Exploring the Associations of the Pre-K CLASS with Children's School Readiness in FACES 2014: Associations for Subgroups and Thresholds of Quality

April 2022
Tutrang Nguyen, Jessica F. Harding, Nikki Aikens, Jeffrey Harrington, and Judy Cannon

This page has been left blank for double-sided copying.

Exploring the Associations of the Pre-K CLASS with Children's School Readiness in FACES 2014: Associations for Subgroups and Thresholds of Quality

This page has been left blank for double-sided copying.

Contents

Tables

Figures

Executive Summary

In early care and education (ECE) settings, the quality of a classroom is sometimes linked with how well children learn and develop (Phillips et al. 2017). Head Start, like many state ECE systems, recognizes high-quality teacher–child interactions as important for children's well-being. Research in Head Start classrooms has found that the Pre-K Classroom Assessment Scoring System (Pre-K CLASS; Pianta et al. 2008a) has limited associations with growth in children's cognitive and social emotional outcomes (Gordon and Peng 2020). In this research, we extend these findings to examine whether Pre-K CLASS dimension scores that are conceptually linked to specific child outcomes are most predictive of that outcome (Perlman et al. 2016), whether quality matters more for Head Start children with different background characteristics who may arrive to classrooms with different experiences and needs (Yoshikawa et al. 2013), and whether quality needs to reach a certain threshold to promote Head Start children's development (Burchinal et al. 2016). Specifically, we examine three research questions:

1. What are the associations between the Pre-K CLASS Total score, three domains, and 10 dimensions and gains in children's school readiness outcomes?

2. Do the associations of the Pre-K CLASS and children's school readiness vary for dual language learners, children of color, or children experiencing poverty?

3. Are there thresholds at which the associations between the Pre-K CLASS and children's school readiness are stronger (or weaker)?

We used data on 982 to 1,517 children from the 2014 Head Start Family and Child Experiences Survey (FACES 2014), which provides nationally representative data about Head Start programs, centers, and classrooms and the children and families they serve. We predicted children's spring school readiness outcomes from direct child assessments and teacher reports based on the Pre-K CLASS, controlling for fall scores on the outcome of interest and other child, family, classroom, and teacher covariates. We also examined whether associations between the Pre-K CLASS and children's school readiness outcomes vary for children who are dual language learners, children of color, and children experiencing poverty. Finally, we examined whether classroom quality scores need to reach a particular level, or threshold, to influence children's school readiness outcomes. That is, we examined whether the association between classroom quality and children's outcomes varies in classrooms above and below the median on quality scores.

Overall, we found limited associations between children's school readiness outcomes and the Pre-K CLASS Total score, domain scores, and dimension scores. These associations were limited regardless of whether children were dual language learners, children of color, or experiencing poverty. We also did not find evidence to support that classroom quality needs to reach a certain threshold to be associated with children's outcomes.

We discuss multiple potential explanations for these limited findings, including potential challenges with measuring classroom quality, the content of the Pre-K CLASS, and other analytic considerations. We then present recommendations for measuring classroom quality, including:

- Consider using the Pre-K CLASS in conjunction with other measures

- Reconsider existing definitions and measures of classroom quality to ensure they are culturally responsive

- Consider using the Pre-K CLASS for continuous quality improvement

- Support research and developed of new and revised measures of classroom quality

Although the Pre-K CLASS provides a theoretically grounded approach to defining classroom quality, there is still more work to be done to better understand and improve on this widely administered measure of quality.

I. Introduction

Head Start, like many state early care and education (ECE) systems, recognizes high-quality teacher–child interactions as important for children's well-being. Scores from observations of classroom quality are used for accountability purposes as part of the Head Start Designation Renewal System, which establishes conditions to determine whether programs qualify for continued Head Start funding (Head Start Early Childhood Learning & Knowledge Center 2020). Similarly, many states currently include classroom quality scores in their Quality Rating and Improvement Systems (Build Initiative 2021), which can result in higher subsidy payments for ECE centers.

What is Head Start?

Head Start is a national program that helps young children from families with low-income get ready to succeed in school. It works to promote their early learning, health, nutrition, and their family's well-being. Head Start connects families with medical, dental, and mental health services to be sure that children are receiving the services they need to develop well. Head Start also tries to involve parents in their children's learning and development, and to help parents make progress on their own goals, such as housing stability, continuing education, and financial security (Administration for Children and Families 2020). Head Start operates by providing grants to local public and private nonprofit and for-profit agencies. The agencies in turn deliver comprehensive child development services to children and families who are economically disadvantaged.

What is Head Start FACES?

Starting in 1997, the Head Start Family and Child Experiences Survey (FACES) provides information at the national level about Head Start programs, centers, and classrooms and about the children and families Head Start serves. This report uses data from FACES 2014. More information on the study methodology and measurement in FACES 2014 is available in the FACES 2014–2015 Data Tables and Study Design report (Aikens et al. 2017).

The Pre-K Classroom Assessment Scoring System (Pre-K CLASS; Pianta et al. 2008) is the tool used as part of the Head Start Designation Renewal System to measure quality. It is also the second most used observation tool in Quality Rating and Improvement Systems, with half of states using it for rating or quality improvement purposes (Build Initiative 2021). The Pre-K CLASS assesses the level of responsiveness and sensitivity of teachers, the extent to which teachers provide and scaffold in-depth learning, and the overall organization of the classroom that teachers provide. The Pre-K CLASS includes three domain scores that are made up of scores on 10 dimensions (Table 1).

Table 1. Description of Pre-K CLASS domains

Domain	Aspect of process quality assessed by domain and dimensions included
Emotional Support	Social and emotional functioning in the classroom. Dimensions include positive climate, negative climate, teacher sensitivity, and regard for student perspectives.
Instructional Support	Quality of instructional practices used in the classroom. Dimensions include concept development, quality of feedback, and language modeling.
Classroom Organization	Teacher's ability to organize the classroom to make efficient use of class time. Dimensions include behavior management, productivity, and instructional learning formats.

The Bill & Melinda Gates Foundation funded Mathematica to explore associations between the Pre-K CLASS and Head Start children's school readiness outcomes using data from the Head Start Family and Child Experiences Survey (FACES 2014). We examine three research questions:

1. What are the associations between the Pre-K CLASS Total score, three domains, and 10 dimensions and gains in children's school readiness outcomes?

2. Do the associations of the Pre-K CLASS and children's school readiness vary for dual language learners, children of color, or children experiencing poverty?

3. Are there thresholds at which the associations between the Pre-K CLASS and children's school readiness are stronger (or weaker)?

We first discuss prior research to motivate each of these research questions. We then discuss the methods and results. We conclude with a substantial discussion of our findings in the context of research on classroom quality and provide some recommendations for measuring classroom quality moving forward.

A. Associations between the Pre-K CLASS and children's school readiness outcomes

In ECE settings, the quality of a classroom has been associated with how well children learn and develop (Phillips et al. 2017). A wealth of empirical studies (Hamre et al. 2013; Mashburn et al. 2008), meta-analyses (Perlman et al. 2016), and literature reviews (Aikens et al. 2021; Burchinal 2018) have examined whether there are linear associations between the Pre-K CLASS and children's school readiness outcomes. This research indicates that Pre-K CLASS scores are associated with children's school readiness outcomes in some studies, but the magnitude of these associations is small to modest (Burchinal 2018; Perlman et al. 2016). In other studies, the Pre-K CLASS is not associated with children's outcomes (Aikens et al. 2021).

FACES includes a large sample of classrooms with classroom quality observations and extensive direct child assessments of school readiness outcomes that allows for examination of the relationship between classroom quality and children's outcomes. Prior analyses of FACES data have found limited associations between the Pre-K CLASS and children's school readiness outcomes. For example, Gordon and Peng (2020) found few significant associations between the Pre-K CLASS and children's outcomes using FACES 2009 and 2014 data. Specifically, for FACES 2009, 4 of 32 associations between the Pre-K CLASS Total and domain scores and children's cognitive and social emotional outcomes were significant and for FACES 2014, 1 of these 32 associations was significant.

Some research (for example, Burchinal et al. 2016; Hamre et al. 2014; Soliday Hong et al. 2019) has shown unique associations between Pre-K CLASS domain scores and children's school readiness outcomes, but limited work has examined the association of the Pre-K CLASS dimension scores with children's outcomes (Perlman et al. 2016). It may be that Pre-K CLASS dimension scores that are conceptually linked to specific child outcomes are most predictive of that outcome. For example, scores on the language modeling dimension—which focus on the teacher's use of language facilitation techniques—may be more predictive of language outcomes than the broader Instructional Support domain scores that focus on the quality of instructional practices used in the classroom. Indeed, a generalizability study of the Pre-K CLASS as used in FACES found that dimensions explained some of the variance in a classroom's quality, suggesting the dimensions were rated in different ways across classrooms (Malone et al. 2021). Therefore, we replicate and extend prior findings by examining whether the Pre-K CLASS Total score, domain scores, and dimension scores are associated with children's readiness for school. By

identifying certain dimensions that are most predictive of children's school readiness outcomes, programs could potentially focus on those and provide more targeted supports to improve classroom quality.

B.   Variation in associations of the Pre-K CLASS and children's school readiness

Young children arrive to classrooms with different experiences and needs. Prior work suggests that the quality of the classroom context may matter more for some children than others (for example, Yoshikawa et al. 2013). A literature review found eight recent studies that examined differences in associations between Pre-K CLASS scores and children's outcomes by children's characteristics (Aikens et al. 2021). These studies found that the associations between Pre-K CLASS scores and children's outcomes were stronger for some children, including native English-speaking children, children without Individualized Education Programs (IEPs), non-immigrant children, and children in families with higher incomes. Other studies found associations between Pre-K CLASS scores and outcomes for specific groups of children that were not found with their peers—for example, for boys, children in rural and small urban communities, and children with stronger self-regulation skills. In sum, this set of work suggests associations may vary for children with different characteristics.

Head Start includes a diverse group of children, and yet research using data from FACES 2014 has not explored whether associations between classroom quality and children's school readiness outcomes vary for children with different characteristics. Therefore, we examine whether the associations of the Pre-K CLASS and children's outcomes vary for dual language learners,[1] children of color,[2] and children experiencing poverty.[3] Prior research suggests these subgroups may benefit more from high-quality ECE (Yoshikawa et al. 2013). If we can better understand the extent of the associations of classroom quality with outcomes for children with different backgrounds, this could help shape decisions with financial and programmatic consequences. For example, it could inform how ECE systems target resources to improve the quality of teacher–child interactions for children who stand to benefit from it the most.

C.   Thresholds at which the associations between the Pre-K CLASS and children's school readiness are stronger (or weaker)

Because of the small to modest linear associations between classroom quality scores and children's school readiness outcomes, researchers have suggested there might be thresholds above which quality is more strongly related to outcomes (Zaslow et al. 2010). However, research on thresholds of quality using the Pre-K CLASS is mixed. In a meta-analysis of four different data sets, Burchinal and colleagues (2016) found that Instructional Support was more strongly associated with children's language and literacy skills in classrooms of higher quality than lower quality. However, they did not find evidence of thresholds on Emotional Support for children's social skills or behavioral problems. In contrast, Hatfield and colleagues (2016) did not find evidence of thresholds for Instructional Support, but did find stronger associations

---

[1] We considered children to be dual language learners if the language that is usually or always spoken to the child in the home was any language other than English.

[2] Children of color include Black/African American, non-Hispanic/Latino; Hispanic/Latino; and multiple races or others, non-Hispanic/Latino.

[3] We define households with low-income as households with income at or below 100 percent of the federal poverty line. FACES uses household income, and not family income (which is used to determine Head Start eligibility), to capture the entirety of the resources available to children at home. Household income captures the income of all members of the same household, but family income captures only the reported income of the child's parents or legal guardians.

between Emotional Support and children's inhibitory control and phonological skills and between Classroom Organization and children's phonological skills and print knowledge in higher-quality classrooms. Recent research found some evidence of thresholds for Classroom Organization and children's teacher-reported behavioral outcomes, but no evidence of thresholds for Instructional Support or Emotional Support (Harding et al. forthcoming).

In earlier rounds of Head Start FACES specifically, Burchinal and colleagues (2016) examined thresholds for Instructional Support and Emotional Support and found mixed support for thresholds. There were no thresholds on Instructional Support for children's language, literacy, and math outcomes in FACES 2006. For FACES 2009, Instructional Support was more strongly associated with language skills in higher-quality classrooms, but there was no evidence of thresholds for literacy or math. There was also no evidence of thresholds on Emotional Support for children's problem behaviors or social skills in FACES 2009, and this was not examined for FACES 2006.

Understanding thresholds in the association between classroom quality and children's school readiness 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, 2016). Therefore, we examine whether there are thresholds at which quality is more strongly associated with children's outcomes in FACES 2014. By identifying thresholds in quality that classrooms need to meet or exceed to influence school readiness outcomes, resources could be targeted toward raising classrooms above those thresholds.

II. Methods

A. Data

We used data from Head Start FACES 2014, which provides nationally representative data about Head Start programs, centers, and classrooms and the children and families they serve. Stratified sampling was used to select 60 programs based on the 2012–2013 Head Start Program Information Report database. Within each participating program, 2 centers were randomly selected to participate on average (N = 119); within each participating center, 2 classrooms were randomly selected to participate on average (N = 247); and 12 children in each classroom were selected at random, for a total of 2,462 children with consent. Most children who were enrolled in the fall were still enrolled in their Head Start program in the spring (N = 2,206). Multiple types of data were collected, including staff and parent surveys, classroom observations, and direct assessments of children. More information on the study methodology and measurement in FACES 2019 is available in the user manual (Kopack Klein et al. 2017) and FACES Spring 2019 Data Tables and Study Design report for Head Start children and families (Doran et al. 2022).

For the current study, we included children with (1) parent survey data in fall 2014 or spring 2015, (2) teacher–child report or child direct assessment data in fall 2014 and spring 2015, (3) teacher survey data in spring 2015, and (4) classroom observation data in spring 2015 (N = 1,736). This sample was weighted to represent children who were enrolled in the fall and are still participating in the spring. To retain most of the full sample, we used missing data dummy codes for some categorial demographic variables.[4] However, the sample included in this analysis is smaller because we included children with full information on the following key data elements: classroom observation scores (N = 1,727) , race and ethnicity (N = 1,727), age (N = 1,727), sex (N = 1,727), dual language learner status (N = 1,726), classroom size (N = 1,726), part-day classroom status (N = 1,726), teacher experience (N = 1,668), teacher education (N = 1,563), and teacher race and ethnicity (N = 1,517). Finally, the sample included in the regression analyses varied based on whether children had outcome data for the individual assessments in the fall and spring (N = 982 to 1,517).[5]

---

[4] We used dummy variables for categorical variables with missing data (IEP status and poverty) to retain a larger sample. By including a separate dummy variable, children missing that information will have a different mean outcome value than non-missing children, without influencing the relationship between the other categories and the outcome. If the incidence of children with missing information is not missing at random, the coefficient on the missing value category can account for that and provide a more accurate picture.

[5] Only children who were 4 years old at the time of the assessment completed the pencil-tapping task to measure executive function, which is why the analytic sample for that assessment is much smaller.

Figure 1. Analysis sample
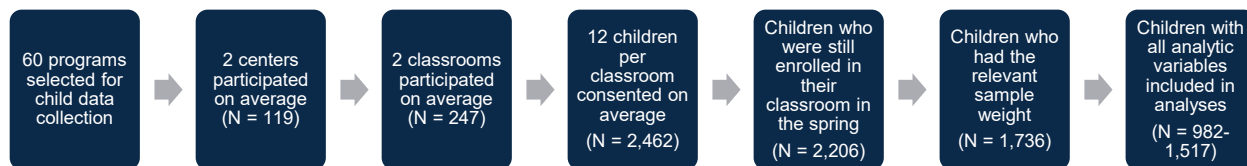
Sample characteristics are shown in Table 2, including for children in classrooms above and below the median on the Pre-K CLASS Total score.[6] About one-third of children were Hispanic/Latino and White, non-Hispanic/Latino, and slightly fewer than one-third of children were Black/African American, non-Hispanic/Latino. Nearly 20 percent of children lived in households in which a language other than English was primarily spoken. More than 60 percent of children lived in households with income at or below the federal poverty line. Descriptively, child and family characteristics were similar in classrooms above and below the median on quality. However, one exception is that more Black/African American, non-Hispanic/Latino children were in classrooms of lower quality than in classrooms with higher quality. In terms of children's classrooms and teachers, about three-quarters of children had teachers with a bachelor's degree. Descriptively, most classroom and teacher characteristics were similar. However, two exceptions are that more classrooms with White/Caucasian, non-Hispanic/Latino teachers and part-day classrooms were rated as above-median quality than below-median quality.[7]

---

[6] The sample included in analyses differed from the full sample of children with the relevant sample weight who were weighted to represent children who were still enrolled in their classroom in the spring. The included sample was more likely to be White, non-Hispanic/Latino or Black/African American, non-Hispanic/Latino. They were also more likely to be female, speak English at home, be in a full-day classroom, and have teachers who are Black/African American, non-Hispanic/Latino. The analytic sample differs because only children tested in English in the spring were included so as to have consistent outcome variables.

[7] We did not examine whether there were statistically significant differences between characteristics of classrooms below and above median quality. Instead, we describe differences larger than 10 percentage points.

Table 2. Sample characteristics

	All classrooms		Below-median quality classrooms		Above-median quality classrooms	
	Mean/ Percentage	SD	Mean/ Percentage	SD	Mean/ Percentage	SD
Child and family characteristics						
Age in months at spring assessment	55.96	0.35	56.02	0.54	55.91	0.46
Pre-test conducted in Spanish	5%		5%		5%	
Male child	49%		48%		49%	
Household poverty						
100% or below	61%		62%		60%	
Missing	9%		10%		9%	
Race and ethnicity						
Black/African American, non-Hispanic/Latino	27%		35%		19%	
Hispanic/Latino	35%		33%		37%	
White/Caucasian, non-Hispanic/Latino	30%		24%		35%	
Multiple races or others, non-Hispanic/Latino	8%		9%		8%	
Home language						
English only	82%		82%		83%	
Spanish only or multilingual	18%		18%		17%	
Individualized Education Program (IEP)						
Has IEP	6%		5%		7%	
Missing IEP	4%		3%		5%	
Teacher and classroom characteristics						
Teacher race and ethnicity						
Black/African American, non-Hispanic/Latino	30%		41%		19%	
Latino/Hispanic, non-Hispanic/Latino	23%		23%		23%	
White/Caucasian, non-Hispanic/Latino	44%		34%		53%	
Multiple races or other	3%		2%		4%	
Lead teacher has a bachelor's degree or higher	73%		73%		72%	
Teacher years of teaching experience	14.68	0.77	13.80	0.92	15.46	1.24
Part-day classroom	48%		39%		55%	
Class size	17.46	0.24	17.31	0.27	17.58	0.32

Note: N = 1,454 (all participants included in analyses of the Peabody Picture Vocabulary Test–Fourth Edition). Statistics are weighted to represent all children who were enrolled in Head Start in fall and were still enrolled in spring. Descriptive statistics were computed at the child level. In this table, the cut point for below- and above-median classroom quality is based on the Pre-K CLASS Total score of 4.20. Standard deviations are presented for continuous variables.

B. Measures

1. Classroom quality

The Pre-K CLASS is a widely used classroom quality observation tool that examines teacher–child interactions. It consists of three domains, and each domain is composed of multiple dimensions (see Table 1). The Pre-K CLASS is measured on a seven-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.

Trained classroom observers conducted observations in spring 2015. Observations generally took place in the morning for three to four hours. Three or four 20-minute cycles were coded for the Pre-K CLASS.[8] Correlations and descriptive statistics for the domains are in Table 3. All three domain scores were highly correlated with the Total score. Classroom Organization and Emotional Support were highly correlated, whereas they were only moderately correlated with Instructional Support. Dimension scores were typically highly correlated within domain (Table A.1). The median Instructional Support score was in the low range on the developers' thresholds, whereas Classroom Organization and Emotional Support were in the mid-range.

Table 3. Correlations and descriptive statistics for Pre-K CLASS scores

	Total	Emotional Support	Instructional Support	Classroom Organization
Total	1.00			
Emotional Support	0.82	1.00		
Instructional Support	0.79	0.38	1.00	
Classroom Organization	0.86	0.80	0.43	1.00
Median	4.20	5.40	2.30	4.80
SD	0.58	0.51	0.90	0.72
Range (Min-Max)	2.24-6.48	2.83-7.00	1.00-6.33	2.58-6.78

Note: N = 1,454 (all participants included in analyses of the Peabody Picture Vocabulary Test–Fourth Edition). Statistics are weighted to represent all children who were enrolled in Head Start in fall and were still enrolled in spring. Descriptive statistics were computed at the child level. All correlations statistically significant at $p \leq .01$.

2. Children's school readiness outcomes

Children's outcomes were assessed in the fall and spring of each year in a one-on-one assessment with trained assessors and via teacher reports on individual children. Descriptions of all outcomes are in Table 4. Assessors asked children questions and showed them corresponding pictures on a second computer screen. All children began the direct assessment with two English language screening measures; children were then routed to an English, Spanish, or "non-English" assessment pathway. We included children who were assessed in English or Spanish in the fall in analyses for outcomes for which they completed

---

[8] About 85 percent of the classrooms have four cycles of Pre-K CLASS data.

the same assessment.[9] We only included children who were assessed in English in the spring so that all children had the relevant outcome variables. Teachers also reported on children's skills and development in fall and spring of the Head Start year using a web or paper questionnaire. All outcomes were standardized so that coefficients can be interpreted as effect sizes.

Table 4. Description of children's outcome assessments

Skill and assessment	Description
Cognitive skills	
Expressive vocabulary: Expressive One-Word Picture Vocabulary Test–4 (EOWPVT–4; Martin and Brownell 2010)	The EOWPVT–4 measures children's expressive vocabulary by assessing how well children can name the objects, actions, or concepts presented in full-color pictures in English.
Receptive vocabulary: Peabody Picture Vocabulary Test–Fourth Edition (PPVT–4 (Dunn et al. 2006)	The PPVT–4 measures children's English receptive vocabulary knowledge by asking children to say, or indicate by pointing, which of four pictures best shows the meaning of a word the assessor says aloud.
Early writing: Woodcock Johnson III Tests of Achievement Spelling subtest (Woodcock et al. 2001).	The Spelling subtest measures children's early writing and spelling ability. The first six items measure fine motor coordination and prewriting skills, such as drawing lines and copying letters. The remaining items measure the child's skill in providing written responses when asked to write specific upper- or lowercase letters. Later parts of the subtest ask the child to write specific words and phrases and punctuation marks.
Letter-word knowledge: Woodcock Johnson III Tests of Achievement Letter-Word Identification subtest (Woodcock et al. 2001).	The Letter-Word Identification subtest measures children's alphabet knowledge, print concepts/conventions, and sight word recognition by assessing children's skills in identifying isolated letters and words that appear in different letter fonts on the pages of the test book.
Early math: Woodcock Johnson III Tests of Achievement Applied Problems Subtest (Woodcock et al. 2001).	The Applied Problems subtest captures math skills in the areas of number concepts and quantities, number relationships and operations, counting, and reasoning/problem solving. Children analyze and solve practical problems involving numbers.
Social emotional skills	
Executive function: Pencil tapping (Blair 2002; Diamond and Taylor 1996; Smith-Donald et al. 2007) is an adaptation of a peg-tapping task (Blair 2002; Diamond and Taylor 1996). The task was translated into Spanish.	The pencil-tapping task requires the child to remember a rule (working memory) to do the opposite of what the assessor does, that is, tap one time when the assessor taps two times and tap two times when the assessor taps one time (inhibitory control). Only children who were 4 years old at the time of the assessment completed the pencil-tapping task.
Approaches to learning (U.S. Department of Education 2002)	Teachers rated each child in on a scale of 1 ("never") to 4 ("very often") on six items that assess a child's motivation, attention, organization, persistence, and independence in learning.
Behavioral problems: Items come from an abbreviated adaptation of the Personal Maturity Scale (Entwisle et al. 1987) and from the Behavior Problems Index (Peterson and Zill 1986).	Teachers responded to questions about the frequency of 14 negative child behaviors associated with learning problems and later grade retention, using a scale of 1 ("never") to 3 ("very often").

[9] Children assessed in English or Spanish in the fall were included for analyses of the Peabody Picture Vocabulary Test–Fourth Edition and pencil-tapping assessments and for teacher reports of children's approaches to learning, behavioral problems, and social skills (with fall language of assessment included as a control). Only children assessed in English in fall and spring (95 percent of children in our analytic sample) were included in analyses of the Woodcock Johnson III Tests of Achievement and Expressive One-Word Picture Vocabulary Test–4.

Skill and assessment	Description
Social skills: Items come from the Personal Maturity Scale (Entwisle et al. 1987) and the Social Skills Rating System (Gresham and Elliott 1990; Elliott et al. 1988).	Teachers responded to questions about the frequency of 12 cooperative classroom behaviors, using a scale of 1 ("never") to 3 ("very often").

Note: We used W/Growth Score Value (GSV) and raw scores in the analysis because these have more variability. These scores indicate absolute rather than relative performance. W/GSV scores permit measurement of change or gains in performance on the same scale over time.

We show means and standard deviations for all outcomes in Table 5, including by the subgroups examined in research question 2. We did not examine whether differences were statistically significant, but mean classroom quality was descriptively similar for all subgroups (dual language learners, children of color, children experiencing poverty), whereas children's outcomes seemed to vary across groups.

3.   Covariates

To address that children with different characteristics may be more likely to be in high- or low-quality classrooms, we accounted for several baseline child, family, teacher, and classroom characteristics. We also controlled for the fall score on the respective outcome to see whether classroom quality is associated with changes or gains in children's school readiness outcomes. Child and family covariates included the child's age in months at the time of the spring assessment, whether the child was assessed in Spanish in the fall, race and ethnicity (Black, non-Hispanic/Latino; Hispanic/Latino; other, non-Hispanic/Latino; White, non-Hispanic/Latino; with White, non-Hispanic/Latino as the reference group), sex, whether the child only speaks English at home, whether the child has an IEP, and whether the child was living in a household with income at or below the poverty line. Teacher characteristics included race and ethnicity (Black, non-Hispanic/Latino; Hispanic/Latino; other, non-Hispanic/Latino; White, non-Hispanic/Latino; with White, non-Hispanic/Latino as the reference group), whether the teacher has a bachelor's degree or higher, and years of teaching experience. Classroom characteristics included whether the classroom was part-day and the number of children in the classroom.

C.   Analysis

1.   What are the associations between the Pre-K Class Total score, three domains, and 10 dimensions and gains in children's school readiness outcomes?

First, we examined the relationship between the Pre-K CLASS and children's school readiness outcomes by estimating four-level hierarchical linear models of children (Level 1) nested within classrooms (Level 2) nested within centers (Level 3) nested within programs (Level 4). Our key classroom quality predictors are the Pre-K CLASS Total and domain scores. We also examined the association between the Pre-K CLASS dimensions and children's outcomes. All regressions controlled for all covariates.[10] Continuous variables were standardized so that coefficients can be interpreted as effect sizes.

---

[10] Across all analyses, we adjusted the standard errors estimates to account for the design effect related to unequal weighting of the sample. Standard hierarchical linear modeling software cannot account for the design effect attributable to unequal weighting of the sample. Therefore, we manually adjusted the standard error estimates to account for the design effect related to unequal weighting of the sample (Kopack Klein et al. 2017). We did so by multiplying the standard error by the square root of the design effect (1.19). We then calculated an adjusted *p*-value that we reported for all applicable analyses.

Table 5. Descriptive statistics of classroom quality scores and outcomes by subgroups

	Dual language learners[a]		Non-dual language learners		Children of color[b]		Non-children of color		Children experiencing poverty[c]		Children not experiencing poverty		Total	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Classroom quality														
Total	4.31	0.58	4.19	0.58	4.21	0.60	4.21	0.55	4.16	0.60	4.31	0.53	4.21	0.58
Emotional Support	5.45	0.52	5.41	0.51	5.41	0.52	5.43	0.50	5.36	0.53	5.54	0.44	5.42	0.51
Instructional Support	2.60	0.94	2.44	0.89	2.49	0.93	2.43	0.85	2.44	0.90	2.52	0.90	2.47	0.90
Classroom Organization	4.88	0.68	4.72	0.72	4.73	0.73	4.77	0.69	4.69	0.74	4.88	0.65	4.74	0.72
Spring outcomes														
Expressive vocabulary	38.50	17.63	53.17	18.69	49.03	17.74	58.13	19.37	51.19	19.23	53.66	18.24	52.22	18.86
Receptive vocabulary	104.64	15.59	118.36	16.52	114.09	17.62	118.73	16.08	112.95	16.94	122.88	15.67	115.91	17.18
Early writing	365.45	34.32	363.04	35.33	363.88	35.23	362.35	35.17	363.26	34.68	363.57	35.96	363.33	35.22
Letter-word knowledge	329.32	27.37	329.93	30.07	330.38	29.56	328.83	29.92	329.33	30.05	330.74	29.03	329.84	29.66
Early math	389.66	26.56	393.62	27.42	390.63	27.72	398.22	25.70	392.23	27.67	394.14	26.84	393.07	27.37
Executive function	9.34	5.81	9.95	5.37	9.71	5.42	10.05	5.52	9.64	5.45	10.30	5.45	9.84	5.46
Approaches to learning	2.15	0.64	1.91	0.76	1.92	0.74	1.99	0.75	1.94	0.74	1.97	0.76	1.95	0.75
Behavioral problems	2.55	3.75	4.22	4.66	4.00	4.60	3.82	4.48	3.75	4.51	4.36	4.65	3.93	4.56
Social skills	19.11	4.36	17.39	4.91	17.64	4.81	17.75	4.96	17.64	4.77	17.78	5.08	17.68	4.87
Percentage of sample	19%		81%		61%		39%		70%		30%		100%	

Note: N = 1,454 (all participants included in analyses of the Peabody Picture Vocabulary Test–Fourth Edition, but sample sizes vary for specific outcomes). Statistics are weighted to represent all children who were enrolled in Head Start in fall and were still enrolled in spring. Descriptive statistics were computed at the child level.

[a]We considered children to be dual language learners if the language that is usually or always spoken to the child in the home was any language other than English.

[b]Children of color include Black/African American, non-Hispanic/Latino; Hispanic/Latino; and multiple races or others, non-Hispanic/Latino.

[c]Children experiencing poverty are in households at or below 100 percent of the poverty line.

2. Do the associations of the Pre-K CLASS and children's school readiness vary for dual language learners, children of color, or children experiencing poverty?

Second, we examined whether associations between the Pre-K CLASS Total score and children's school readiness outcomes vary for children who are dual language learners (children whose parents primarily speak to them in Spanish at home versus English), children of color (versus White, non-Hispanic/Latino), and children experiencing poverty (versus those with household poverty above 100 percent of the poverty line). We conducted the same four-level regression models including an interaction term between the Pre-K CLASS Total score and each subgroup. We used group-mean centered variables, which rescales variables to reflect the child's relative standing on the predictor with respect to other children in their own classroom. We also conducted these models using the Pre-K CLASS domain scores to ensure the patterns did not vary.

3. Are there thresholds at which the associations between the Pre-K CLASS and children's school readiness are stronger (or weaker)?

Third, we examined whether classroom quality scores need to reach a particular level, or threshold, to influence children's school readiness outcomes. Specifically, we fit four-level spline models, controlling for all covariates. The splines are linear regression models that are estimated for classrooms that fall in the low and high portion of the quality range. The first slope is for the association of quality and children's outcomes in classrooms falling in the lower-quality range, and the second slope is for classrooms falling in the higher-quality range. The difference between these two slopes are tested to see if there is a threshold effect.

We chose the median of each classroom quality score to be the "knot," or threshold, at which the slope was allowed to change because of the distribution of our data and because these cut points provided the most power for our analyses (see Table 3).[11] These chosen cut points corresponded approximately to the mid-range of quality for all classroom quality measures. One exception is Pre-K CLASS Instructional Support, which had a cut point of 2.3 in the low-quality range, consistent with prior literature documenting the lower scores on this domain (for example, Burchinal 2018).

---

[11] We also examined thresholds that were previously used in the literature, and if necessary, we adjusted to make sure at least 10 percent of the sample was in each group. For the Total score, we used a threshold of 4.5 (Purtell and Ansari 2018). For Emotional Support, we used a threshold of 5.25; for Classroom Organization, we used a threshold of 5; for Instructional Support, we used a threshold of 2.75 (Burchinal et al. 2010; 2016).

III. Results

A. What are the associations between the Pre-K CLASS Total score, three domains, and 10 dimensions and gains in children's school readiness outcomes?

Overall, there were 2 statistically significant associations out of 36 associations between the Pre-K CLASS Total and domain scores and children's nine school readiness outcomes. Controlling for fall scores and other covariates, children's early math scores were positively associated with the Pre-K CLASS Total score and Instructional Support domain score (Table 6). By chance, we would expect about 2 associations to be statistically significant, so we conclude there were limited associations of the Pre-K CLASS with children's outcomes, consistent with Gordon and Peng (2020).

There were 4 statistically significant associations out of 90 associations between the 10 Pre-K CLASS dimension scores and children's nine school readiness outcomes (Table A.2). Consistent with the domain results, all three dimensions of Instructional Support were positively associated with children's early math scores, suggesting there is limited specificity among these dimensions. Negative climate was negatively associated with children's expressive vocabulary scores, suggesting teacher negativity could be detrimental to children's learning. However, again, we would expect about 5 associations to be significant by chance. Therefore, we conclude there were limited associations of the Pre-K CLASS dimensions with children's outcomes.

B. Do the associations of the Pre-K CLASS and children's school readiness vary for dual language learners, children of color, and children experiencing poverty?

Because the associations were similar for the Pre-K CLASS Total score as the domain scores, we focus on examining whether the association of the Total score and children's school readiness outcomes varies by subgroup (Table 7). There was no evidence that Pre-K CLASS scores mattered more for dual language learners or children of color. For children experiencing poverty, there was one statistically significant result such that for this group of children, classroom quality was more strongly associated with increases in receptive vocabulary.[12] Overall, there is limited evidence that Pre-K CLASS scores are more or less strongly associated with children's outcomes for dual language learners, children of color, or children experiencing poverty.

---

[12] We also confirmed that there were no consistent patterns across subgroups using the Pre-K CLASS domain scores. Although 6 of 54 associations were significant, we would expect about 3 significant associations by chance, and the patterns were not consistent across subgroups or outcomes.

Table 6. Results of regressing school readiness outcomes on classroom quality measures

	Total Pre-K CLASS score	Emotional Support	Instructional Support	Classroom Organization	
	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	N
Expressive vocabulary	0.00	0.00	0.01	-0.01	1,213
	(0.02)	(0.02)	(0.02)	(0.02)	
Receptive vocabulary	0.01	0.00	0.01	0.01	1,454
	(0.02)	(0.02)	(0.02)	(0.02)	
Early writing	0.03	0.00	0.05	0.02	1,355
	(0.04)	(0.04)	(0.04)	(0.04)	
Letter-word knowledge	0.00	-0.04	0.03	-0.01	1,343
	(0.04)	(0.04)	(0.04)	(0.04)	
Early math	0.05*	0.02	0.06**	0.03	1,354
	(0.02)	(0.02)	(0.02)	(0.02)	
Executive function	0.01	0.00	0.02	0.01	982
	(0.04)	(0.04)	(0.04)	(0.04)	
Approaches to learning	0.03	0.01	0.03	0.02	1,516
	(0.05)	(0.05)	(0.05)	(0.04)	
Behavioral problems	0.05	0.02	0.05	0.04	1,517
	(0.04)	(0.04)	(0.04)	(0.04)	
Social skills	0.00	-0.02	0.02	-0.01	1,514
	(0.05)	(0.05)	(0.05)	(0.05)	

Note: Statistics are weighted to represent all children who were enrolled in Head Start in fall and were still enrolled in spring. Outcomes were standardized so coefficients can be interpreted as effect sizes. All models control for child, teacher, and classroom characteristics. Child characteristics include the fall score on the respective outcome, whether the child was assessed in English or Spanish in the fall, child race and ethnicity, child sex, child age in months at time of spring assessment, IEP status, and poverty status. Teacher characteristics include race and ethnicity, education, and years of teaching experience. Classroom characteristics include size and part-day status. We also included dummy variables for missing IEP status and poverty status to decrease missing data.

+ $p \leq 0.10$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$. SE = standard error.

Table 7. Results of regressing school readiness outcomes on the Pre-K CLASS total score for dual language learners, children of color, and children experiencing poverty

	Dual language learners[a]	Children of color[b]	Children experiencing poverty[c]
	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Expressive vocabulary	-0.01	0.04	0.03
	(0.10)	(0.08)	(0.05)
Receptive vocabulary	-0.05	0.01	0.08*
	(0.05)	(0.06)	(0.04)
Early writing	-0.13	0.00	0.03
	(0.08)	(0.09)	(0.06)
Letter-word knowledge	0.00	0.13	-0.02
	(0.08)	(0.08)	(0.05)
Early math	-0.15+	-0.08	0.02
	(0.08)	(0.09)	(0.06)
Executive function	-0.06	0.15	0.06
	(0.08)	(0.12)	(0.07)
Approaches to learning	0.05	0.02	-0.03
	(0.07)	(0.08)	(0.06)
Behavioral problems	-0.03	-0.04	0.07
	(0.07)	(0.08)	(0.05)
Social skills	0.04	-0.07	-0.04
	(0.07)	(0.08)	(0.06)

Note: Statistics are weighted to represent all children who were enrolled in Head Start in fall and were still enrolled in spring. Coefficients and standard errors are for the interaction between the subgroup and Pre-K CLASS Total score. Outcomes were standardized so coefficients can be interpreted as effect sizes.

[a] We considered children to be dual language learners if the language that is usually or always spoken to the child in the home was any language other than English.

[b] Children of color include Black/African American, non-Hispanic/Latino; Hispanic/Latino; and multiple races or others, non-Hispanic/Latino.

[c] Children experiencing poverty are in households at or below 100 percent of the poverty line.

+ $p \leq 0.10$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$. SE = standard error.

C. Are there thresholds at which the associations between the Pre-K CLASS and children's school readiness are stronger (or weaker)?

Again, we focus on Pre-K CLASS Total scores to examine whether there are thresholds in the association between quality and children's school readiness. In Table 8, we present the results from estimating the spline regression models, with the spline knot set equal to the median cut point of the Total score (see Table 3). The "Below median quality" column corresponds to the slopes and standard errors for classrooms below the median quality level on the Total score. The "Above median quality" column corresponds to the slopes and standard errors for classrooms above the median quality level on the Total

score. The "Differ?" column indicates whether the difference between the slopes for the below- and above-median quality classrooms was statistically significantly different.

Overall, there were limited threshold effects.[13,14] Classroom quality was significantly associated with children's early math in classrooms above the median on quality, but this was only different from the association in classrooms below the median on quality at trend-level.

Table 8. Threshold regression results, using median-split cut points

	Pre-K CLASS Total score		
	Below median quality (SE)	Above median quality (SE)	Differ?
Expressive vocabulary	0.00	-0.03	No
	(0.01)	(0.08)	
Receptive vocabulary	-0.02*	-0.04	No
	(0.01)	(0.07)	
Early writing	0.00	0.01	No
	(0.02)	(0.12)	
Letter-word knowledge	-0.02	-0.06	No
	(0.02)	(0.11)	
Early math	0.01	0.16*	Yes+
	(0.02)	(0.08)	
Executive function	0.02	0.08	No
	(0.02)	(0.12)	
Approaches to learning	-0.06**	-0.15	No
	(0.02)	(0.12)	
Behavioral problems	-0.01	0.09	No
	(0.02)	(0.09)	
Social skills	-0.04*	-0.15	No
	(0.02)	(0.13)	

Note: Statistics are weighted to represent all children who were enrolled in Head Start in fall and were still enrolled in spring. Outcomes were standardized so coefficients can be interpreted as effect sizes. 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, teacher, and classroom characteristics. Child characteristics include the fall score on the respective outcome, whether the child was assessed in English or Spanish in the fall, child race and ethnicity, child sex, child age in months at time of spring assessment, months between fall and spring assessments, IEP status, and poverty status. Teacher characteristics include race and ethnicity, education, and years of teaching experience. Classroom characteristics include size and part-day status. We also included dummy variables for missing IEP status and poverty status to decrease missing data.

+ $p \leq 0.10$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

SE = standard error.

___

[13] We also examined thresholds for the domain scores, and only 1 of the 27 associations was significant, also indicating limited support for thresholds of quality (see Table A.3).

[14] Results were similar using the thresholds with the pre-determined cutoffs.

IV. Discussion

Overall, we found limited associations between children's school readiness outcomes and the Pre-K CLASS Total score, domain scores, and dimension scores. These associations were limited regardless of whether children were dual language learners, children of color, or experiencing poverty. We also did not find evidence to support that classroom quality needs to reach a certain threshold to be associated with children's outcomes. For interpretation of these findings, it is important to note that the sample in our analyses is not representative of all Head Start children because we did not include children for whom we did not have outcome data. Therefore, we cannot generalize these findings to all Head Start children.

Next, we discuss potential explanations for these limited findings, including potential challenges with measuring classroom quality, the content of the Pre-K CLASS, and other analytic considerations.

A. Potential challenges with measuring classroom quality using the Pre-K CLASS

Variation in Pre-K CLASS scores may be due to the conditions under which classrooms are observed, rather than only because of differences in classroom quality. CLASS scores may vary because of the time of day or day of the week when classrooms are observed, the observers who conduct the observation, and the activities going on in the classroom (Mashburn et al. 2017). Past research using the CLASS in kindergarten classrooms found that observers accounted for about 5 to 16 percent of the variance in domain scores (Mantzicopoulos et al. 2018). Using Pre-K CLASS data from FACES 2014, one study estimated the amount of variance in domain scores that is attributable to classrooms, observers, and cycles (Malone et al. 2021). This particular study found that there was substantial variation in Classroom Organization and Instructional Support domains across Pre-K CLASS cycles, suggesting that the quality of these domains differs across classroom activities. Indeed, evidence from other studies indicates that structure and content make a difference in CLASS scores. For example, pre-K teachers had higher Instructional Support scores in large-group and free-choice settings (Cabell et al. 2013).

Pre-K CLASS scores often have a restricted range. In FACES 2014, less than 1 percent of classrooms were rated in the developer's high range on Instructional Support, whereas less than 2 percent of classrooms were rated in the developer's low range on Classroom Organization and Emotional Support (Aikens et al. 2016). Other researchers similarly note that several scores on the Pre-K CLASS were not used by any observers and suggest this is a potential issue with the Pre-K CLASS (Burchinal 2018; Styck et al. 2021). Limited variation makes it difficult to differentiate between most classrooms, and therefore reduces the utility of classroom quality scores. It is possible that classrooms would need to span the full range of values for Pre-K CLASS scores to show associations with outcomes. For example, Pre-K CLASS scores for Instructional Support might need to be in the higher ranges (6 to 7) to detect significant associations with children's outcomes. However, finding samples for which Pre-K CLASS Instructional Support scores are high is rare. Boston Public Schools' widely recognized pre-K program, in which teachers have master's degrees and use high-quality content-specific curricula in literacy and math, had average Pre-K CLASS Instructional Support scores of 4.30 (Weiland et al. 2013). Researchers have suggested that one way to increase the range of the scores is by either rewriting or adding items for the different domains (Gordon and Peng 2020). For example, Instructional Support could include more items that capture instructional practices on the easier end, whereas Emotional Support and Classroom Organization could include more items on the harder end to allow more room for growth.

B. Limitations of the Pre-K CLASS in measuring aspects of quality related to children's outcomes

Current classroom quality measures may not be capturing the curriculum and instructional content. There might be additional classroom quality elements that are currently not reflected in commonly used measures of classroom quality that are also important for children's outcomes. For example, the Pre-K CLASS does not assess the content of what is taught in the classroom. The curriculum used and the content children are exposed to in the classroom may be more influential in supporting gains in children's math, language, and executive function skills than the more general instructional practices that the Pre-K CLASS measures (Mashburn et al. 2016; McGuire et al. 2016; Sarama et al. 2016; Whittaker et al. 2020). Relatedly, classroom quality measures may not be detecting enough teacher–child interactions and instructional content in the classroom. Children spend only a small part of the day on academic activities, and it is not clear how much interaction or content observers can capture within the standard two-hour observation window of the Pre-K CLASS. Past research indicates that preschool children tend to spend only 3 to 13 percent of the day in small-group activities, which is when teachers and children typically have more opportunities to engage in back-and-forth exchanges (Buysse et al. 2016; Cabell et al. 2013; Farran et al. 2017; Fuligni et al. 2012).

Individual child experiences in the classroom may matter more for children's outcomes. Quality measured at the classroom level, capturing what the average child experiences, also may not capture what is most important for children's learning. Recently, researchers have shown that the quality of what an individual child experiences in the classroom is more predictive of school readiness outcomes (Burchinal et al. 2021; Hanno et al. 2021; Pianta et al. 2020). Child-level observations may provide more insight into children's learning and development that would be missed if aggregated to the classroom level. Even if teachers are providing high-quality learning experiences overall, not all children may be experiencing it in the same way (Chien et al. 2010; Rojas and Abenavoli 2021). Whereas classroom quality is typically measured by ratings of the whole classroom, behavior counts capture individual differences in classroom experiences. There are currently several measures that focus on the quality of the individual learning experience with behavior counts and ratings (Bilbrey et al. 2007; Downer et al. 2010) but these are not often used at scale because of the intensity of time and resources. Research using these measures has found that behavior counts more strongly relate to gains in children's outcomes than classroom level ratings of quality. For example, behavior counts of children individually observed spending more time in literacy, sound-related, and whole-group activities were related to gains in language and academic skills (Burchinal et al. 2021).

C. Other challenges for finding associations between classroom quality and children's outcomes

The benefits of classroom quality may be cumulative. It may be that classroom quality is more important for children's learning and development when aggregated across multiple years (Pianta et al. 2016). That is, high-quality interactions that children experience in a single year may be less impactful than high-quality interactions that are sustained over time, such as during pre-K through the early grades. Previous research has shown the benefits of cumulative classroom quality. In one study, the cumulative benefits of high-quality Emotional Support and Classroom Organization in both pre-K and kindergarten were associated with gains in children's social-emotional skills at the end of kindergarten (Broekhuizen et al. 2016). Similarly, higher-quality Emotional Support and Instructional Support in both pre-K and kindergarten was associated with higher levels of language, literacy, and math skills (Carr et al. 2019;

Cash et al. 2019). Taken together, this research suggests that an accumulation of high-quality classroom experiences over multiple years may be more important than a high-quality experience at a single point in time. This may explain the limited associations of the Pre-K CLASS with children's outcomes we found in the current study. In particular, for this sample, the fall scores were measured an average of six months before the spring outcomes, meaning there was only a short period during which classroom quality could influence children's outcomes. It may be that gains from classroom quality take longer to be realized.

Selection bias may be an issue in examining associations between classroom quality and children's outcomes. Children are not randomly sorted into classrooms, and their opportunities for and engagement in high-quality classroom experiences are associated with socioeconomic status, race and ethnicity, or parents' choices (Chaudry et al. 2017; Currie and Thomas 2000; Lee and Loeb 1995; Rothstein 2009). For example, more advantaged families often attend higher-quality settings and less advantaged families attend lower-quality settings (Hillemeier et al. 2013; Perlman et al. 2016). Other studies' findings of associations between classroom quality and children's outcomes might have been because of selection bias. This might be an issue in the current study; however, our sample included children who all qualify for and choose to attend a Head Start center, and are likely similar to one another as compared with other samples that may be spread across different program auspices. The Head Start FACES data that we used for these analyses implicitly control for some of the differences related to selection described above, such as socioeconomic status, race and ethnicity, or parents' choices by only including children who choose to attend a Head Start center.

There could be limitations to measures of children's outcomes. It may be that what is currently taught to children in the classroom may not be captured by the outcomes measured (Martone and Sireci 2009; Porter and Smithson 2001). For example, it is possible that children in high quality classrooms are indeed making school readiness gains in the classroom, but it may not be reflected in their scores on the assessments because what they are learning is not being measured on these assessments. In the upper grades, some researchers have found that when teachers align their instruction to criterion-based standards assessments, the associations with the CLASS domains tend to be larger (Allen et al. 2013). It may also be that the current measures of children's outcomes are too narrow. These assessments might not provide a complete picture of children's achievement because of the small sample of knowledge that is tested. Relatedly, many measures of children's outcomes have generally focused on rote skills, and there has not been as much attention to assessing children's higher-order skills that require mental manipulation of information (Burchinal 2018; Snow and Van Hemel 2008).

Challenges with sample size. Finally, it may be that we did not have a large enough sample size, and therefore limited power, for our subgroup and threshold analyses to detect associations between the Pre-K CLASS and child outcomes. Head Start FACES is one of the largest studies of ECE classrooms, yet we still may have lacked sufficient power. For comparison, a study conducted in Ecuador also examined associations between the CLASS and child outcomes, but in kindergarten, for a sample of more than 23,000 children in more than 400 classrooms (Araujo et al. 2016). The authors found effect sizes of 0.07 to 0.11, which are not far off in magnitude from the associations we found. Their larger sample size likely allowed for greater precision in their estimates to detect effects of the CLASS. If the expected associations of the CLASS with children's outcomes are in the general range of a tenth of a standard deviation and often null like others have reported (Burchinal 2018; Perlman et al. 2016), then it is likely we would need an even larger sample than we currently have in these analyses, particularly to examine subgroups and thresholds.

D. Recommendations

Given the limited associations of the Pre-K CLASS with Head Start children's outcomes, we suggest potential next steps for measuring classroom quality.

Consider using the Pre-K CLASS in conjunction with other measures. There is broad consensus in the field that early and effective teacher-child interactions are important and form the foundation of children's learning and development (Hamre and Pianta 2005; Hamre et al. 2013; Mashburn et al. 2008). It is likely that no single measure can fully capture classroom quality across contexts (Kane et al. 2013). The Pre-K CLASS may best be used alongside other measures that provide related information about other aspects of quality such as the curriculum used, how teachers conduct ongoing monitoring and assessment of children's progress, and individualized teaching. For example, by using the Pre-K CLASS with a more individualized child-level measure, we can better understand and guide teachers' practice about when and how to adjust interactions with children to meet their individual needs. In addition, by collecting information on the curriculum used and drawing on information gathered from the Pre-K CLASS, we can begin to better understand how the CLASS domains and dimensions are reflected in curricula. If certain curricula do not provide sequenced content within engaging activities, then we would not expect teachers to be rated highly on Instructional Support. It would not necessarily mean that low Instructional Support is attributable only to the teacher.

Reconsider existing definitions and measures of classroom quality to ensure they are culturally responsive. Public ECE programs, such as Head Start, serve a diverse group of children. It might be that what is traditionally considered high quality in the field might need to be conceptualized in a different way that is specific to the experiences of all children. There may be some aspects of quality that are universal (for example, higher levels of teacher education or sensitive and responsive teachers) and some aspects that are specific to dual language learners or children of color (for example, equitable and culturally-responsive interactions with teachers). Current measures of classroom quality might need to be revised or updated to reflect this, and tested empirically with different samples to ensure that they are valid for diverse populations.

Consider using the Pre-K CLASS for continuous quality improvement. Our results, in addition to evidence by other researchers (Casabianca et al., 2015; Sabol et al., 2020; Styck et al. 2020), support the recommendation that Pre-K CLASS scores should not be used for high-stakes decisions (Mashburn 2017). Any measure that is going to be used for high-stakes purposes must be held to high measurement standards because of the consequences that stem from the scores. Rather, the Pre-K CLASS could be used as a tool for quality improvement. For example, it may be more appropriate to use the scores as a formative tool in a lower-stakes context to support teachers' professional development, as has been suggested by others (Pianta et al. 2008b; Sandilos et al. 2017). By using it for such purposes, the tool can be used to identify specific teaching practices that can be strengthened and provide teachers with professional development supports. Regardless of whether the Pre-K CLASS is used as a formative or summative tool, it is important to make sure that different observers are making the same judgments about what they are seeing in the classroom and how it should be scored. Teachers should be advancing their professional development and improving their teaching based on reliable observations of their practices and not based on which observer is assessing them.

Support research and development of new and revised measures of classroom quality. Reports of modest or null associations between commonly used measures of classroom quality and children's school readiness outcomes calls for an increased focus on measures that may more strongly predict children's

outcomes. This might mean developing new and improved measures of quality in addition to revising the current measures in place. For example, future measurement work may be necessary to determine the best way to capture accurate estimates of classroom quality to predict children's outcomes with existing measures. It might be that there is a particular number of observation occasions or longer observation periods needed to attain acceptable reliability. Or, in addition to capturing how teachers are interacting with children with measures such as the CLASS, it will also be important to know about the quality of teachers' instruction depending on what they are teaching. New measures could be developed that focus on the quality of teachers' interactions with children when engaged in specific activities. For example, such a measure could capture how self-regulation skills and strategies are taught to children and how the teacher uses teachable moments that might require children's self-regulation. Expanded measures of classroom quality could capture some of this nuance in addition to more general teaching practices.

V. Conclusions

The Pre-K CLASS is predicated on the assumption that high-quality interactions with teachers shape children's learning and development. However, the models in the current analyses showed limited associations between classroom quality and children's school readiness outcomes. We also found limited significant findings when examining these associations for key subgroups and when examining whether there were thresholds in these associations. Although the Pre-K CLASS provides a theoretically grounded approach to defining classroom quality, there is still more work to be done to better understand and improve on this widely administered measure of quality.

References

Administration for Children and Families. "Head Start Programs." 2020. Available at https://www.acf.hhs.gov/ohs/about/head-start. Accessed September 2020.

Aikens, N., C. Bush, P. Gleason, L. Malone, and L. Tarullo. "Tracking Quality in Head Start Classrooms: FACES 2006 to FACES 2014 Technical Report." OPRE Report 2016-95. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2016.

Aikens, N., A. Kopack Klein, E. Knas, M. Reid, A. M. Esposito, M. Manley, L. Malone, et al. "Descriptive Data on Head Start Children and Families from FACES 2014: Fall 2014 Data Tables and Study Design." OPRE Report #2017-97. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2017.

Aikens, N., T. Nguyen, and J. F. Harding. "Recent Evidence on the Pre-K CLASS and Children's Outcomes." Princeton, NJ: Mathematica, 2021.

Allen, J., A. Gregory, A. Mikami, J. Lun, B. Hamre, and R. Pianta. "Observations of Effective Teacher–Student Interactions in Secondary School Classrooms: Predicting Student Achievement with the Classroom Assessment Scoring System—Secondary." *School Psychology Review*, vol. 42, no. 1 2013, pp. 76–98.

Araujo, M., P. C. Caridad, Y. Cruz-Aguayo, and N. Schady. "Teacher Quality and Learning Outcomes in Kindergarten." *The Quarterly Journal of Economics*, vol. 131, no. 3, 2016, pp. 1415–1453.

Blair, C. "School Readiness: Integrating Cognition and Emotion in a Neurobiological Conceptualization of Children's Functioning at School Entry." *American Psychologist*, vol. 57, no. 2, 2002, pp. 111–127.

Bilbrey, C., E. Vorhaus, and D. C. Farran. "Teacher Observation in Preschool." Nashville, TN: Peabody Research Institute, Vanderbilt University, 2007.

Broekhuizen, M. L., I. L. Mokrova, M. R. Burchinal, and P. T. Garrett-Peters. "Classroom Quality at Pre-Kindergarten and Kindergarten and Children's Social Skills and Behavior Problems." *Early Childhood Research Quarterly*, vol. 36, 2016, pp. 212–222.

Build Initiative. "Quality Compendium." 2021. Available at https://qualitycompendium.org/. Accessed September 2021.

Burchinal, M. "Measuring Early Care and Education Quality." *Child Development Perspectives*, vol. 12, no. 1, 2018, pp. 3–9.

Burchinal, M., K. Garber, T. Foster, M. Bratsch-Hines, Ximena Franco, and Ellen Peisner-Feinberg. "Relating Early Care and Education Quality to Preschool Outcomes: The Same or Different Models for Different Outcomes?" *Early Childhood Research Quarterly*, vol. 55, 2021, pp. 35–51.

Burchinal, M., N. Vandergrift, R. Pianta, and A. Mashburn. "Threshold Analysis of Association Between Child Care Quality and Child Outcomes for Low-Income Children in Pre-Kindergarten Programs." *Early Childhood Research Quarterly*, vol. 25, no. 2, 2010, pp. 166–176. https://doi.org/10.1016/j.ecresq.2009.10.00.

Burchinal, M., Y. Xue, A. Auger, H. C. Tien, A. Mashburn, E. Peisner-Feinberg, E. W. Cavadel, et al. "Quality Thresholds, Features, and Dosage in Early Care and Education: Secondary Data Analyses of Child Outcomes: III. Testing for Quality Thresholds and Features in Early Care and Education." *Monographs of the Society for Research in Child Development*, vol. 81, no. 2, 2016, pp. 46–63.

Buysse, V., E. Peisner-Feinberg, E. Soukakou, A. Fettig, J. Schaaf, and M. Burchinal. "Using Recognition & Response (R&R) to Improve Children's Language and Literacy Skills: Findings from Two Studies." *Early Childhood Research Quarterly*, vol. 36, 2016, pp. 11–20.

Cabell, S. Q., J. DeCoster, J. LoCasale-Crouch, B. K. Hamre, and R. C. Pianta. "Variation in the Effectiveness of Instructional Interactions Across Preschool Classroom Settings and Learning Activities." *Early Childhood Research Quarterly,* vol. 28, 2013, pp. 820–830. doi.org/10.1016/j.ecresq.2013.07.007.

Carr, R. C., I. L. Mokrova, L. Vernon-Feagans, and M. R. Burchinal. "Cumulative Classroom Quality During Pre-Kindergarten and Kindergarten and Children's Language, Literacy, and Mathematics Skills." *Early Childhood Research Quarterly,* vol. 47, 2019, pp. 218–228.

Casabianca, J. M., J. R. Lockwood, and D. F. McCaffrey. Trends in Classroom Observation Scores. *Educational and Psychological Measurement*, vol. 75, no. 2, 311–337, 2015.

Cash, A. H., A. Ansari, K. J. Grimm, and R. C. Pianta. "Power of Two: The Impact of 2 Years of High Quality Teacher Child Interactions." *Early Education and Development*, vol. 30, no. 1, 2019, pp. 60–81.

Chaudry, A., T. Morrissey, C. Weiland, and H. Yoshikawa. "Universal Preschool. Cradle to Kindergarten: A New Plan to Combat Inequality." New York: Russell Sage Foundation, 2017.

Chien, N. C., C. Howes, M. Burchinal, R. C. Pianta, S. Ritchie, D. M. Bryant, R. M. Clifford, et al. "Children's Classroom Engagement and School Readiness Gains in Prekindergarten." *Child Development*, vol. 81, no. 5, 2010, pp. 1534–1549.

Currie, J., and D. Thomas. "School Quality and the Longer-Term Effects of Head Start." *Journal of Human Resources*, vol. 35, no. 4, 2000, pp. 755–774.

Diamond, A., and C. Taylor. "Development of an Aspect of Executive Control: Development of the Abilities to Remember What I Said and to 'Do as I Say, Not as I Do.'" *Developmental Psychobiology*, vol. 29, no. 4, 1996, pp. 315–334.

Doran, E., N. Aikens, M. Dang, A. Li, A. Kopack Klein, N. Reid, M. Scott, et al. "Understanding Head Start Children and Families in Spring 2020: Descriptive Tables and Study Design FACES 2019 Report Design." OPRE Report 2022-16. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2022.

Downer, J. T., L. M. Booren, O. K. Lima, A. E. Luckner, and R. C. Pianta. "The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary Reliability and Validity of a System for Observing Preschoolers' Competence in Classroom Interaction." *Early Childhood Research Quarterly*, vol. 25, 2010, pp. 1–16.

Dunn, L. M., L. L. Dunn, and D. M. Dunn. *Peabody Picture Vocabulary Test, Fourth Edition Examiner's Manual and Norms Booklet*. Circle Pines, MN: American Guidance Service, 2006.

Elliott, S. N., F. M. Gresham, T. Freeman, and G. McCloskey. "Teacher and Observer Ratings of Children's Social Skills: Validation of the Social Skills Rating Scales." *Journal of Psychoeducational Assessment*, vol. 6, no. 2, 1988, pp. 152–161.

Entwisle, D. R., K. L. Alexander, and L. S. Olson. *Children, Schools, and Inequality*. Boulder, CO: Westview Press, 1997.

Farran, D. C., D. Meador, C. Christopher, K. T. Nesbitt, and L. E. Bilbrey. "Data-Driven Improvement in Prekindergarten Classrooms: Report from a Partnership in an Urban District." *Child Development*, vol. 88, no. 5, 2017, pp. 1466–1479.

Fuligni, A. S., C. Howes, Y. Huang, S. Soliday Hong, and S. Lara-Cinisomo. "Activity Settings and Daily Routines in Preschool Classrooms: Diverse Experiences in Early Learning Settings for Low-Income Children." *Early Childhood Research Quarterly*, vol. 27, no. 2, 2012, pp. 198–209.

Gordon, R. A., and F. Peng. "Evidence Regarding the Domains of the CLASS Pre-K in Head Start Classrooms." *Early Childhood Research Quarterly*, vol. 53, 2020, pp. 23–39.

Gresham, F. M., and S. N. Elliot. *Social Skills Rating System*. Circle Pines, MN: American Guidance Service, 1990.

Hamre, B., B. Hatfield, R. Pianta, and F. Jamil. "Evidence for General and Domain-Specific Elements of Teacher–Child Interactions: Associations with Preschool Children's Development." *Child Development,* vol. 85, no. 3, 2014, pp. 1257–1274.

Hamre, B. K. and R.C. Pianta. Can Instructional and Emotional Support in the First-Grade Classroom Make a Difference for Children at Risk of School Failure? *Child Development*, vol. 76, no. 5, pp.949–967, 2005.

Hamre, B. K., R. C. Pianta, J. T. Downer, J. DeCoster, A. J. Mashburn, S. M. Jones, J. L. Brown, et al. Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal*, vol. 113, no. 4, 461–487, 2013.

Hanno, E. C., K. E. Gonzalez, S. M. Jones, and N. K. Lesaux. "Linking Features of Structural and Process Quality Across the Landscape of Early Education and Care." *AERA Open,* vol. 7, 2021, article no. 23328584211044519.

Harding, J. F., T. Nguyen, and N. Aikens. "Thresholds of Classroom Quality." Princeton, NJ: Mathematica, forthcoming.

Hatfield, B. E., M. R. Burchinal, R. C. Pianta, and J. Sideris. "Thresholds in the Association Between Quality of Teacher–Child Interactions and Preschool Children's School Readiness Skills." *Early Childhood Research Quarterly*, vol. 36, 2016, pp. 561–571.

Head Start Early Childhood Learning and Knowledge Center. "Designation Renewal System." 2020. Available at https://eclkc.ohs.acf.hhs.gov/designation-renewal-system. Accessed September 2020.

Hillemeier, M. M., P. L. Morgan, G. Farkas, and S. A. Maczuga. "Quality Disparities in Child Care for At-Risk Children: Comparing Head Start and Non-Head Start Settings." *Maternal and Child Health Journal*, vol. 17, no. 1, 2013, pp. 180–188.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment." Seattle, WA: Bill & Melinda Gates Foundation, 2013.

Kopack Klein, A., B. L. Carlson, N. Aikens, A. Bloomenthal, J. West, L. Malone, M. Dang, et al. "Head Start Family and Child Experiences Survey (FACES 2014): User's Manual." Princeton, NJ: Mathematica Policy Research, 2017.

Lee, V. E., and S. Loeb. "Where Do Head Start Attendees End Up? One Reason Why Preschool Effects Fade Out." *Educational Evaluation and Policy Analysis*, vol. 17, no. 1, 1995, pp. 62–82.

Malone, L., C. Baxter, and S. Atkins-Burnett. "Advancing the Measurement of Early Care and Education Quality to Support Young Children's Development." Paper presented at the Society for Research on Child Development Biennial Meeting, webinar, 2021.

Mantzicopoulos, P., B. F. French, H. Patrick, J. S. Watson, and I. Ahn. "The Stability of Kindergarten Teachers' Effectiveness: A Generalizability Study Comparing the Framework for Teaching and the Classroom Assessment Scoring System." *Educational Assessment*, vol. 23, no. 1, 2018, pp. 24–46.

Martin, N., and R. Brownell. "Expressive One-Word Picture Vocabulary Test-4th Edition." Novato, CA: Academic Therapy Publications, 2010.

Martone, A., and Sireci, S. G. "Evaluating Alignment Between Curriculum, Assessment, and Instruction." *Review of Educational Research*, vol. 79, no. 4, 2019, pp. 1332–1361.

Mashburn, A. J. "Evaluating the Validity of Classroom Observations in the Head Start Designation Renewal System." *Educational Psychologist*, vol. 52, no. 1, 2017, pp. 38–49.

Mashburn, A., L. M. Justice, A. McGinty, and L. Slocum. "The Impacts of a Scalable Intervention on the Language and Literacy Development of Rural Pre-Kindergartners." *Applied Developmental Science*, vol. 20, no. 1, 2016, pp. 61–78.

Mashburn, A. J., R. C. Pianta, B. K. Hamre, J. T. Downer, O.A. Barbarin, D. Bryant, M. Burchinal, et al. Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills. *Child Development*, vol. 79, no. 3, 732–749, 2008.

McGuire, Patrick R., Mable Kinzie, Kateri Thunder, and Robert Berry. "Methods of Analysis and Overall Mathematics Teaching Quality in At-Risk Prekindergarten Classrooms." *Early Education and Development*, vol. 27, no. 1, 2016, pp. 89–109.

Perlman, M., O. Falenchuk, B. Fletcher, E. McMullen, J. Beyene, and P.S. Shah. "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*, vol. 11, no. 12, 2016, pp. 1–33.

Peterson, J., and N. Zill. "Marital Disruption, Parent-Child Relationships, and Behavior Problems in Children." *Journal of Marriage and the Family*, vol. 48, 1986, pp. 295–307.

Phillips, D. A., M. W. Lipsey, K. A. Dodge, R. Haskins, D. Bassok, M. R. Burchinal, G. J. Duncan, et al. "Puzzling It Out: The Current State of Scientific Knowledge on Pre-Kindergarten Effects – A Consensus Statement." Washington, DC: Brookings Institution, 2017.

Pianta, R. C., A. J. Mashburn, J.T. Downer, B.K. Hamre, and L. Justice. "Effects of Web-Mediated Professional Development Resources on Teacher–Child Interactions in Pre-Kindergarten Classrooms." *Early Childhood Research Quarterly*, vol. 23, no. 4, 2008, pp. 431–451.

Pianta, R., J. Downer, and B. Hamre. "Quality in Early Education Classrooms: Definitions, Gaps, and Systems." *The Future of Children*, 2016, pp. 119–137.

Pianta, R. C., J. E. Whittaker, V. Vitiello, E. Ruzek, A. Ansari, T. Hofkens, and J. DeCoster. "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*, vol. 66, no. 101084, 2020, pp. 1–10.

Pianta, R. C., K. M. La Paro, and B. K. Hamre. *Classroom Assessment Scoring System™: Pre-K Manual*. Baltimore, MD: Paul H. Brookes Publishing, 2008.

Purtell, K. M. and A. Ansari. "Classroom Age Composition and Preschoolers' School Readiness: The Implications of Classroom Quality and Teacher Qualifications." *AERA Open*, vol. 4, no. 1, 2018, pp. 2332858418758300.

Porter, A. C. and Smithson, J. L. "Defining, Developing, and Using Curriculum Indicators." CPRE Research Report Series. Philadelphia, PA: Consortium for Policy Research in Education, 2001.

Rojas, N. M. and R. M. Abenavoli. "Preschool Teacher-Child Relationships and Children's Expressive Vocabulary Skills: The Potential Mediating Role of Profiles of Children's Engagement in the Classroom." *Early Childhood Research Quarterly,* vol. 56, 2021, pp. 225–235.

Rothstein, J. "Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables." *Education Finance and Policy*, vol. 4, no. 4, 2009, pp. 537–571.

Sabol, T. J., E. C. Ross, and A. Frost. Are all Head Start Classrooms Created Equal? Variation in Classroom Quality within Head Start Centers and Implications for Accountability Systems. *American Educational Research Journal*, vol. 57, no. 2, 2020, pp. 504–534.

Sandilos, L. E., S. W. Shervey, J. C. DiPerna, P. Lei, and W. Cheng. Structural Validity of CLASS K–3 in Primary Grades: Testing Alternative Models. *School Psychology Quarterly*, vol. 32, no. 2, 2017, pp. 226–239.

Sarama, J., D. H. Clements, C. B. Wolfe, and M. E. Spitler. "Professional Development in Early Mathematics: Effects of an Intervention Based on Learning Trajectories on Teachers' Practices." *Nordic Studies in Mathematics Education*, vol. 21, no. 4, 2016, pp. 29–55.

Smith-Donald, R., C. Raver, T. Hayes, and B. Richardson. "Preliminary Construct and Concurrent Validity of the Preschool Self-Regulation Assessment (PSRA) for Field-Based Research." *Early Childhood Research Quarterly*, vol. 22, 2007, pp. 173–187.

Snow, C. E., and S. B. Van Hemel. *Early Childhood Assessment: Why, What, and How*. Washington, DC: The National Academies Press, 2008.

Soliday Hong, S. L., T. J. Sabol, M. R. Burchinal, L. Tarullo, M. Zaslow, and E. Peisner-Feinberg. "ECE Quality Indicators and Child Outcomes: Analyses of Six Large Child Care Studies." *Early Childhood Research Quarterly*, vol. 49, 2019, pp. 202–217.

Styck, K. M., C. J. Anthony, L. E. Sandilos, and J. C. DiPerna. "Examining Rater Effects on the Classroom Assessment Scoring System." *Child Development*, vol. 92, no. 3, 2021, pp. 976–993.

U.S. Department of Education. "Early Childhood Longitudinal Study—Kindergarten Class of 1998–99 (ECLS–K), Psychometric Report for Kindergarten Through First Grade." NCES 2002–05. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2002.

Weiland, C., K. Ulvestad, J. Sachs, and H. Yoshikawa. "Associations Between Classroom Quality and Children's Vocabulary and Executive Function Skills in an Urban Public Prekindergarten Program." *Early Childhood Research Quarterly*, vol. 28, no. 2, 2013, pp. 199–209.

Whittaker, J. V., M. B. Kinzie, V. Vitiello, J. DeCoster, C. Mulcahy, and E. A. Barton. "Impacts of an Early Childhood Mathematics and Science Intervention on Teaching Practices and Child Outcomes." *Journal of Research on Educational Effectiveness,* vol. 13, no. 2, 2020, pp. 177–212.

Woodcock, R.W., K. McGrew, and N. Mather. *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Publishing, 2001.

Yoshikawa, H., C. Weiland, J. Brooks-Gunn, M. R. Burchinal, L. M. Espinosa, W. T. Gormley, J. Ludwig, et al. "Investing in Our Future: The Evidence Base on Preschool Education." Washington, DC: Society for Research in Child Development, 2013.

Zaslow, M., R. Anderson, Z. Redd, J. Wessel, L. Tarullo, and M. Burchinal. "Quality Dosage, Thresholds, and Features in Early Childhood Settings: A Review of the Literature." OPRE 2011-5. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2010.

Appendix A

Supplemental Tables

Table A.1. Correlations and descriptive statistics for Pre-K CLASS dimensions

	Concept develop-ment	Quality of feedback	Language modeling	Positive climate	Negative climate	Teacher sensitivity	Regard of student per-spectives	Behavior manage-ment	Pro-ductivity	Instruc-tional learning formats
Concept development	1.00									
Quality of feedback	0.83	1.00								
Language modeling	0.71	0.87	1.00							
Positive climate	0.59	0.55	0.52	1.00						
Negative climate	0.51	0.45	0.52	0.78	1.00					
Teacher sensitivity	0.52	0.57	0.58	0.86	0.78	1.00				
Regard of student perspectives	0.53	0.54	0.54	0.86	0.66	0.84	1.00			
Behavior management	0.48	0.49	0.52	0.77	0.67	0.82	0.74	1.00		
Productivity	0.51	0.54	0.55	0.84	0.61	0.88	0.71	0.78	1.00	
Instructional learning formats	0.55	0.54	0.55	0.79	0.62	0.71	0.74	0.68	0.88	1.00
Median	2.00	2.50	2.50	5.50	7.00	5.00	4.75	5.00	5.00	4.25
SD	0.91	0.95	0.92	0.47	0.48	0.48	0.51	0.42	0.43	0.56
Range (Min-Max)	1.00-6.33	1.00-6.75	1.00-6.25	3.00-7.00	3.67-7.00	2.67-7.00	1.67-7.00	2.33-7.00	2.50-7.00	2.00-6.33

Note: Statistics are weighted to represent all children who were enrolled in Head Start in fall and were still enrolled in spring. All correlations statistically significant at $p < .05$. Negative climate was reverse coded. Descriptive statistics were computed for children (N = 1,454) included in analyses of the Peabody Picture Vocabulary Test–Fourth Edition.

Table A.2. Results of regressing school readiness outcomes on Pre-K CLASS dimensions

	Expressive vocabulary (SE)	Receptive vocabulary (SE)	Early writing (SE)	Letter-word knowledge (SE)	Early math (SE)	Executive function (SE)	Approaches to learning (SE)	Behavioral problems (SE)	Social skills (SE)
Positive climate	-0.01	0.00	0.01	-0.02	0.02	0.00	0.01	0.02	-0.02
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.04)	(0.05)
Negative climate	-0.04*	-0.02	-0.05	-0.07+	-0.03	-0.01	-0.01	-0.02	0.02
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.04)	(0.04)
Teacher sensitivity	0.02	0.01	0.02	-0.04	0.02	-0.01	0.04	0.01	0.01
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.04)	(0.05)
Regard for student perspectives	0.01	0.02	-0.01	0.00	0.03	0.03	-0.01	0.03	-0.05
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.04)	(0.05)
Concept development	0.02	0.02	0.02	0.04	0.06**	0.04	-0.02	0.07+	-0.01
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.04)	(0.05)
Quality of feedback	0.01	0.01	0.07+	0.02	0.06**	0.01	0.05	0.03	0.04
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.05)	(0.04)	(0.05)
Language modeling	0.01	0.02	0.05	0.01	0.04*	0.00	0.05	0.03	0.02
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.05)	(0.04)	(0.05)
Behavior management	0.01	0.00	0.01	-0.04	0.03	-0.01	0.00	0.04	-0.03
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.04)	(0.05)
Productivity	-0.01	0.01	0.03	0.00	0.02	0.03	0.03	0.03	-0.02
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.04)	(0.05)
Instructional learning formats	-0.02	0.01	0.01	0.02	0.02	0.01	0.03	0.04	0.03
	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.04)	(0.04)	(0.04)	(0.05)
N	1,213	1,454	1,355	1,343	1,354	982	1,516	1,517	1,514

Note: Statistics are weighted to represent all children who were enrolled in Head Start in fall and were still enrolled in spring. All models control for child, teacher, and classroom characteristics. Child characteristics include the fall score on the respective outcome, whether the child was assessed in English or Spanish in the fall, child race and ethnicity, child sex, child age in months at time of spring assessment, months between fall and spring assessments, IEP status, and poverty status. Teacher characteristics include race and ethnicity, education, and years of teaching experience. Classroom characteristics include size and part-day status. We also included dummy variables for missing IEP status and poverty status to decrease missing data.

+ $p \leq 0.10$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

SE = standard error.

Table A.3. Threshold regression results for Pre-K CLASS domains, using median-split cut points

	Emotional Support			Classroom Organization			Instructional Support			
	Below median quality (SE)	Above median quality (SE)	Differ?	Below median quality (SE)	Above median quality (SE)	Differ?	Below median quality (SE)	Above median quality (SE)	Differ?	N
Expressive vocabulary	0.00	0.04	No	0.00	-0.04	No	-0.01	0.00	No	1,213
	(0.01)	(0.12)		(0.01)	(0.08)		(0.04)	(0.05)		
Receptive vocabulary	0.00	-0.03	No	-0.02*	-0.06	No	0.00	0.01	No	1,454
	(0.01)	(0.09)		(0.01)	(0.07)		(0.02)	(0.04)		
Early writing	0.00	-0.02	No	-0.01	-0.03	No	-0.06	0.01	No	1,355
	(0.02)	(0.15)		(0.02)	(0.12)		(0.05)	(0.07)		
Letter-word knowledge	0.00	-0.14	No	0.00	-0.07	No	-0.02	0.01	No	1,343
	(0.01)	(0.14)		(0.02)	(0.11)		(0.05)	(0.06)		
Early math	0.00	0.09	No	0.01	0.09	No	0.01	0.10*	Yes*	1,354
	(0.01)	(0.12)		(0.01)	(0.09)		(0.04)	(0.05)		
Executive function	0.01	0.02	No	-0.02	-0.09	No	0.02	0.06	No	982
	(0.02)	(0.14)		(0.02)	(0.12)		(0.05)	(0.06)		
Approaches to learning	-0.04*	-0.24	No	-0.02	-0.06	No	0.00	0.02	No	1,516
	(0.02)	(0.17)		(0.02)	(0.14)		(0.05)	(0.07)		
Behavioral problems	0.01	0.15	No	-0.01	0.10	No	-0.02	0.05	No	1,517
	(0.01)	(0.13)		(0.02)	(0.11)		(0.04)	(0.06)		
Social skills	-0.03	-0.22	No	-0.03	-0.19	No	0.02	0.01	No	1,514
	(0.02)	(0.18)		(0.02)	(0.14)		(0.05)	(0.07)		

Note: Statistics are weighted to represent all children who were enrolled in Head Start in fall and were still enrolled in spring. 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, teacher, and classroom characteristics. Child characteristics include the fall score on the respective outcome, whether the child was assessed in English or Spanish in the fall, child race and ethnicity, child sex, child age in months at time of spring assessment, months between fall and spring assessments, IEP status, and poverty status. Teacher characteristics include race and ethnicity, education, and years of teaching experience. Classroom characteristics include size and part-day status. We also included dummy variables for missing IEP status and poverty status to decrease missing data.

+ $p \leq 0.10$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

SE = standard error.