of being assigned to the fair-quality group.

Table A4. Means and standard deviations of quality measures by early childhood classroom quality group and percent of variance of each quality measure explained by group membership, 2002/03

Measure	Full sample (n = 1,309–1,349[a])	Good-quality group (n = 787–819[a])	Fair-quality group (n = 397–403[a])	Poor-quality group (n = 95–102[a])	Percentage of measure variance explained by group membership
Self-reported measures					
Student–teacher relationship[b] (scale of 1–5)					
Conflict	1.1 (0.46)	1.1 (0.45)	1.1 (0.48)	1.2 (0.44)	0.6
Closeness	2.8 (0.33)	2.8 (0.32)	2.8 (0.35)	2.7 (0.40)	0.9
Instructional activities[c]					
Literacy activities (scale of 0–12)	7.4 (3.06)	7.6 (3.09)	7.4 (1.00)	6.1 (2.98)	1.6
Math activities (scale of 0–7)	5.4 (2.05)	5.6 (2.04)	5.4 (0.94)	4.2 (2.04)	3.7
Other activities (scale of 0–4)	3.6 (0.74)	3.8 (0.75)	3.6 (0.97)	2.9 (0.76)	11.6
Observation-based measures					
Classroom environment[d] (scale of 1–7)					
Space and furnishings	5.0 (1.04)	5.5 (1.04)	4.4 (1.02)	3.5 (1.13)	40.8
Personal care routines	5.2 (1.46)	6.0 (1.42)	4.2 (1.51)	2.9 (1.50)	48.2
Language-reasoning	5.0 (1.29)	5.7 (1.26)	4.3 (1.33)	2.6 (1.35)	49.6
Activities	4.4 (1.17)	5.0 (1.16)	3.5 (1.17)	2.7 (1.25)	43.5
Interactions	5.6 (1.27)	6.3 (1.25)	5.1 (1.32)	2.4 (1.31)	65.2
Program structure	5.4 (1.61)	6.2 (1.61)	4.3 (1.61)	2.6 (1.62)	51.2
Sensitivity in teacher interactions with children[e] (scale of 1–4)					
Harshness	2.7 (0.31)	2.9 (0.30)	2.7 (0.71)	2.0 (0.34)	52.3
Sensitivity	2.3 (0.60)	2.6 (0.58)	2.0 (0.88)	1.1 (0.66)	52.7

Note: Numbers in parentheses are standard deviations.

a. The latent class analysis model failed to assign classrooms to groups only when data were missing for all 13 measures of quality. As a result, the number of classrooms with values for each of the individual quality measures varies. The range of sample sizes for each group represents the smallest and largest number of classrooms with values over the individual measures of quality. The overall group membership remains the same as in the descriptions of the groups in the text: 901 for the good-quality group, 436 for the fair-quality group, and 116 for the poor-quality group.

b. Uses the Student–Teacher Relationship Scale (Pianta, 2001).

c. From the spring 2003 Head Start Impact Study teacher survey.

d. Uses the Early Childhood Environment Rating Scale—Revised (Harms, Clifford, & Cryer, 1998).

e. Uses the Arnett Caregiver Interaction Scale (Arnett, 1989).

Source: Authors' analysis of 2002/03 data from the Head Start Impact Study.

For this study, classrooms were assigned to the classroom quality group for which they had the highest probability of membership. The largest number of classrooms were grouped into the good-quality group (n = 901), with substantially fewer classrooms grouped into the fair-quality (n = 436) and poor-quality (n = 116) groups. Twenty-nine classrooms were unable to be grouped into a quality pattern because of lack of data on the measures used to assign classrooms to quality groups. The number of classrooms assigned to each classroom quality group and the minimum, maximum, and average probabilities of membership for each classroom quality group are displayed in table A5.

Table A5. The number of early childhood classrooms assigned to each quality pattern group differed, but the average probability of membership was similar across all quality groups, 2002/03

Group assignment	*n*	Percent of sample	Minimum probability	Maximum probability	Mean probability
Good quality	901	62	0.35	1.00	0.94
Fair quality	436	30	0.35	1.00	0.94
Poor quality	116	8	0.37	1.00	0.95

Note: Twenty-nine classrooms were unable to be grouped into a quality pattern because of lack of data on the measures used to assign classrooms to quality groups.

Source: Authors' analysis of 2002/03 data from the Head Start Impact Study.

Note

1. Although the intraclass correlation likely exceeds the traditionally accepted level of .01 for the measures used and it would generally be appropriate to account for the nonindependence of the data (Cohen, Cohen, West, & Aiken, 2003), the study team determined that the use of a multilevel latent class analysis model for the data was beyond the scope of the study. This decision was made largely because the effects of ignoring nested structures "have not yet been fully resolved" (Muthén & Asparouhov, 2011) due to the fact that "multilevel mixture modeling is a rather new area of statistical methodology" (Vermunt, 2011, p. 78). Furthermore, model specification is not a trivial problem in multilevel latent class analysis. One key concern is the difficulty in achieving convergence in multilevel latent class analysis model estimation (Van Landeghem, De Fraine, & Van Damme, 2005). Moreover, Chen (2012) suggests that accounting for higher level structure will improve standard errors in estimated group means and classification of cases, particularly when groups are balanced, which is not the case in this study. Research on this methodology, combined with the observation that the sample includes a predominance of single-classroom centers, led the study team to conclude that ignoring the higher structure in the data is appropriate assumption for this exploratory study.

References

Arnett, J. (1989). *Arnett Caregiver Interaction Scale*. Retrieved December 2, 2015, from http://www.eec.state.ma.us/docs1/qris/20110121_arnett_scale.pdf.

Bloom, H. S., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study*. New York: MDRC. Retrieved April 24, 2015, from http://www.mdrc.org/sites/default/files/quantifying_variation_in_head_start.pdf.

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 Developmental Science, 12*(3), 140–153.

Burchinal, M. R., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes: A meta-analysis and coordinated analyses of data from large-scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Measuring quality in early childhood settings* (pp. 11–31). Baltimore, MD: Brookes Publishing.

Burchinal, M. R., Kainz, K., Cai, Y., Tout, K., Zaslow, M., Martinez-Beck, I., et al. (2009). *Early care and education quality and child outcomes* (Research-to-Policy Research-to-Practice Brief: OPRE Research-to-Policy Brief 1). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation, and Child Trends. http://eric.ed.gov/?id=ED510748

Chen, Q. (2012). The impact of ignoring a level of nesting structure in multilevel mixture model: A Monte Carlo study. *SAGE Open, 2*(1).

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*, 3rd edition. Mahwah, NJ: Earlbaum.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Hoboken, NJ: John Wiley and Sons, Inc.

Daniels, D. H., & Shumow, L. (2003). Child development and classroom teaching: A review of the literature and implications for educating teachers. *Applied Developmental Psychology, 23*(5), 495–526.

Harms, T., Clifford R., & Cryer, D. (1998). *Early Childhood Environment Rating Scale: Revised Edition (ECERS-R)*. New York, NY: Teachers College Press.

Keys, T. D., Farkas, G., Burchinal, M. R., Duncan, G. J., Vandell, D. L., Li, W., et al. (2013). Preschool center quality and school readiness: Quality effects and variation by demographic and child characteristics. *Child Development, 84*(4), 1171–1190.

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

McCann, C. (2014). *Americans say they support federal dollars for pre-k*. Washington. DC: New America EdCentral. Retrieved November 15, 2015, from http://www.edcentral.org/americans-say-they-support-federal-dollars-pre/.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models: Inference and applications to clustering.* New York, NY: Marcel Dekker.

Muthén B. O., & Asparouhov T. (2011). Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. In Hox J. J., Roberts J. K. (Eds.), *Handbook for advanced multilevel analysis* (pp. 15–40). New York, NY: Routledge/Taylor & Francis.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide, seventh edition*. Los Angeles, CA: Muthén & Muthén.

National Institute of Child Health and Human Development Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development, 71*(4), 960–980. http://eric.ed.gov/?id=EJ620372

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*(5), 1534–1553. http://eric.ed.gov/?id=EJ643710

Pianta, R. C. (2001). *Student-teacher relationship scale: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Downer, J., et al. (2012). *Third grade follow-up to the Head Start Impact Study: Final report.* Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation. http://eric.ed.gov/?id=ED539264

Van Landeghem G., De Fraine B., & Van Damme J. (2005). The consequence of ignoring a level of nesting in multilevel analysis: A comment. *Multivariate Behavioral Research*, 40, 423–434.

Vermunt J. K. (2011). Mixture models for multilevel data sets. In Hox J. J., Roberts J. K. (Eds.), *Handbook for advanced multilevel analysis* (pp. 59–81). New York, NY: Routledge/Taylor & Francis.

Vermunt, J. K., & Magidson, J. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research, 20*(1), 37–44.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*(6), 2112–2130.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Zaslow, M. J., et al (2013). *Investing in our future: The evidence base on preschool education*. Ann

Arbor, MI: Society for Research in Child Development. Retrieved April 8, 2015, from http://fcd-us.org/sites/default/files/Evidence%20Base%20on%20Preschool%20Education%20FINAL.pdf.

The Regional Educational Laboratory Program produces 7 types of reports

Making Connections
Studies of correlational relationships

Making an Impact
Studies of cause and effect

What's Happening
Descriptions of policies, programs, implementation status, or data trends

What's Known
Summaries of previous research

Stated Briefly
Summaries of research findings for specific audiences

Applied Research Methods
Research methods for educational settings

Tools
Help for planning, gathering, analyzing, or reporting data or research