Do Early Childhood Environment Rating Scale-3 and Pre-K Classroom Assessment Scoring System Scores Need to Reach Thresholds to Predict Children's School Readiness?

Jessica F. Harding, Tutrang Nguyen & Nikki Aikens

View supplementary material ⧉

Published online: 05 Aug 2025.

Submit your article to this journal ⧉

View related articles ⧉

View Crossmark data ⧉

Routledge
Taylor & Francis Group

Do Early Childhood Environment Rating Scale-3 and Pre-K Classroom Assessment Scoring System Scores Need to Reach Thresholds to Predict Children's School Readiness?

Jessica F. Harding ⓘ, Tutrang Nguyen, and Nikki Aikens

Nutrition, Health, and Human Services Division, Mathematica Inc

ABSTRACT

Research Findings: We examined associations between young children's school readiness and the Early Childhood Environment Rating Scale Third Edition (ECERS-3) and the Pre-K Classroom Assessment Scoring System (Pre-K CLASS), and whether these measures need to reach thresholds to promote children's outcomes. We only found one significant linear association (of 30) between the ECERS-3 and Pre-K CLASS and children's outcomes. We only detected three significant threshold findings (of 30). These suggested ECERS-3 Math Activities and Learning Opportunities scores were stronger predictors of literacy skills in classrooms above the median on quality than in lower-quality classrooms. This provides the first evidence that there are some limited and small associations between the ECERS-3 and children's learning when quality meets certain thresholds. We also found evidence that Pre-K CLASS Classroom Organization scores were associated with children's behavioral outcomes when classroom quality was above the median. We did not find evidence of thresholds for other quality subscales or outcomes, or when using conceptually selected thresholds, supporting recent findings that associations between classroom quality and children's outcomes are limited-even when examining thresholds. *Practice or policy*: Early childhood monitoring systems should consider whether and how to set relevant thresholds for quality given limited associations between observed quality and children's outcomes.

Decades of research suggest that high-quality classrooms in early care and education (ECE) settings can support healthy development in children, including development of their academic and social-emotional skills (Hamre, 2014; NICHD Early Child Care Research Network, 2003; Mashburn et al., 2008). Of central importance to a high-quality classroom is what happens inside – namely, the interactions between the teacher and children. In a high-quality classroom, teachers engage children with learning strategies tailored to their developmental level and use an appropriate curriculum to structure the learning experience (Weiland et al., 2018). A well-trained and highly skilled teacher tailors interactions to fit the needs of the child – using responsive language, engaging them in classroom activities, fostering independence, and creating a language-rich environment. Effective ECE teachers proactively prevent and redirect challenging behavior and respond to children's needs with respect, warmth, and empathy.

Increasingly, ECE systems have acknowledged the importance of high-quality teacher–child interactions in promoting children's well-being. For example, many states currently include classroom quality scores in their Quality Rating and Improvement Systems (QRISs) (Build Initiative, 2021).

Quality scores are also used for accountability as part of the Head Start Designation Renewal System (Head Start Early Childhood Learning and Knowledge Center, 2020). In some of these systems, quality scores have to reach certain thresholds for ratings or continued funding, but it is unclear whether these thresholds are important for children's development (Mashburn, 2017; Tout et al., 2009).

The current study is designed to understand the extent to which associations exist between classroom quality and children's school readiness outcomes as assessed by two of the measures used most commonly to define ECE quality, the Early Childhood Environment Rating Scale Third Edition (ECERS-3; Harms et al., 2015) and Pre-K Classroom Assessment Scoring System (CLASS; Pianta et al., 2008). Specifically, we examine whether these measures need to reach thresholds of quality to promote children's school readiness. Understanding thresholds in the association between classroom quality and child outcomes is important for answering policy-relevant questions about what level of quality will produce stronger associations with children's outcomes (Burchinal et al., 2010; Zaslow et al., 2016). If there are thresholds in quality that classrooms need to meet or exceed if they are to influence school readiness, and if additional causal evidence supported these thresholds, then the best use of resources might focus on raising classrooms above that threshold.

Conceptual and theoretical framework

The conceptualization of quality is generally broken into distinct aspects of the classroom environment: structural quality – regulable elements such as teachers' education and training, classroom ratios, or group sizes, and process quality – social, emotional, physical, and instructional elements of teacher–child interactions (Pianta et al., 2005). In this model, structural quality is necessary but not enough for promoting children's school readiness (NICHD Early Child Care Research Network, 2002). Process quality is assumed, and has been shown, to be associated with more favorable developmental outcomes for children (NICHD Early Child Care Research Network, 2003; Mashburn et al., 2008). At the center of process quality is sensitive and responsive interactions that consistently take place between children and teachers (Hamre, 2014). Developmental and educational theories emphasize that children's learning and development are influenced by their interactions with others – particularly with adults (Bronfenbrenner & Morris, 2006; Vygotsky et al., 1978). The quality of children's proximal-level interactions with their teachers or caregivers are thought to be the drivers by which they learn and develop (Hamre et al., 2013). For example, interactions characterized by warmth, safety, and connection create a secure and supportive environment for children to positively engage in social interactions and learning activities. Some of the most widely used measures of classroom quality today, two of which are examined in the current study, were designed to reflect these prominent theories.

Measures of ECE classroom quality and their associations with children's outcomes

Although there is no single definition of high quality in ECE, and therefore no single tool to measure it, there are widely used tools to assess and report on the quality of early childhood programs. Two of the most common suite of measurement tools are the Environmental Rating Scale (ERS; Harms & Clifford, 1980) and the CLASS. For both, trained observers who have demonstrated they can reliably use the tool spend several hours observing ECE classrooms. Observers rate the classroom's quality on several dimensions.

ECERS-3 and children's outcomes
The ECERS-3 is part of the ERS, the most commonly used suite of observational tools in states' QRISs (BUILD Initiative & Child Trends, 2019). About three-quarters of states use the ERS for rating or quality improvement (Build Initiative, 2021). Many researchers, evaluators, and state agency staff are either transitioning or considering a transition to the ECERS-3 from its predecessor, the Early Childhood Environment Rating Scale – Revised (ECERS-R; Harms et al., 2005). Like the ECERS-R,

the ECERS-3 measures quality in terms of access to enriching activities and a healthy and safe classroom environment. However, the ECERS-3 places more emphasis than the ECERS-R did on the role of the teacher in helping children develop cognitive and social skills. For example, the ECERS-3 has new items about language, literacy, and math, and items focus less on accessible materials and more on how adults use these materials (Hestenes et al., 2019). In addition, whereas the ECERS-R incorporates information from direct observations and interviews with teachers, the ECERS-3 relies solely on direct observation. A study of the ECERS-R and ECERS-3 in one state found a correlation of 0.60 between scores from observations conducted a few weeks apart, with the ECERS-3 showing lower mean scores and greater variability than the ECERS-R, suggesting that these tools are different from one another (Hestenes et al., 2019).

Most research on the ECERS-R from the 1990s and early 2000s found positive associations between higher scores on the ECERS-R and children's development (for a review, see R. C. Pianta, 2012). However, more contemporary research has found limited or small positive associations between the ECERS-R and children's outcomes (Auger et al., 2014; Brunsek et al., 2017; Gordon et al., 2013; Sabol & Pianta, 2014), perhaps due to better accounting for the selection of different families into higher-quality settings. The ECERS-3 was expected to improve the prediction of child outcomes (Harms et al., 2015). Early et al. (2018) used the ECERS-3 to examine associations with child outcomes. The study examined 30 associations and found 5 significant associations with the total score and subscales identified by the authors' factor analysis: total scores and Teacher Interactions were associated with executive function skills; Math Activities were associated with social skills; and Learning Opportunities were associated with executive function and early math skills. Of those that were statistically significant, the associations were small, with effect sizes ranging from 0.06 to 0.08.

Pre-K CLASS and children's outcomes

The Pre-K CLASS assesses the teacher's level of responsiveness and sensitivity, the extent to which teachers provide and scaffold in-depth learning, and the teacher's overall organization of the classroom. After the ERS, it is the most commonly used observation tool in QRISs, with half of states using it for rating or quality improvement (Build Initiative, 2021). Past research has shown that Pre-K CLASS scores are related to children's school readiness outcomes, but the magnitude of the linear associations is generally small (Aikens et al., 2021; Burchinal, 2018; Perlman et al., 2016), with effect sizes typically smaller than 0.10. Moreover, recent research has found limited significant associations between the Pre-K CLASS and children's outcomes (Guerrero-Rosada et al., 2021; Nguyen et al., 2022).

Taken together, most of the recent literature on associations between classroom quality measures and children's outcomes has not found associations with outcomes, and even when associations are found, the effect sizes are small. In addition, research about associations between quality and children's outcomes is challenged by measurement and selection issues. In terms of measurement, classroom quality scores can vary because of the time of the day (Finders et al., 2021), activities going on in the classroom (Mashburn, 2017), and rater differences (Mantzicopoulos et al. 2018). In terms of selection, children are not randomly sorted into classrooms, and their opportunities to attend high-quality classroom experiences are associated with socioeconomic status, race and ethnicity, and parents' choices (Burchinal & Nelson, 2000; Coley et al., 2014). For example, more advantaged families often have access to and attend higher-quality settings whereas less advantaged families attend lower-quality settings that are available to them (Hillemeier et al., 2013). Therefore, associations between classroom quality and children's outcomes could be driven by unobserved variables.

Thresholds in the associations between classroom quality and children's outcomes

Because of the small linear associations between classroom quality measures and children's outcomes, the field has questioned whether there might be thresholds above which quality is more strongly related to outcomes (Hatfield et al., 2016). If there is evidence for thresholds in quality, this can mean that there are

no associations between quality and children's outcomes below the threshold and that classrooms need to meet a threshold of quality to be associated with children's outcomes. As with the linear associations between quality and children's outcomes, it is important to note that threshold findings could occur because the association of quality and children's outcomes is stronger for children in higher-quality classrooms because of other characteristics of those children and their families, not just the higher-quality learning environments.

Few studies have examined thresholds with the Pre-K CLASS, and evidence is mixed as to whether the Pre-K CLASS is a stronger predictor of children's school readiness outcomes in higher-quality classrooms. Some studies found stronger associations between classroom quality and children's outcomes in classrooms with moderate to high levels of either Instructional Support, Emotional Support, and/or Classroom Organization compared with classrooms rated as low on these domains (Burchinal, 2018; Hatfield et al., 2016). Using data from a study of prekindergarten programs spanning 11 states, Burchinal et al. (2010) found stronger associations between teachers' Instructional Support and children's academic achievement in higher-quality prekindergarten classrooms than in lower-quality classrooms. However, a recent study examining thresholds with the Pre-K CLASS in nationally representative Head Start data did not find evidence of thresholds (Nguyen et al., 2022).

Published studies have yet to examine thresholds using the ECERS-3, in part because it is relatively new. Consequently, data are limited. However, Burchinal et al. (2016) examined the thresholds of the ECERS-R. In a meta-analysis across six datasets, the authors found that the ECERS-R Interaction score was a stronger positive predictor of children's receptive language, and social competence scores in higher-quality classrooms than it was in lower-quality classrooms.

The current study

The current study examines the associations between two widely used measures of classroom quality and children's school readiness outcomes: the Pre-K CLASS and the ECERS-3. We reanalyze data from the ECERS-3 Validation Study by Early and colleagues (2018) that is publicly available on OpenICPSR. The study includes more than 100 classrooms across three states observed using the ECERS-3 and Pre-K CLASS, as well as direct assessments and teacher-reports of more than 400 children's outcomes. We first examine the linear associations with children's school readiness (that is, cognitive, social-emotional, and executive function outcomes in pre-K) to see whether we replicate the original study findings. We then expand the original study by examining thresholds in the associations between classroom quality and growth in children's outcomes. This is the first study we are aware of that has examined thresholds using the ECERS-3, and few studies have examined threshold findings with the Pre-K CLASS. More research is needed to determine the appropriateness of benchmarks for determining quality with both measures. In particular, we examine associations with children's outcomes for multiple thresholds, including some Pre-K CLASS thresholds that Head Start programs are required to meet to avoid having to compete for continued funding (ECLKC, 2023). The study adds to the literature by addressing two overarching research questions: (1) What are the associations between measures of classroom quality and growth in children's school readiness outcomes? (2) Are there thresholds at which the associations between classroom quality measures and growth in children's school readiness are stronger (or weaker)? We hypothesized that any linear associations between quality and child outcomes would be small and that there would be some evidence of associations between quality and child outcomes in higher-quality classrooms that meet certain thresholds.

Material and methods

Data

We used data from Early and colleagues' (2018) ECERS-3 Validation Study, which were collected in the 2015–2016 academic year in three states (Georgia, Pennsylvania, and Washington) as part of their

QRIS data collection. More information is documented in Early et al. (2018). We provide a brief overview below.

As part of the data collection, each state was asked to recruit 40 centers to participate in the collection of classroom observations and child outcomes. This sample included 119 randomly selected classrooms, one from each of the 119 centers. Five children in each classroom were selected at random from all eligible children in the room. During the pretest in the fall, 575 children participated. During the posttest in the spring, 491 children (85%) from the pretest participated. For the primary results, we focus on the classrooms that had either the ECERS-3 or Pre-K CLASS data collected ($n = 119$) and children in these classrooms who had direct assessments and teacher-reported assessments in the fall and spring. We imputed 20 datasets using multiple imputation via chained equations in Stata with all study variables. We imputed continuous variables using predictive mean matching, categorical variables using multinomial logistic regression, and binary variables using logistic regression. We then conducted analyses using the sample with non-missing outcome data for each of the six outcomes, following Von Hippel's (2007) multiple imputation then deletion approach. Multiple imputation then deletion can reduce noise by reducing the risk of including problematic imputations and usually offers more efficient estimates than an ordinary multiple imputation approach because it reduces the variance between imputed datasets (Von Hippel, 2007). Because imputed outcomes do not contain new information about the association between predictors and the outcomes, there is limited benefit to including these imputed outcome values (Von Hippel, 2007). The multiply imputed sample included children with each spring direct assessment ($n = 487–491$) and teacher-reported assessment ($n = 454$). The final classroom analytic sample ranged from 108 to 119 classrooms.

We conducted sensitivity analyses that used a different missing data approach to see whether estimates were sensitive to the missing data approach. We included only children with information about race or ethnicity ($n = 484$), home language ($n = 478$), the primary parent's (typically, the mother) years of education ($n = 470$), teacher's race or ethnicity ($n = 466$), and classroom auspice ($n = 456$). We used missing data dummy codes for categorical variables with high rates of missing data (Individualized Education Plan status = 9% missing, poverty status = 10% missing) to decrease the amount of missing data. Including a separate missing value dummy variable allows children missing that information to have a different mean outcome value than children with nonmissing data without influencing the relationship between the other categories and the outcome. If missing information is not missing at random, the coefficient on the missing value category can account for that and provide a more accurate estimate. However, we could not use missing dummy variables when there were very small amounts of missing data (i.e., on race or ethnicity, home language, the primary parent's years of education, teacher's race or ethnicity, and classroom auspice). The sample included for the "complete case" analyses for each of the six outcomes varied based on whether children had at least one direct assessment ($n = 458–464$) or teacher-reported assessment ($n = 422$) in both the fall and spring.

Measures

Classroom quality

This study uses two measures of classroom quality: the ECERS-3 and the Pre-K CLASS. Extensive procedures were in place to train observers and establish reliability for both measures across all three states (for details, see Early et al., 2018). Observations were conducted between November 2015 and April 2016 and were generally conducted on the same day (81%) by two independent observers. When they could not be scheduled on the same day, they were almost always conducted within 3 days of each other (18%).

The ECERS-3 is designed for use with classrooms that mostly serve children ages 3 to 5. The ECERS-3 includes 35 items that the developers organized into six subscales (Health Practices, Space for Gross Motor Play, Staff–Child Interactions, Individualized Teaching and Learning, Understanding Written Numbers, and Becoming Familiar with Print) and a total score. Early et al. (2018) conducted a factor analysis of the 35 items to derive four subscales: Learning Opportunities, Gross Motor,

Teacher Interactions, and Math Supports. The item scores for the ECERS-3 are based on a set of yes/no indicators under each item. Scores are based on a scale of 1–7, with 1 as inadequate quality, 3 as minimal quality, 5 as good quality, and 7 as excellent quality. Observations generally lasted for 3 hours, as recommended by the ECERS-3 developers. All observers were trained and certified as reliable by the ECERS-3 publishers. Paired observations were conducted for 86 visits in which data collectors agreed on consensus scores. On average, 91% (SD = 6) of their original scores were within one scale point of the consensus score (Early et al., 2018).

The second measure of classroom quality, the Pre-K CLASS, is a widely used classroom quality observation tool that focuses on teacher andchild interactions. It has three domains: Emotional Support, Instructional Support, and Classroom Organization. Each domain comprises multiple dimensions. The Pre-K CLASS is measured on a 7-point scale, with higher scores indicating higher quality. Scores of 1 or 2 indicate the classroom is low quality; 3, 4, or 5 indicate the classroom is in the mid-range of quality; and 6 or 7 indicate the classroom is high quality. Observers rated the classrooms on the dimensions about every 30 min during the morning of the observation. Although the goal was to collect six 30-min observation cycles in each classroom, about 26% of the classrooms had between four and five observation cycles. All observers were trained and certified as reliable by the CLASS publishers. Paired observations were conducted for 11 visits. On average, 95% (SD = 6) of the scores were within one scale point of each other (Early et al., 2018).

Children's outcomes

Cognitive skills. Direct assessments of children's cognitive skills were conducted in the fall and spring of the school year. Subtests from the Woodcock-Johnson IV Tests of Cognitive Abilities (Schrank et al., 2014) were used to measure children's cognitive skills. The Picture Vocabulary subtest was used to assess children's expressive vocabulary skills by presenting them with a series of increasingly complex images and asking them to name the image they were looking at. The Letter-Word Identification subtest measures children's literacy skills by asking them to first identify single letters and then increasingly complex words. The Applied Problems subtest measures children's ability to analyze and solve practical problems using basic math skills, such as counting, addition, and subtraction. We used W scores instead of standard scores in the analysis because they have more variability (Levine et al., 2001).

Social-emotional skills. One teacher-reported measure was used to assess children's social-emotional skills. Teachers completed 38 items from the *Devereux Early Childhood Assessment Preschool Program, 2nd Edition* (DECA-P2; LeBuffe & Naglieri, 2013), indicating how often the child demonstrated certain behaviors over the past 4 weeks. Teachers reported this on a 5-point scale ranging from "never" to "very frequently." The DECA-P2 provides two scores: (1) the Total Protective Factors score, based on 27 items that assess the child's social skills, including initiative, self-regulation, and ability to form and maintain positive connections with others; and (2) the Behavioral Concerns score, based on 11 items that assess the extent to which the child exhibits behavioral problems that might require referral or intervention. In the current sample, the Cronbach's alpha was 0.95 for Total Protective Factors and 0.87 for Behavioral Concerns.

Executive function. Assessors administered the Head-Toes-Knees-Shoulders (HTKS; Ponitz et al., 2008) task, in which children are asked to play a game where they are first asked to follow the spoken instructions to touch a body part, and then asked to switch and touch a body part that contradicts what they were told. The rules become more complex and increase in difficulty as children progress through the task. The HTKS is an assessment of children's executive function, including inhibitory control (children must inhibit the dominant response of imitating the assessor), working memory (children must remember the rules of the task), and attentional focusing (children must focus attention to the directions being presented by the assessor). If the child produces the correct response immediately, their score is 2. If they self-correct immediately, without prompting, their score is 1. If they do not

touch the correct part of their body at all, their score is 0. The developers report Cronbach's alphas between .92 and .94 (McClelland et al., 2014).

Child, family, and teacher covariates

To address the fact that children with different characteristics may be more likely to be in high- or low-quality classrooms, we accounted for several baseline characteristics of children, families, teachers, and classrooms. We controlled for the child's age in months at the time of the fall assessment, the months between the fall and spring assessments, their race or ethnicity (Black, non-Hispanic; Hispanic; other, non-Hispanic, or missing; White; with White as the reference group), gender, whether the child only speaks English at home, whether the child has an IEP or whether this information is missing, the state the child lives in, parent's years of education, teacher's race or ethnicity (Black, non-Hispanic; Hispanic; other, non-Hispanic, or missing; White; with White as the reference group), whether the teacher has a bachelor's degree or higher, whether children are in households with incomes below 185% of the poverty line or whose information on poverty status is missing; and whether the child is in a Head Start or state pre-K program. Finally, we controlled for the fall score on the respective outcomes to see whether classroom quality is associated with changes or gains in children's school readiness outcomes.

Sample description

Table 1 presents descriptive information about the sample. On average, the children were 52 months old in the fall, their parents had at least a high-school diploma (14 years of education) and they spoke only English at home. About half were from families with incomes at 185% of the poverty line or less. Most of the children in our sample were either White/Caucasian, non-Hispanic/Latinx, or of multiple races. The majority of teachers in the sample were White/Caucasian, non-Hispanic/Latinx, and more than half of the classrooms had lead teachers with a bachelor's degree or higher. Almost half of the classrooms were part of a state-funded pre-K program.

There were some differences in the characteristics of children and teachers, and of classrooms that were above and below the median on quality on the ECERS-3 total score based on independent sample t-tests. In classrooms above the median on quality compared with classrooms below the median on quality, children's parents had fewer years of education, there were fewer children in poverty (48 versus 50%), fewer children spoke only English at home (90 versus 94%), fewer children were missing information about having an IEP (9 versus 10%), there were fewer Black/African American, non-Hispanic/Latinx teachers (8 versus 18%), fewer Hispanic/Latinx teachers (2 versus 6%), more White/Caucasian, non-Hispanic/Latinx teachers (86 versus 69%), more teachers with a BA or higher (64 versus 51%) and classrooms were more likely to be Head Start (12 versus 4%) or state-funded pre-K (53 versus 40%).

Analysis approach

We first examined the association between the classroom quality measures and growth in children's outcomes by estimating two-level hierarchical linear models of children (Level 1) nested within classrooms (Level 2). Because each center only included one classroom, children are nested within classrooms and centers. Our key classroom quality predictors are the ECERS-3 and Pre-K CLASS scores. For the ECERS-3, we used three of the four subscales derived from Early et al. (2018), including Learning Opportunities, Teacher Interactions, and Math Supports. The authors also derived a fourth subscale, Gross Motor, which we do not include in our analyses because we did not hypothesize it to be related to the outcomes of interest. For the Pre-K CLASS, we used the three domain scores. Each quality subscale was entered separately into regressions predicting each child's outcome.

We used a different approach to conducting this analysis than the one used in the original study by Early et al. (2018). Specifically, we used an imputation and then deletion model for imputing outcome

Table 1. Sample characteristics of children in above and below median quality classrooms.

	Children in all classrooms		Children in below median quality classrooms		Children in above median quality classrooms		Above/below median comparison
	Mean or %	SD	Mean or %	SD	Mean or %	SD	p-value
Child and family characteristics							
Age at pretest (months)	51.92	6.26	51.56	6.40	52.25	6.13	.231
Months between pretest and posttest	6.93	0.42	6.93	0.42	6.90	0.42	.454
Male child	49%		51%		48%		.465
Parent education (years)	14.44	2.33	14.80	2.35	14.12	2.27	.002**
Family poverty							
185% of Federal Poverty Line or below	46%		50%		48%		.009**
Missing	10%		10%		9%		.899
Child race/ethnicity							
Black/African American, non-Hispanic/Latino	17%		18%		15%		.320
Latino/Hispanic	6%		5%		7%		.159
White/Caucasian, non-Hispanic/Latino	58%		61%		55%		.245
Multiple races or others (including Asian and Native American), non-Hispanic/Latino	20%		17%		22%		.121
Speaks only English at home	92%		94%		90%		.048*
Individualized Education Plan (IEP)							
Has IEP	6%		5%		6%		.338
Missing IEP	9%		10%		9%		.015*
Teacher and classroom characteristics							
Teacher race/ethnicity							
Black/African American, non-Hispanic/Latino	13%		18%		8%		< .001***
Latino/Hispanic	4%		6%		2%		.022*
White/Caucasian, non-Hispanic/Latino	77%		69%		86%		< .001***
Multiple races or others (including Asian and Native American)	5%		6%		5%		.540
Lead teacher has a BA or higher	58%		51%		64%		.005**
Classroom auspice							
Head Start	8%		4%		12%		< .001***
State-funded pre-K	47%		40%		53%		.005**
State							
State 1	36%		35%		37%		.597
State 2	33%		30%		35%		.193
State 3	31%		35%		28%		.062
n	464		223		241		

In this table, the cut point for classroom quality below and above the median is based on the ECERS-3 total score, which is a score of 3.41. Standard deviations are only presented for continuous variables. According to Early et al. (2018), "Classrooms can belong to multiple classroom auspice categories. Classrooms with blended funding were counted as Head Start and/or state-funded pre-K if any enrolled children were funded with those sources. In a public school refers to a physical location in public school building where older children were also attending."

data as suggested by Von Hippel (2007) instead of including imputed outcome data. In addition, we used a two-level model (with children nested within classrooms) instead of a three-level model (with time nested within children nested within classrooms) because children's outcomes were only measured at two follow-up points. Singer and Willett (2003) recommend there are at least three follow-up points for multilevel longitudinal analyses. This analysis allowed us to see how robust the original study findings were to different analytic decisions.

We then examined whether classroom quality scores need to reach a particular level, or threshold, to be associated with growth in child outcomes. Specifically, we fit spline regression models in which we estimated two linear slopes. The first slope was for the lower-quality range, and the second slope was for the higher-quality range. We chose the median of each classroom quality score to be the "knot," or threshold, at which the slope was allowed to change. Since there is no agreement upon the best method for selecting quality thresholds (Weiland et al., 2013; Zaslow et al., 2011), we examined

the distribution of our data and chose the median cut points for our quality measures (Table 2). The chosen cut points corresponded approximately to the mid-range of quality for all classroom quality measures. One exception is Pre-K CLASS Instructional Support. With a cut point of 2.17, it is in the low-quality range, consistent with prior literature documenting lower scores on this domain (Burchinal, 2018).

We also present sensitivity analyses using predetermined quality thresholds. For these analyses, we started with conceptually chosen thresholds that have been used in other studies (4.5 for the ECERS-3 scales, 2.75 for Pre-K CLASS Instructional Support, and 5 for Pre-K CLASS Emotional Support and Classroom Organization; Burchinal et al., 2011; Weiland et al., 2013). For distributional reasons, we then adjusted the conceptual cut points for some of the scales (3.5 for ECERS-3 Interactions, 2 for ECERS-3 Math Activities, 2.5 for Pre-K CLASS Instructional Support, and 5.5 for Pre-K CLASS Emotional Support; results using these thresholds are presented in Supplemental Table 3). The conceptual cut points used for Instructional Support and Classroom Organization are the same as those that Head Start classrooms are required to meet to avoid having to compete for continued funding (ECLKC, 2023). We could not examine the Head Start-relevant cutoff of 5 for Emotional Support because 85% of the classrooms were above that threshold. Although the conceptual cut points we used for the ECERS-3 Interactions and Math Activities score vary substantially from the 4.5 used for the ECERS-R in other studies, other threshold analyses have not used the ECERS-3, which has been found to have a lower mean (Hestenes et al., 2019). Ultimately, the median cut points provided the most power for our analyses by creating equal-sized subgroups. We also conducted all analyses using complete case analysis with missing dummies as a sensitivity check on the primary results.

To address concerns that classroom quality may be associated with other family, child, and classroom characteristics, we included – in all of the models described above – control variables that are associated with both quality and outcomes. All key variables of interest were standardized to have a mean of 0 and a standard deviation of 1 so coefficients can be interpreted as effect sizes in standard deviation units.

Results

Table 2 presents the descriptive statistics for the ECERS-3 total and subscale scores and the Pre-K CLASS total and domain scores. The median for the ECERS-3 total score was 3.41, which is considered to be between minimal and good quality. Median scores for the ECERS-3 subscales were also in the minimal to good range of quality, with 2.91 for Math Activities, 3.40 for Learning Opportunities, and 4.20 for Interactions subscales. The median for the Pre-K CLASS total score was 4.51, which is considered to be in the mid-range of quality. Median scores for the Pre-K CLASS domains were in the low to middle ranges of quality, with 2.17 for Instructional Support, 5.44 for Classroom Organization, and 5.75 for Emotional Support. Table 3 presents the descriptive statistics for fall and spring child outcomes. Scores on all six outcomes increased from fall to spring. This indicates children are making progress across the year, but teachers' ratings of their behavioral concerns are also

Table 2. Descriptive statistics of classroom quality scores.

	Mean	Median	SD	Range (Min-Max)
ECERS-3 (*n* = 119)				
Total	3.40	3.41	0.76	1.66–5.15
Interactions	3.65	4.20	1.02	1.02–5.73
Math Activities	2.32	2.91	1.04	1.00–6.67
Learning Opportunities	3.51	3.40	0.95	1.18–5.36
CLASS (*n* = 118)				
Total	4.41	4.51	0.74	2.62–6.13
Instructional Support	2.26	2.17	0.69	1.00–5.06
Emotional Support	5.68	5.75	0.79	3.63–7.00
Classroom Organization	5.29	5.44	1.04	2.75–6.92

Table 3. Descriptive statistics of child outcomes in the fall and spring.

	Fall			Spring			
	Mean	SD	Range (Min-Max)	Mean	SD	Range (Min-Max)	n
WJ IV Picture Vocabulary	456.84	13.67	374–489	462.79	12.13	374–495	464
WJ IV Letter-Word	329.84	25.79	272–449	346.51	28.84	272–506	464
WJ IV Applied Problems	402.05	23.48	324–448	415.97	19.14	341–466	463
HTKS	19.14	25.16	0–90	34.38	32.10	0–94	458
DECA Total Protective Factors	51.13	9.98	28–72	53.44	9.70	28–72	422
DECA Behavioral Concerns	48.16	10.30	29–72	48.73	10.12	29–72	422

DECA = Devereux Early Childhood Assessment; HTKS = Head-Toes-Knees-Shoulders; WJ = Woodcock-Johnson.

Table 4. Correlations between quality scales and child outcomes in the spring.

	1	2	3	4	5	6	7	8	9	10	11
1. ECERS-3: Interactions											
2. ECERS-3: Math Activities	.56***										
3. ECERS-3: Learning Opportunities	.77***	.65***									
4. CLASS: Instructional Support	.38***	.26***	.44***								
5. CLASS: Emotional Support	.54***	.41***	.52***	.54***							
6. CLASS: Classroom Organization	.41***	.33***	.43***	.52***	.86***						
7. WJ IV Picture Vocabulary: Spring	.02	.09+	.07	.05	.08+	.10*					
8. WJ IV Letter-Word: Spring	−.04	.07	.04	−.00	.00	.05	.47***				
9. WJ IV Applied Problems: Spring	.08+	.14**	.18***	.13**	.18***	.20***	.56***	.54***			
10. HTKS: Spring	.04	.05	.11*	.11*	.16***	.17***	.46***	.44***	.60***		
11. DECA Total Protective Factors: Spring	.07	.01	.06	.06	.10*	.16***	.26***	.23***	.34***	.31***	
12. DECA Behavioral Concerns: Spring	−.10*	−.04	−.05	−.09+	−.12**	−.16***	−.13**	−.18***	−.22***	−.22***	−.67***

n = 446–570. DECA = Devereux Early Childhood Assessment; HTKS = Head-Toes-Knees-Shoulders; WJ = Woodcock-Johnson.
$^{+}p < .10$; $*p < .05$; $**p < .01$; $***p < .001$.

increasing. Table 4 shows the bivariate correlations between classroom quality and children's outcomes. Because DECA Total Protective Factors and Behavioral Concerns were correlated at .67, and because the significance of their correlations with children's outcomes were similar, we only included DECA Behavioral Concerns in analyses to reduce the number of comparisons examined.

Results of the linear model

The linear model results, which allow us to examine the associations between the various classroom quality scores and school readiness outcomes, are presented in Table 5. Absolute effect sizes were all smaller than 0.11. For the ECERS-3, one of 15 associations was significant: higher scores on Interactions were associated with lower scores on behavior problems (as measured by DECA Behavioral Concerns). For the Pre-K CLASS, none of the predictors were statistically significantly associated with child outcomes. However, there were marginal associations between the Emotional Support score and higher math skills (as measured by WJ IV Applied Problems) and executive function skills (as measured by the Head-Toes-Knees-Shoulders task) in children. In addition, there were marginal associations between Classroom Organization and lower scores on behavior problems

(as measured by DECA Behavioral Concerns). These findings were consistent in the complete case analyses, except that the association between ECERS-3 Interactions and behavioral problems became only marginally significant (Supplemental Table 1).

Results of the threshold model

Table 6 has the results from estimating the spline regression models, with the spline knot set equal to the median cut point (Table 2) of each respective classroom quality measure. The "Below median quality" column corresponds to the slopes and standard errors for classrooms below the median quality level on the corresponding measure. The "Above median quality" column corresponds to the slopes and standard errors for classrooms above the median quality level on the corresponding measure. The "Differ?" column reveals whether the difference between the slopes for the classrooms below and above median quality was statistically significant.

Analyses using the ECERS-3 provide some evidence of thresholds for Math Activities and Learning Opportunities scores and children's cognitive outcomes, with effect sizes between 0.09 and 0.15. Among the median-split spline regression models for the ECERS-3, the ECERS-3 Math Activities and Learning Opportunities subscales were stronger positive predictors of literacy skills (WJ Letter-Word) for children in classrooms above median quality than they were in below median quality classrooms (Math Activities: b = 0.15, $p < .05$; difference between slopes significant at $p < .05$; Learning Opportunities: b = 0.09, $p < .10$; difference between slopes significant at $p < .10$). Although not statistically significant, there were trends that suggested larger coefficients in classrooms above the median of quality on Math Activities scores. There were no significant thresholds for the ECERS-3 and social-emotional or executive function outcomes. In addition, there were no significant threshold findings for the ECERS-3 using the conceptual cut points (Supplemental Table 3).

Analyses using the Pre-K CLASS provided evidence of thresholds for Classroom Organization and teacher-reported social-emotional outcomes, but no evidence of thresholds for any of the other quality domains and outcomes. The Classroom Organization domain was a stronger predictor of lower behavioral problems (DECA Behavioral Concerns) (b = −0.47, $p < .001$; difference between slopes significant at $p < .01$) in above median quality classrooms than it was in below median quality classrooms. The finding of stronger associations between Classroom Organization and lower behavioral problems in higher-quality classrooms was also evident using the conceptual cut points (Supplemental Table 3).

All threshold results were consistent with the complete case results (Supplemental Table 2). In addition, examinations of scatterplots (Supplemental Figures 1–4) suggest that threshold results were not driven by outliers.

Discussion

Overall, we detected one statistically significant linear association (of 30) between ECERS-3 Interactions and decreases in behavioral problems. We did not detect any other statistically significant associations between the ECERS-3 or Pre-K CLASS and growth in children's cognitive, social-emotional, or executive function outcomes. We did detect three (of 30) small threshold findings. These show that there were associations between Math Activities and Learning Opportunities scores and children's literacy skills when quality was above the median and that there were no associations between quality and child outcomes in classrooms below these thresholds. Although associations were small, this provides the first evidence of thresholds for the ECERS-3. We also found evidence that suggests Pre-K CLASS Classroom Organization scores were associated with children's social-emotional outcomes in classrooms above certain thresholds. However, we found no evidence of thresholds for other quality subscales or outcomes, or for most conceptually selected thresholds, suggesting that thresholds are limited. We discuss these findings in turn and then describe limitations, including that threshold associations could be driven by the fact that children in higher-quality

Table 5. Results of regressing school readiness outcomes on classroom quality measures with multiply imputed data.

	WJ IV Picture Vocabulary	WJ IV Letter-Word	WJ IV Applied Problems	HTKS	DECA Behavioral Concerns
ECERS-3					
Interactions	0.03	0.01	0.02	0.03	−0.10*
	(0.03)	(0.02)	(0.04)	(0.03)	(0.05)
Math Activities	0.05	0.03	0.06	0.04	−0.03
	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Learning Opportunities	0.02	0.04	0.04	0.04	0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Pre-K CLASS					
Instructional Support	0.03	0.03	0.04	0.05	−0.01
	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)
Emotional Support	0.00	−0.01	0.06+	0.07+	−0.08
	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)
Classroom Organization	0.02	0.01	0.05	0.06	−0.10+
	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)
N (child)	491	491	490	487	454
N (classroom)	119	119	119	119	108

Standard errors in parentheses. All models control for child, family, teacher, and classroom characteristics, including the fall score on the respective outcome, state in which child lives, parent years of education, race/ethnicity, gender, age in months at time of spring assessment, months between fall and spring assessments, IEP status, home language, parent years of education, poverty status, teacher race/ethnicity and education, and type of program. To account for missing data, all predictor variables were imputed using multiple imputation via chained equations. For all models, the parameter estimates have been standardized so that they represent the amount of growth on the dependent variable, in standard deviations, associated with a one standard deviation change on the quality measure. DECA = Devereux Early Childhood Assessment; ECERS-3 = Early Childhood Environment Rating Scale Third Edition; HTKS = Head Toes Knees Shoulders; Pre-K CLASS = Classroom Assessment Scoring System WJ = Woodcock-Johnson Tests of Achievement. $^+p < .10$; $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$.

classrooms have different characteristics that could contribute to associations between quality and outcomes.

Linear associations with classroom quality

We found one significant linear association (of ECERS-3 Interactions with lower behavioral problems). This finding differs somewhat from the results in Early et al. (2018), which found 3 different statistically significant results out of the 15 we tested for the ECERS-3 and 3 statistically significant results out of the 15 we tested for the Pre-K CLASS. These differences may be the result of using a two-level model that controls for fall scores, instead of a three-level model with scores nested within children, or restricting our sample to children with data on outcomes. We did this because we wanted to avoid using imputed values of the dependent variables to reduce noise (Von Hippel, 2007). Therefore, our sample was smaller than the one used in Early et al. (2018), which may contribute to different patterns of significance. These results indicate that the significance of the original study findings was sensitive to analytic decisions, which limits their reproducibility. Despite differences in the patterns of significance, both studies found effect sizes below 0.11. This is consistent with recent research that finds weak associations between classroom quality measures and children's school readiness outcomes (Aikens et al., 2021; Brunsek et al., 2017; Guerrero-Rosada et al., 2021; Nguyen et al., 2022; Perlman et al., 2016; Sabol & Pianta, 2014).

Thresholds for the ECERS-3

We found two threshold findings that indicate the ECERS-3 was associated with growth in children's outcomes in higher-quality classrooms, although associations were small. Specifically, the ECERS-3 Math Activities and Learning Opportunities scores were significantly stronger predictors of growth in children's literacy scores in classrooms above the median on quality than in classrooms below the

Table 6. Spline regression results, using median-split cut points with multiply imputed data.

	WJ IV Picture Vocabulary			WJ IV Letter-Word			WJ IV Applied Problems			HTKS			DECA Behavioral Concerns		
	Below median quality	Above median quality	Differ?	Below median quality	Above median quality	Differ?	Below median quality	Above median quality	Differ?	Below median quality	Above median quality	Differ?	Below median quality	Above median quality	Differ?
ECERS-3															
Interactions	−0.04	−0.02		−0.02	0.03		−0.01	0.01		−0.04	−0.06		−0.02	−0.11	
	(0.03)	(0.06)		(0.02)	(0.05)		(0.03)	(0.07)		(0.03)	(0.06)		(0.04)	(0.09)	
Math Activities	0.02	0.12		0.02	0.15*	*	0.00	0.10		−0.03	0.01		−0.02	−0.07	
	(0.04)	(0.08)		(0.03)	(0.07)		(0.04)	(0.08)		(0.04)	(0.08)		(0.05)	(0.12)	
Learning Opportunities	−0.01	0.01		0.01	0.09+	+	−0.03	0.00		0.00	0.03		0.04	0.08	
	(0.03)	(0.06)		(0.03)	(0.05)		(0.03)	(0.06)		(0.03)	(0.06)		(0.04)	(0.08)	
Pre-K CLASS															
Instructional Support	−0.02	0.00		0.02	0.12		−0.01	0.05		−0.04	0.01		−0.03	−0.04	
	(0.05)	(0.09)		(0.04)	(0.08)		(0.05)	(0.10)		(0.05)	(0.10)		(0.07)	(0.13)	
Emotional Support	−0.02	−0.11		0.00	0.00		0.01	0.14		−0.03	−0.13		0.01	−0.08	
	(0.02)	(0.14)		(0.02)	(0.12)		(0.02)	(0.15)		(0.02)	(0.14)		(0.03)	(0.20)	
Classroom Organization	0.01	0.07		−0.01	−0.05		−0.04+	−0.14		−0.01	−0.05		−0.06+	−0.47***	***
	(0.02)	(0.11)		(0.02)	(0.10)		(0.02)	(0.12)		(0.02)	(0.12)		(0.03)	(0.16)	
N (child)	491			491			490			487			454		
N (classroom)	119			119			119			119			108		

Standard errors in parentheses. The "Differ?" column tests whether the slopes for the below and above median quality classrooms are statistically significantly different from each other. All models control for child, family, teacher, and classroom characteristics, including the fall score on the respective outcome, state in which child lives, race/ethnicity, gender, age in months at time of spring assessment, months between fall and spring assessments, IEP status, home language, parent years of education, poverty status, teacher race/ethnicity and education, and type of program. To account for missing data, all predictor variables were imputed using multiple imputation via chained equations. For all models, the parameter estimates have been standardized so that they represent the amount of growth on the dependent variable, in standard deviations, associated with a one standard deviation change on the quality measure. DECA = Devereux Early Childhood Assessment; ECERS-3 = Early Childhood Environment Rating Scale Third Edition; HTKS = Head Toes Knees Shoulders; Pre-K CLASS = Classroom Assessment Scoring System WJ = Woodcock-Johnson Tests of Achievement. $^+p < .10$; $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$.

median on quality. ECERS-3 Learning Opportunities and Math Activities scores were associated with children's literacy scores at least at the trend level in classrooms above the median on quality, suggesting that quality as measured by the ECERS-3 is associated with children's cognitive outcomes only in classrooms that reach a certain level of quality.

Although the ECERS-3 subscales are different from the ECERS-R subscales, Burchinal et al. (2016) found thresholds in the association between the acquisition of language and social skills and the ECERS-R Interaction factor. This provides some limited evidence that classrooms may need to reach a certain level of quality if we are to detect associations with child outcomes. However, we did not find thresholds for other ECERS-3 subscale scores or other child outcomes, indicating that threshold associations are limited. In addition, any observed threshold associations could be because of differences in the characteristics of children and families in the higher-quality classrooms, as discussed more in the limitations. Additional research using the ECERS-3 and future updated versions of the measure can replicate and build on these initial findings.

Thresholds for the Pre-K CLASS

We found one threshold finding for the Classroom Organization domain of the Pre-K CLASS. Specifically, compared with classrooms in the lower range of quality, higher levels of Classroom Organization were more strongly associated with teacher-reported decreases in children's behavioral problems in classrooms that ranked above the median on quality. It may be that classrooms need to have a certain level of order and structure to promote children's social-emotional well-being. Alternatively, classrooms in which children were reported as better-behaved may be more likely to be rated as high on Classroom Organization. In previous studies, threshold associations were found for Classroom Organization and children's cognitive inhibitory control (Weiland et al., 2013). However, another study found no evidence of thresholds for any Pre-K CLASS domains and children's social-emotional outcomes (Nguyen et al., 2022). We also did not replicate past findings of thresholds for Instructional Support and children's language and literacy outcomes (Burchinal et al., 2016) or Emotional Support and children's literacy and social-emotional outcomes (Hatfield et al., 2016). We also did not find any evidence for the thresholds used to determine whether Head Start programs need to compete for renewal. Threshold associations may vary across studies due to differences in the thresholds used, the distribution of Pre-K CLASS scores in the sample, or the different outcomes examined.

Potential explanations for limited findings

Although we found some small associations between classroom quality and children's outcomes in higher-quality classrooms, most linear and threshold associations were null. The limited associations between classroom quality and child outcomes may be because of limitations in the measures of quality. For example, classroom quality scores may vary because of the time of day or day of the week when classrooms are observed, the observers, and the activities going on in the classroom (Finders et al., 2021; Mashburn, 2017). Limited inter-rater reliability of scores could also limit the utility of these scores. Although observers were highly trained, past research has found that observers accounted for about 5–16% of the variance in Pre-K CLASS domain scores in kindergarten (Mantzicopoulos et al. 2018). In addition, classroom quality scores often have a restricted range, which makes it difficult to differentiate between most classrooms (Aikens et al., 2016; Gordon & Peng, 2020). Moreover, the Pre-K CLASS and ECERS-3 are both global measures of quality and may not measure the elements of classroom quality that improve children's cognitive outcomes, such as the content being taught (Burchinal et al., 2010; Zaslow et al., 2010). These measures may also not be culturally relevant for students and teachers of color (Curenton et al., 2020). Finally, both measures capture quality measured at the classroom level, whereas the quality of what an individual child experiences in the classroom

may be more predictive of school readiness outcomes (Burchinal et al., 2021; Hanno et al., 2021; Pianta et al., 2020).

Given recent evidence of limited associations between measures of classroom quality and children's outcomes, researchers are emphasizing a need for new measures of classroom quality (Maier et al., 2020), including measures that are more culturally appropriate (Curenton et al., 2020). Future researchers could consider what should be measured more carefully to refine hypotheses. For example, measures of classroom quality such as the Pre-K CLASS and the ECERS-3 may be capturing broader classroom experiences that are not directly aligned with the measured outcomes. Associations might be detected if the observational tools used were more proximal to students' classroom experiences and aligned directly to the curricular and instructional focus of their classrooms.

Limitations

This research provides some limited evidence of thresholds in the association between the two most used classroom quality observations and children's cognitive and social-emotional outcomes, although most associations were null and all associations were small. Unobserved variables could explain the associations between classroom quality and children's outcomes in higher-quality classrooms, such as systematic differences in children's family and neighborhood environments outside of the classroom. Another limitation is that small sample sizes can limit our ability to detect statistically significant associations. The sample size was further limited by the decision to not impute missing spring outcome data to reduce noise. To maximize the power to detect threshold associations, we used median split cut points that typically represented the midpoint of quality. Although we also examined predetermined thresholds, the distribution of our data meant using the exact thresholds from prior research was not possible because only a small proportion of classrooms met the thresholds used in prior research, especially for the ECERS-3 that has a different distribution to the ECERS-R (Hestenes et al., 2019). This means that our findings cannot support specific quality thresholds found in prior research. In addition, the reliability checks on the Pre-K CLASS were limited so poor measurement could have contributed to limited associations. Finally, these data are not nationally representative: they are limited to convenience samples of children in three states who were assessed in English. Future research using different and larger samples can strengthen the research on quality thresholds. Early childhood monitoring systems may want to consider whether and how to set thresholds for quality given limited associations between observed quality and children's outcomes.

Disclosure statement

Funding

ORCID

Jessica F. Harding http://orcid.org/0000-0002-7052-6660

References

Aikens, N., Bush, C., Gleason, P., Malone, L., & Tarullo, L. (2016, October). *Tracking quality in head start classrooms: FACES 2006 to FACES 2014*. OPRE Brief #2016-82. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Aikens, N., Nguyen, T., & Harding, J. F. (2021). *How much does the pre-K CLASS relate to children's readiness for school skills? Early childhood literature scan brief*. Mathematica.

Auger, A., Farkas, G., Burchinal, M. R., Duncan, G. J., & Vandell, D. L. (2014). Preschool center care quality effects on academic achievement: An instrumental variables analysis. *Developmental Psychology*, *50*(12), 2559–2571. https://doi.org/10.1037/a0037995

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In W. Damon & R. M. Lerner (Ed.), *Handbook of child psychology*: *Vol. 1 theoretical models of human development* (5th ed. pp. 994–1028). Wiley.

Brunsek, A., Perlman, M., Falenchuk, O., McMullen, E., Fletcher, B., Shah, P. S., & Manalo, E. (2017). The relationship between the early childhood environment rating scale and its revised form and child outcomes: A systematic review and meta-analysis. *PLOS ONE*, *12*(6), e0178512. https://doi.org/10.1371/journal.pone.0178512

BUILD Initiative. (2021). *Quality compendium*. https://qualitycompendium.org

BUILD Initiative & Child Trends. (2019). *A catalog and comparison of quality initiatives*. http://qualitycompendium.org/

Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives*, *12*(1), 3–9. https://doi.org/10.1111/cdep.12260

Burchinal, M., Garber, K., Foster, T., Bratsch-Hines, M., Franco, X., & Peisner-Feinberg, E. (2021). Relating early care and education quality to preschool outcomes: The same or different models for different outcomes? *Early Childhood Research Quarterly*, *55*, 35–51. https://doi.org/10.1016/j.ecresq.2020.10.005

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

Burchinal, M. R., & Nelson, L. (2000). Family selection and child care experiences: Implications for studies of child outcomes. *Early Childhood Research Quarterly*, *15*(3), 385–411.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, *25*(2), 166–176. https://doi.org/10.1016/j.ecresq.2009.10.004

Burchinal, M., Zaslow, M., Tarullo, L., Votruba-Drzal, E., & Miller, P. (2016). Quality thresholds, features, and dosage in early care and education: Secondary data analyses of child outcomes. *Monographs of the Society for Research in Child Development*, *81*(2), 1–126.

Coley, R. L., Votruba-Drzal, E., Collins, M. A., & Miller, P. (2014). Selection into early education and care settings: Differences by developmental period. *Early Childhood Research Quarterly*, *29*(3), 319–332.

Curenton, S. M., Iruka, I. U., Humphries, M., Jensen, B., Durden, T., Rochester, S. E., Sims, J., Whittaker, J. V., & Kinzie, M. B. (2020). Validity for the assessing classroom sociocultural equity scale (ACSES) in early childhood classrooms. *Early Education & Development*, *31*(2), 284–303. https://doi.org/10.1080/10409289.2019.1611331

Early, D. M., Sideris, J., Neitzel, J., LaForett, D. R., & Nehler, C. G. (2018). Factor structure and validity of the early childhood environment rating scale-third edition (ECERS-3). *Early Childhood Research Quarterly*, *44*, 242–256. https://doi.org/10.1016/j.ecresq.2018.04.009

ECLKC. (2023). *Use of classroom assessment scoring system (CLASS®) in head start*. Retrieved February 2, 2024, from https://eclkc.ohs.acf.hhs.gov/designation-renewal-system/article/use-of-classroom-assessment-scoring-system-class-head-start

Finders, J. K., Budrevich, A., Duncan, R. J., Purpura, D. J., Elicker, J., & Schmitt, S. A. (2021). Variability in preschool CLASS scores and children's school readiness. *AERA Open*, *7*, 23328584211038938. https://doi.org/10.1177/23328584211038938

Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of child care quality and relations to child development. *Developmental Psychology*, *49*(1), 146–160. https://doi.org/10.1037/a0027899

Gordon, R. A., & Peng, F. (2020). Evidence regarding the domains of the CLASS PreK in head start classrooms. *Early Childhood Research Quarterly*, *53*, 23–39. https://doi.org/10.1016/j.ecresq.2020.01.008

Guerrero-Rosada, P., Weiland, C., McCormick, M., Hsueh, J., Sachs, J., Snow, C., & Maier, M. (2021). Null relations between CLASS scores and gains in children's language, math, and executive function skills: A replication and extension study. *Early Childhood Research Quarterly*, *54*, 1–12. https://doi.org/10.1016/j.ecresq.2020.07.009

Hamre, B. K. (2014). Teachers' daily interactions with children: An essential ingredient in effective early childhood programs. *Child Development Perspectives*, *8*(4), 223–230. https://doi.org/10.1111/cdep.12090

Hamre B. K, Pianta R. C, Downer J., Mashburn A. J., Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal*, *113*(4), 461–487. https://doi.org/10.1086/669616

Hanno, E. C., Gonzalez, K. E., Jones, S. M., & Lesaux, N. K. (2021). Linking features of structural and process quality across the landscape of early education and care. *AERA Open*, *7*, 23328584211044519. https://doi.org/10.1177/23328584211044519

Harms, T., & Clifford, R. (1980). *Early childhood environment rating scale (ECERS)*. Teachers College Press.

Harms, T., Clifford, R., & Cryer, D. (Ed.). (2005). *Early childhood environment rating scale* (rev. ed.). (Teachers College Press.

Harms, T., Clifford, R., & Cryer, D. (2015). *Early childhood environment rating scale* (3rd ed.). Teachers College Press.

Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Thresholds in the association between quality of teacher-child interactions and preschool children's school readiness skills. *Early Childhood Research Quarterly*, *36*, 561–571. https://doi.org/10.1016/j.ecresq.2015.09.005

Head Start Early Childhood Learning and Knowledge Center. (2020). *Designation renewal system*. https://eclkc.ohs.acf.hhs.gov/designation-renewal-system

Hestenes, L. L., Rucker, L., Wang, Y. C., Mims, S. U., Hestenes, S. E., & Cassidy, D. J. (2019). A comparison of the ECERS-R and ECERS-3: Different aspects of quality? *Early Education and Development*, *30*(4), 496–510. https://doi.org/10.1080/10409289.2018.1559681

Hillemeier, M. M., Morgan, P. L., Farkas, G., & Maczuga, S. A. (2013). Quality disparities in child care for at-risk children: Comparing head start and non-head start settings. *Maternal and Child Health Journal*, *17*(1), 180–188.

LeBuffe, P. A., & Naglieri, J. A. (2013). The Devereux Early childhood assessment for preschoolers (DECA-P2) (2nd ed.). Kaplan Early Learning Company.

Levine, J. A., Pollack, H., & Comfort, M. E. (2001). Academic and behavioral outcomes among the children of young mothers. *Journal of Marriage & Family*, *63*(2), 355–369.

Maier, M. F., Hsueh, J., & McCormick, M. (2020). *Rethinking classroom quality: What we know and what we are learning*. MDRC.

Mantzicopoulos, P., French, B. F., Patrick, H., Watson, J. S., & Ahn, I. (2018). The stability of kindergarten teachers' effectiveness: A generalizability study comparing the framework for teaching and the classroom assessment scoring system. *Educational Assessment*, *23*(1), 24–46.

Mashburn, A. J. (2017). Evaluating the validity of classroom observations in the head start designation renewal system. *Educational Psychologist*, *52*(1), 38–49. https://doi.org/10.1080/00461520.2016.1207539

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., Burchinal, M., Early, D., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, *79*(3), 732–749. https://doi.org/10.1111/j.1467-8624.2008.01154.x

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in Psychology*, *5*, 3–14. https://doi.org/10.3389/fpsyg.2014.00599

Nguyen, T., Harding, J. F., Harrington, J., & Cannon, J. (2022). *Exploring limited associations of the Pre-K class with children's school readiness in FACES 2014: Exploring associations for subgroups and thresholds of quality*. Mathematica.

NICHD Early Child Care Research Network. (2002). Child-care structure → process → outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science*, *13*(3), 199–206. https://doi.org/10.1111/1467-9280.00438

NICHD Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten? *Child Development*, *74*(4), 976–1005. https://doi.org/10.1111/1467-8624.00582

Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., Shah, P. S., & van Wouwe, J. P. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the classroom assessment scoring system) in early childhood education and care settings and child outcomes. *PLOS ONE*, *11*(12), e0167660. https://doi.org/10.1371/journal.pone.0167660

Pianta, R. C. (2012). *Implementing observation protocols: Lessons for K-12 education from the field of early childhood*. Center for American Progress.

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system manual: Pre-K*. Paul H. Brookes Publishing Co.

Pianta, R. C., Whittaker, J. E., Vitiello, V., Ruzek, E., Ansari, A., Hofkens, T., & DeCoster, J. (2020). Children's school readiness skills across the pre-K year: Associations with teacher-student interactions, teacher practices, and exposure to academic content. *Journal of Applied Developmental Psychology*, *66*, 101084. https://doi.org/10.1016/j.appdev.2019.101084

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, *9*(3), 144–159. https://doi.org/10.1207/s1532480xads0903_2

Ponitz, C. E. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly*, *23*(2), 141–158. https://doi.org/10.1016/j.ecresq.2007.01.004

Sabol, T. J., & Pianta, R. C. (2014). Do standard measures of preschool quality used in statewide policy predict school readiness? *Education Finance and Policy*, *9*(2), 116–164. https://doi.org/10.1162/EDFP_a_00127

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-johnson IV tests of achievement*. Riverside Publishing Company.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). *Issues for the next decade of quality rating and improvement systems (OPRE issue brief No. 9)*. U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Von Hippel, P. T. (2007, 4). Regression with missing YS: An improved strategy for analyzing multiply imputed data. *Sociological Methodology*, *37*(1), 83–117. https://doi.org/10.1111/j.1467-9531.2007.00180.x

Vygotsky, L. S., Cole, M., John-Steiner, V., Scribner, S., & Souberman, E. (1978). *The development of higher psychological processes*. Harvard University Press.

Weiland, C., McCormick, M., Mattera, S., Maier, M., & Morris, P. (2018). Preschool curricula and professional development features for getting to high-quality implementation at scale: A comparative review across five trials. *AERA Open*, *4*(1), 2332858418757735. https://doi.org/10.1177/2332858418757735

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, *28*(2), 199–209. https://doi.org/10.1016/j.ecresq.2012.12.002

Zaslow, M., Burchinal, M., Tarullo, L., Votruba-Drzal, E., & Miller, P. (2016). Quality thresholds, features, and dosage in early care and education: Secondary data analyses of child outcomes. *Monographs of the Society for Research in Child Development*, *81*(2), 1–126. https://doi.org/10.1111/mono.12240

Zaslow, M., Martinez-Beck, I., Tout, K., & Halle, T. (Eds.). (2011). *Quality measurement in early childhood settings*. Brookes Publishing.

Zaslow, M., Tout, K., Halle, T., Whittaker, J. V., & Lavelle, B. (2010). *Toward the identification of features of effective professional development for early childhood educators*. U.S. Department of Education, Office of Planning, Evaluation, and Policy Development.