TYPE Original Research
PUBLISHED 30 April 2024
DOI 10.3389/feduc.2024.1370573

OPEN ACCESS

EDITED BY
Cheng Yong Tan,
The University of Hong Kong,
Hong Kong SAR, China

REVIEWED BY
David Paige,
Northern Illinois University, United States
Jane Ashby,
Mount St. Joseph University, United States

*CORRESPONDENCE
Yusuf Kara
✉ ykara@smu.edu

RECEIVED 15 January 2024
ACCEPTED 08 April 2024
PUBLISHED 30 April 2024

CITATION
Kara Y, Farmer DE, Wright A, Kamata A, Hujar J and Baker SK (2024) Examining the effects of prekindergarten to second grade classroom quality on early elementary achievement.
Front. Educ. 9:1370573.
doi: 10.3389/feduc.2024.1370573

Examining the effects of prekindergarten to second grade classroom quality on early elementary achievement

Yusuf Kara[1]*, Dylan E. Farmer[1], Annie Wright[1], Akihito Kamata[2], Julia Hujar[1] and Scott K. Baker[3]

[1]Center on Research and Evaluation, Southern Methodist University, Dallas, TX, United States, [2]Department of Education Policy and Leadership, Southern Methodist University, Dallas, TX, United States, [3]Wheelock College of Education & Human Development, Boston University, Boston, MA, United States

Previous research has investigated the influence of preschool interventions in early elementary environments with some results suggesting that high quality early preschool experiences may have meaningful effects on student academic outcomes not only immediately following the preschool years but also in the long term. This study aimed to investigate this topic further by using a novel cross-classified model to account for the complex hierarchical structure of a large-scale longitudinal and multilevel dataset. First, we examined the carryover classroom effects from prekindergarten (PK), kindergarten (K), and first grade on later grades' reading and mathematics scores. Second, we explored the effect of PK to second grade classroom quality on students' K, first, and second grade reading and mathematics scores. Our analyses yielded significant PK carryover classroom effects for reading; these results provide evidence that a student's PK classroom continues to impact that student for at least three years after. Additionally, one domain of PK classroom quality —classroom organization—was found to be positively related to students' K, first, and second grade reading outcomes.

KEYWORDS

prekindergarten, classroom quality, fadeout effect, cross-classified models, student achievement

1 Introduction

Federal funding for public prekindergarten (PK) programs has been supported by decades of research that has demonstrated both short- and long-term student benefits from attending such programs (Reynolds et al., 2001, 2010; Campbell et al., 2002; Gormley et al., 2005; Schweinhart et al., 2005; Reynolds and Ou, 2011). However, the body of literature is not in full agreement regarding the conditions that influence whether effects of PK are sustained or fade out. Research related to the impact of PK educational programs has focused primarily on classroom quality in these programs, and quality's impact on whether the benefits are sustained or fade out over time (Gormley et al., 2005; Howes et al., 2008; Mashburn et al., 2008). Despite quality's supposed role in beneficial student outcomes, the literature on this topic is mixed as well (Weiland et al., 2013). This study aims to add to the literature of the effect of quality PK programs on student achievement by utilizing a unique methodological approach that allows for the consideration of carryover random classroom effects from year to year.

1.1 Sustained effects vs. fadeout effect

Favorable outcomes for students who attended a PK program have long been documented in the extant literature (e.g., Campbell et al., 2002; Gormley et al., 2005; Belfield et al., 2006; Magnuson et al., 2007; Reynolds and Ou, 2011; Campbell et al., 2014). Specifically, sustained benefits have been shown for populations of low-income children (Reynolds et al., 2001; Pungello et al., 2010; Duncan and Magnuson, 2013). One well-known longitudinal study documenting the lasting effects of PK program attendance was the HighScope Perry Preschool project study (Schweinhart and Weikart, 1993). By age 27, participants in the program were more likely to have completed high school, had fewer encounters with the justice system, had higher income, were more likely to own their home, and fewer received welfare than those who did not participate in the program. Similarly, the Abecedarian Project was another randomized longitudinal study designed to determine the impact of preschool later in life (Campbell et al., 2002). Results indicated that those who attended the preschool program earned higher scores on measures of intelligence and achievement, had higher educational attainment, and were more likely to attend a four-year college or university.

Despite this evidence documenting lasting benefits of PK, a parallel body of research demonstrates a fadeout effect, showing the diminishing returns of attending PK programs (Hill et al., 2015). Several studies (e.g., Loeb et al., 2007; Magnuson et al., 2007) determined that the average developmental benefits of attending a PK program were small to moderate, with most benefits fading out over the elementary school years. Also, several studies provided evidence that after approximately two years of schooling, differences between those who attended PK and those who did not had nearly disappeared (Barnett, 1995; Puma et al., 2010; Bassok et al., 2019). Through a meta-analysis of over 60 early childhood interventions, Bailey et al. (2017) found a 50 percent decrease in PK impacts on both cognitive and achievement outcomes in kindergarten (average effect size decreased from 0.23 to 0.10), and another 50 percent decrease through first grade (effect size decreased from 0.10 to 0.05). Recently, Durkin et al. (2022) conducted a randomized trial to determine the longitudinal impacts of a public PK program and found that the children who attended the PK program had lower state achievement test scores in third through sixth grade than those who did not attend the PK program. Additionally, the children who attended the PK program also had higher rates of disciplinary referrals, higher rates of referral to special education services, and lower attendance rates than those who did not attend the PK program.

The results of these studies have begged the question, under what circumstances do beneficial PK effects sustain or fade out? Overall, the body of research on this topic argues that for benefits to sustain, both the PK program and the elementary school must be high-quality (Lee and Loeb, 1995; Reynolds et al., 2001, 2010; Pungello et al., 2010; Reynolds and Ou, 2011; Duncan and Magnuson, 2013; Pearman et al., 2019). After a student attends PK, attendance at a low-quality elementary school likely means that they are receiving repeated instruction on content they already know rather than receiving new instruction to support their academic growth (Lee and Loeb, 1995; Engel et al., 2013). Ansari and Pianta (2018) determined that the benefits of preschool were sustained through the end of elementary school when students attended a high-quality elementary school; however, less than a quarter of the benefits were sustained when children subsequently attended a low-quality elementary school. Similarly, Pearman et al. (2019) determined that having both exposure to highly effective elementary school teachers and attending a high-quality elementary school were associated with sustained benefits for PK participants in both mathematics and reading through third grade. The current study contributes to the literature by examining student enrollment in quality classrooms (not just schools).

1.2 Impact of early childhood education quality

The concept of quality in PK has been defined as "the interactive and interpersonal processes that facilitate positive learning experiences" (Pianta et al., 2008). Although some studies have used extant data such as teacher-student ratio as a proxy for classroom-level quality (e.g., Phillipsen et al., 1997; NICHD Early Child Care Research Network, 2002), quality is a construct best measured directly through observation. One common tool for measuring classroom quality is the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008). CLASS is a valid and reliable tool used by certified observers. It assesses three domains comprised of multiple dimensions that focus on different aspects of effective teacher-child interactions. Numerous studies have been conducted validating CLASS across diverse student populations, settings, and grade levels (Mashburn et al., 2008; Hamre et al., 2014). Previous research using CLASS scores as a measure of quality has demonstrated that higher classroom quality scores were able to predict growth in PK student achievement (Howes et al., 2008; Mashburn et al., 2008), academic skill gains in kindergarten and first grade (Hamre and Pianta, 2005; Burchinal et al., 2008), and student engagement (LaParo et al., 2004).

Overall, high-quality early childhood education classrooms have been shown to improve cognitive and social outcomes for children, especially those from low-income families (Reynolds et al., 2001; Campbell et al., 2002; Schweinhart et al., 2005; Pungello et al., 2010; Duncan and Magnuson, 2013). Additionally, when public PK programs were established as high quality, students in these programs demonstrated substantial gains in language, academic skills, and social skills (Gormley et al., 2005; Howes et al., 2008; Mashburn et al., 2008). Importantly, when the quality of a PK program hinged on an emphasis on preliteracy skills and/or structured mathematics tasks, the program effects were stronger (Pianta and Stuhlman, 2004; Mashburn et al., 2008; Fuller et al., 2017).

Mirroring the contradictory findings in the literature between sustained PK effects or fadeout, a similar discrepancy exists in the literature linking quality PK to student outcomes. Weiland et al. (2013) conducted an analysis of associations between multiple rating scales used for measuring classroom quality (one of which was CLASS) and children's academic outcomes. Their results were consistent with multiple other studies that found small or null associations between the two constructs (Burchinal et al., 2010; Zaslow et al., 2010, 2011). Weiland and colleagues hypothesized that these small associations could be due to measures of quality (e.g., observational tools) not adequately measuring the construct, or perhaps that overall PK classroom quality should be linked more to specific content areas rather than interpersonal processes. The second hypothesis was also suggested by Fuller et al. (2017) as well as Pianta et al. (2005). Specifically, they suggested that cognitive gains from PK programs may be absent when teachers exert more energy on

10.3389/feduc.2024.1370573

emotional support and therefore spend less time on nurturing preliteracy skills and early numeracy instruction.

It is also worth noting that varying findings regarding the effects of quality PK programs may have been a result of underutilized methodologies. Namely, the exploration of such distal effects requires availability of robust data sets and application of sophisticated statistical models that fully account for contextual impacts of clusters on individual-level outcomes. In fact, methodological challenges have been cited as a potential reason for the differing results regarding sustained effects of PK versus the fadeout of these effects, as well as the differing results on the impact of PK quality on student academic outcomes (Lipsey et al., 2018; Cash et al., 2019). As elaborated in the following sections, in addition to using a well-collected longitudinal dataset with direct observations of classroom quality, this study utilizes a rigorous model that considers random effects of all classrooms that students were taught in. Specifically, we incorporated the methods of Kwok et al. (2018) to build cross-classified models that account for the complex hierarchical structure of longitudinal data in addition to modeling carryover effects from previous grades' classrooms onto subsequent years. Thus, it is believed that the current work will contribute to the body of the research regarding the long-term impacts of high-quality PK programs on students' academic outcomes.

1.3 Purpose and research questions

The purpose of this study was twofold. First, we aimed to examine the carryover classroom effects from PK, kindergarten (K), and first (F) to later grades' reading and mathematics scores. Second, we explored the effect of PK to second grade (S) classroom quality on students' K, F, and S reading and mathematics scores. Ultimately, we aimed to add to the body of literature on both the fadeout (or lack thereof) of PK effects as well as the literature on the effects of classroom quality as measured by CLASS. We aimed to answer the following research questions:

1. How do student outcomes measured from kindergarten to second grade vary at the classroom level within the same academic year?
2. Do later years' student outcomes show a significant amount of variation across former years' classrooms? Are there any significant carryover classroom effects based on early years' classrooms?
3. What are the unique effects of classroom quality on students' academic outcomes? Do early years' classroom quality measures explain variability in later years' student outcomes? Do observed effects vary for the individual domains of the quality measure?

2 Methods

2.1 Sample

Data for this study were collected in a large urban school district in a Southern state from multiple cohorts of students across various grades. We linked students to their classroom teachers within a year at four grade levels from PK to S. We structured the data at the student level with all measures linked to students, including their individual achievement outcomes and quality measures of the classrooms in which they were taught. We selected a cohort with maximum data availability from PK to S after excluding classrooms with no quality measures. The final analysis sample included 1,040 students who attended PK at this district during the 2015–16 school year, attended K during 2016–17, F during 2017–18, and S during 2018–19. Descriptive statistics of the student demographic variables are summarized in Table 1.

The sample was balanced in terms of gender (50.9% females and 49.1% males). Representative of the district, the sample was majority Hispanic (79.6%), followed by Black or African American (18.1%). In addition, more than half of the students were identified as limited English proficient (LEP) in at least in one grade level (63% on average). Although the numbers changed slightly across years, on average 92% of the students were identified as economically disadvantaged, and 63% were identified as "at risk," based on the state education agency's predefined criteria related to socioeconomic challenges and performance. The analysis sample represents the population of the students within the district in terms of demographic characteristics.

TABLE 1 Descriptive statistics of the student demographic variables.

Variable	Levels	Grade level			
		PK	K	F	S
Gender	Female	529 (50.9%)			
	Male	511 (49.1%)			
Race	Black or Afr. Amr.	188 (18.1%)			
	Hispanic	827 (79.6%)			
	White & other	24 (2.3%)			
Limited Eng. Prof.	Yes	654 (62.9%)	653 (62.8%)	655 (63%)	643 (61.8%)
	No	386 (37.1%)	387 (37.2%)	385 (37%)	387 (37.2%)
Econ. Dis.	Yes	1,015 (97.6%)	913 (87.8%)	921 (88.6%)	960 (93.2%)
	No	25 (2.4%)	127 (12.2%)	119 (11.4%)	70 (6.8%)
At risk	Yes	659 (63.4%)	657 (63.2%)	650 (62.5%)	636 (61.2%)
	No	381 (36.6%)	383 (36.8%)	390 (37.5%)	394 (37.9%)

PK: prekindergarten, K: kindergarten, F: first grade, S: second grade. Some %'s did not add up to 100% at some grade levels due to missing observations.

Among 10,067 students who enrolled in public PK programs in the 2015–16 school year in this district, 72.1% of them were Hispanic/Latino, 54.0% of them were LEP, and 94% of them were economically disadvantaged (Texas Education Agency, 2024). Additionally, our sample included student data from all 149 elementary schools that provided full/half day PK programs in the district.

2.2 Measures

Student outcomes were measured by TerraNova and its Spanish version (SUPERA) for reading and mathematics achievements from K to S. All students took the mathematics test in English, whereas the reading test was taken either in English or Spanish. Both tests are norm-referenced and are administered once-per-year at the end of school year. Tests are reported to have evidence for three types of validity (construct, convergent, and concurrent) as well as high levels of reliability around 90s (Data Recognition Corporation, 2024b). We used the national percentiles provided by those tests and transformed them to standardized (Z) scores assuming raw test scores in the norm population were normally distributed. This scale transformation aimed at normalizing the distribution of the raw percentile scores. Then, we multiplied those Z scores by 10 for a better interpretation of the estimated model parameters, including variance components. We combined TerraNova and SUPERA scores for reading analyses because it was the test publisher's intention that the national percentiles from those two tests are comparable within a single sample (Data Recognition Corporation, 2024a). The transformation of the raw percentile scores to standardized scores was performed after combining TerraNova and SUPERA scores into a single set of outcome variables.

Student outcomes were predicted by classroom quality measured by the CLASS (Pianta et al., 2008) tool. CLASS data for the current analyses were collected on a rolling basis once per academic year during the spring semester. Classrooms from a total of 149 schools were observed by trained observers during the regular school day. More details about the CLASS tool and the classroom observation procedures are explained below.

CLASS is a framework for measuring the quality of teacher–child interactions that may contribute to student gains in academic, social, and behavioral areas. CLASS is a direct observation tool of teacher-child interactions within classrooms. Each classroom observation lasts approximately two hours, or four 30-min cycles comprised of 20 min of observation followed by 10 min of review and scoring. Certified observers use the CLASS tool to rate teachers on each of the ten sub-dimensions that make up the three domains of child-teacher interactions. The resulting three domains are emotional support (ES), classroom organization (CO), and instructional support (IS). ES addresses how teachers help children develop positive relationships, enjoy learning, feel comfortable in the classroom, and demonstrate independence. CO addresses how teachers manage the classroom to promote learning and engagement. IS involves how teachers promote children's thinking and problem solving, use feedback to deepen understanding, and use increasingly complex language skills. Sub-dimensions are rated on a scale of one (lowest) to seven (highest); dimensions are then averaged to create overall scores for each of the three domains. The validity of the three-factor structure of the CLASS tool along with adequate reliability levels (0.68, 0.78, and 0.93 for ES, CO, and IS respectively) are reported in Hamre et al. (2013). The thresholds for high quality in each domain are 5 for ES and CO and 3.25 for IS. Table 2 summarizes descriptive statistics of the CLASS domain scores for overall, low-, and high-quality classrooms.

2.3 Data analysis

The data had a complex hierarchical structure, where students and classrooms at different grades were cross-classified over multiple years. To accommodate the multiyear cross-classification structure in student outcome data, we employed a series of multivariate cross-classified linear models. Note that we may use the terms "year" and "grade" interchangeably in the remainder of the paper since all students matriculated to upper grades by the end of the school year (students who repeated a grade were excluded from the analyses).

TABLE 2 Descriptive statistics of the CLASS Scores per grade-level and domain.

Grade	Classrooms	*n*	Emotional support			Classroom organization			Instructional support		
			M	SD	Range	*M*	SD	Range	*M*	SD	Range
PK	All	1,040	5.81	0.81	2.75–7	5.64	0.83	2.42–6.75	3.04	0.98	1.08–5.83
	Low	642	5.46	0.78	2.75–6.63	5.3	0.86	2.42–6.67	2.41	0.53	1.08–3.75
	High	398	6.38	0.47	5–7	6.2	0.31	5.5–6.75	4.05	0.63	3.25–5.83
K	All	1,040	5.7	0.76	2.38–6.88	5.66	0.79	2.83–7	2.84	0.88	1–5.5
	Low	718	5.47	0.76	2.38–6.63	5.4	0.78	2.83–7	2.36	0.52	1–3.33
	High	322	6.24	0.38	5.19–6.88	6.24	0.45	5.08–6.92	3.9	0.53	3.25–5.5
F	All	1,040	5.77	0.78	1.38–7	5.69	0.84	1.08–7	3.42	1.13	1–6.75
	Low	499	5.34	0.83	1.38–6.81	5.31	0.95	1.08–6.83	2.5	0.49	1–3.25
	High	541	6.18	0.45	5.25–7	6.04	0.51	5–7	4.27	0.85	3.25–6.75
S	All	1,040	5.79	0.72	2.56–7	5.81	0.71	1.92–7	3.48	1.17	1.17–6.67
	Low	507	5.33	0.72	2.56–6.56	5.38	0.7	1.92–6.5	2.51	0.58	1.17–4.67
	High	533	6.23	0.37	5.06–7	6.21	0.43	5.08–7	4.41	0.77	3.25–6.67

PK: prekindergarten, K: kindergarten, F: first grade, S: second grade.

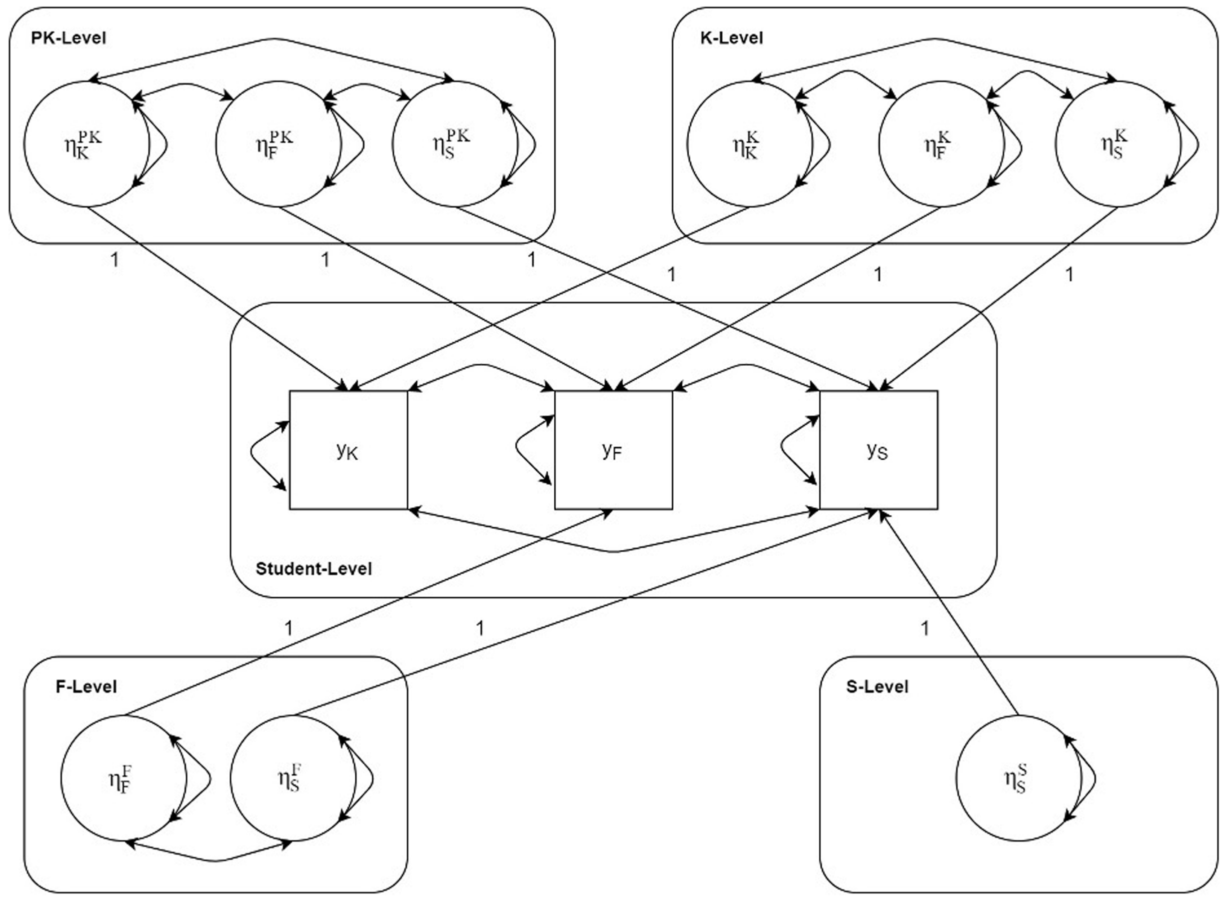

FIGURE 1
Path diagram of the unconditional model.

We used the xxM package (Mehta, 2013) in R (R Core Team, 2024) for the analyses. xxM implements multilevel and cross-classified models within a structural equation modeling (SEM) framework. Theoretically, it can model an unlimited number of data levels and multiple outcomes at a time, which perfectly fit the cross-classified longitudinal structure that we had in our data. In addition to random classroom effects within the same school year (i.e., the random effect of the current classroom), xxM can be used to fit a cross-classified model to estimate carryover effects, which are the random classroom effects on later years' student outcomes (Kwok et al., 2018). Readers are referred to Kwok et al. (2018) and Petscher and Schatschneider (2019) for more examples of fitting complex multilevel models with SEM approach by using xxM.

2.3.1 Unconditional models

Prior to modeling the effect of classroom quality on student outcomes, we analyzed an unconditional model to observe the partition of the variance in student outcomes. Results of this model provided answers to our first and second research questions. Estimation of the variance components from the unconditional model aimed to explore the degree to which student outcomes measured at a particular grade varied among the classrooms of this grade and also among former grades' classrooms. Note that the former variance component refers to within-year random classroom effect, whereas the latter refers to carryover random classroom effect, respectively.

An SEM-based path diagram is provided in Figure 1 to graphically represent within-year and carryover random classroom effects in the unconditional model. As shown in the path diagram, the model has five levels in total: the student level and four grade levels which are PK, K, F, and S. Variables represented by squares at the student level ($y_{K,}$ $y_{F,}$ and y_S) are the observed test scores collected by the end of K, F, and S. Note that there was no student outcome measured at the PK level. These observed scores are modeled as multivariate normal outcomes cross-classified with classrooms in multiple grade levels. All remaining variables represented by circles at the four grade levels (η's) are unobserved (i.e., latent) variables, namely the random classroom effects. Superscripts and subscripts of η's indicate grade levels of classrooms and the grade at which the outcome was measured, respectively. For example, η_K^K is the random effect of K classrooms associated with students' K outcomes, or simply, the within-year random classroom effect for the K outcome. η_F^K and η_S^K are the carryover effects from K classrooms associated with students' F and S outcomes, respectively.

A general equation for the unconditional model can be written in matrix format as

$$y_{P_i} = \gamma_P + \Lambda^K_{P_i}\eta^K_P + \Lambda^F_{P_i}\eta^F_P + \Lambda^S_{P_i}\eta^S_P + e_{P_i},$$

where a subscript P is used to represent the grade level at which the outcome is measured. Thus, y_{P_i} is student i's measured score at grade P, γ_P is the student-level intercept for the score measured at grade P, $\Lambda^K_{P_i}$ is the multiplier (0 or 1) for random classroom effects η^K_P originating from K classrooms and associated with the outcome measured at grade P. Remaining Λ and η terms can be deduced in the same way as the preceded explanation. Finally, e_{P_i} is the error term at the student-level per outcome measured at grade P. The distributional assumptions for the error and classroom-level variance terms can be expressed as $e_P \sim N(0,\Theta)$, $\eta^K_P \sim N(0,\Psi^K)$, $\eta^F_P \sim N(0,\Psi^F)$, and $\eta^S_P \sim N(0,\Psi^S)$. Θ and Ψ^K are 3×3, and Ψ^F is a 2×2 variance–covariance matrices, whereas Ψ^S is a 1×1 scalar value based on the current model. In the unconditional model, estimated variances of the η terms, namely, the current/carryover random classroom effects, were of interest. A large variance would indicate that classrooms at a specific grade differed considerably from each other in terms of the target student outcome. Smaller variances were expected for carryover effects compared to the variance of the random effects within the same school year.

We performed likelihood ratio tests to identify the meaningful sources of classroom-level variation on student outcomes so that the classroom-level quality measures could be used to explain such variability. We conducted likelihood ratio tests in a stepwise manner as follows. For each random effect, we fitted a constrained version of the unconditional model to the same data by excluding the target random effect (i.e., constraining its variance and covariances to zero). Then, we compared the $-2\times$ likelihood values of the full and the constrained models. Finally, we treated a statistically significant difference among the constrained vs. full model as statistical evidence for the significance of an omitted random effect.

2.3.2 Conditional models

To investigate the effect of classroom quality on student outcomes (as per third research question), we further analyzed conditional models, where classroom quality measures were used as predictors of student outcomes. We examined the effect of classroom quality through its relation to the classroom-level random effects at each grade. For example, we examined the effect of K quality in terms of predicting K scores as well as scores at later grades. Thus, the conditional models examined the effect of classroom quality on student outcomes not only within the same school year but also in later years. This allowed us to investigate if being in a high-quality classroom in early years affected students' future achievement scores. However, rather than using quality measures as predictors of all random effects, we only considered statistically significant quality effects identified by the likelihood ratio tests explained in the former section.

We conducted two conditional model analyses for reading and mathematics scores as outcomes, where we used the domain-specific classroom quality scores as predictors. We examined the effect of a specific quality measure on student outcomes through the statistical significance and magnitude of the estimated effect by the models. In addition to monitoring quality effects through estimated coefficients, we also examined the amount of variance explained by those quality measures.

3 Results

3.1 Unconditional model analysis results

Analysis results of the unconditional models are summarized in Table 3. Estimates of the covariances among random effects and estimates of the fixed effects (i.e., the intercepts) are not reported in this table to keep the focus on the variances. The likelihood ratio test results are also included in the table for additional interpretation of the statistical significance for the estimated variances of the random classroom effects. We used a strict criterion (alpha level as 0.01) for statistical significance to avoid type-1 errors in chi-square tests.

As explained earlier, the PK level of the model had only carryover classroom effects due to a lack of an outcome measure at this grade. Contrary to our expectation that PK effects might "fadeout" over time, variance estimates of PK carryover effects for reading did not systematically decrease from K to S. The K and F levels of the model had both carryover and within-school-year random effects. Estimated variances were larger for the within-year effects for these grades. Specifically, the K classroom effect on the K reading outcome (variance of η^K_K) was substantially larger than the effects on later years' outcomes at F (variance of η^K_F) and S (variance of η^K_S), which are defined as carryover classroom effects of K to later grades. This makes sense because the classroom effects were expected to be the highest for the current grade level. Nevertheless, there were also some exceptions where carryover effects from PK were larger than later years' carryover effects. Specifically, the PK carryover effect to S reading scores (η^{PK}_S=4.509) was larger than the same carryover effects from K (η^K_S=0.175) and F (η^F_S=2.124) and even from the within-year random effect at S (η^S_S=1.855). The magnitude of the within-year random classroom effect at S was substantially higher for mathematics compared to reading (10.441 vs. 1.855). On the other hand, the within-year random effect of the F classrooms on F reading scores was found to be twice as high as the same effect on mathematics scores (13.929 vs. 6.268). In other words, the same-year classroom effect was higher for reading at first grade for mathematics at second grade.

Results of the likelihood ratio tests from the unconditional analyses are provided in the last three columns of Table 3. Degrees of freedom (df) values for the chi-square tests differed depending on the number of the random effects modeled within each level. Chi-square tests for the three carryover effects at the PK level were all statistically significant for reading[1], however none of those effects were found to be statistically significant for mathematics. These findings indicate that carryover effects from PK classrooms to later grades' student outcomes were prominent only for reading. None of the remaining carryover effects from K and F classrooms were found to be statistically significant according to chi-square tests. Nevertheless, the random K and F classroom effects within the same school year were statistically significant for both reading and mathematics. At S, the random classroom effect on S outcomes was statistically significant only for mathematics. Random effects with larger variances were more significant contributors to the model-fit than the effects with lower variances, especially the carryover effects from K and F.

1 The p-value for η^{PK}_S was 0.01, when rounded to the third decimal, and chi-square statistics was close to other values within the PK level. Thus, we decided to treat this random effect to be statistically significant as well.

10.3389/feduc.2024.1370573

TABLE 3 Variance estimates and likelihood ratio (LR) test results for the unconditional models.

Model level	Random effect	Variance estimate [CI]		LR (χ^2) test results		
		Reading	Math	df	Reading	Math
PK	η_K^{PK}	5.416 [1.295, 11.935]	3.693 [0.331, 8.860]	3	12.17*	6.29
	η_F^{PK}	7.044 [2.989, 13.217]	4.835 [1.144, 10.138]	3	15.8*	8.24
	η_S^{PK}	4.509 [0.994, 9.874]	4.081 [0.576, 9.806]	3	11.26**	7.55
K	η_K^{K+}	16.776 [10.169, 25.386]	17.882 [11.874, 25.629]	3	47.69*	85.53*
	η_F^{K}	0.339 [0.001, 4.423]	1.045 [0.006, 4.720]	3	1.52	4.56
	η_S^{K}	0.175 [0.001, 3.945]	1.966 [0.127, 6.957]	3	0.9	7.21
First	η_F^{F+}	13.929 [8.800, 19.129]	6.268 [2.992, 10.528]	2	36.32*	22.55*
	η_S^{F}	2.124 [0.018, 6.278]	3.451 [0.001, 9.034]	2	4.91	3.7
Second	η_S^{S+}	1.855 [0.001, 5.312]	10.441 [5.871, 16.336]	1	1.72	34.57*

+Random classroom effects at the same grade level as the outcome. Remaining variance estimates belong to carryover effects. *$p < 0.01$. **$p = 0.0104$. CI stands for 95% confidence interval. PK: prekindergarten, K: kindergarten, F: first grade, S: second grade.

TABLE 4 Estimates of quality effects from the conditional models.

Model level	Random effect+	Effect of domain quality		
		Domain	Reading	Math
PK	η_K^{PK}	Es	0.460	–
		Co	1.417*	
		Is	−1.640*	
	η_F^{PK}	Es	−0.148	
		Co	1.310*	
		Is	−0.745	
	η_S^{PK}	Es	−0.571	
		Co	1.453*	
		Is	−0.556	
K	η_K^{K}	Es	−1.036*	−1.041*
		Co	1.606*	1.354*
		Is	−0.785	−0.134
F	η_F^{F}	Es	0.442	0.277
		Co	0.629	0.248
		Is	−0.245	0.021
S	η_S^{S}	Es	-	−0.802
		Co		1.490*
		Is		−0.444

+The effect of classroom quality was examined only for random effects that had a significant LR test result. *Significant based on 95% CI. PK: prekindergarten, K: kindergarten, F: first grade, S: second grade. Es: Emotional Support, Co: Classroom Organization, Is: Instructional Support.

Overall, unconditional model results pointed to rich findings for our research questions. As per first research question, we observed that mathematics scores measured at K, F, and S showed significant variations between the classrooms of these grades. For reading, however, within-year classroom-level variation was significant only at K and F. As per second research question, results showed that reading scores measured at K, F, and S showed significant amount of variation between students' PK classrooms. This confirms that PK classroom effects for reading were carried over all subsequent observed years. Nevertheless, this was true only for reading, and no other carryover effects from K and F classrooms to later years' outcomes were significant. Moreover, the magnitudes of these PK-level variations were even larger than the within-year variation for second grade reading scores.

3.2 Conditional model analysis results

Estimated classroom quality effects on student outcomes as part of the conditional models are summarized in Table 4. Recall that we used the classroom-level quality measures to explain classroom-level variation in student outcomes, which are quantified by within-year and carryover random classroom effects. We examined the effect of classroom quality only for random effects that were determined to be statistically significant by the likelihood ratio tests performed as part of the unconditional model analyses. Thus, we used the information from the unconditional models to identify meaningful classroom-level outcome variations that might be explained by classroom-level quality measures.

For carryover effects, classroom-level variation refers to the variation of student outcomes (e.g., reading scores at S) between students' former grades' classrooms (e.g., PK classrooms). Thus, if outcomes at later grades did not significantly vary based on students' former classrooms, then it was not reasonable to investigate the effect of these classrooms' quality on later years' outcomes. The only carryover effects that were found to be significant were PK classroom effects for reading on all subsequent grades. All other remaining random classroom effects at K, F, and S were within the same school year (see results in Table 3). Thus, we examined the carryover effects of PK classroom quality on later year outcomes.

3.2.1 PK carryover results: effect of PK classroom quality on later year outcomes

The quality of CO at PK was a positive, significant predictor of all subsequent grades' reading scores. This means that higher levels of CO measured in PK classrooms were associated with higher reading scores at all later grades. ES at PK was not a significant predictor of reading scores at any following grades. Contrary to CO, IS at PK was a negative and significant predictor of reading scores at K. Thus, lower PK-level IS scores were found to be related to higher reading scores at K and vice versa.

3.2.2 Within-year results: effect of K, F, and S classroom quality on K, F, and S outcomes

ES in K classrooms was a negative, significant predictor of both K reading and mathematics scores. This implied that at K, low ES quality scores were related to higher reading and mathematics scores, which sounds contradictory. CO, however, was a positive, significant predictor of both reading and mathematics scores in the same year. IS in K was not a significant predictor of K reading or mathematics scores. As per F, none of the quality measures were significant predictors of student reading or mathematics scores measured by the end of F. At S, only CO was a positive and significant predictor of mathematics scores. Thus, it was found that higher CO measures at S were related to higher mathematics scores in S classrooms.

3.2.3 Variance explained by quality scores

Changes in variances of the random effects by the inclusion of the domain-specific quality scores as predictors are summarized in Table 5. These values can be interpreted as the amount of variance (in student outcomes) explained by the classroom quality scores. Also, these results can be used to complement the significance and effect-magnitude results presented in Table 4. An interesting result presented here is that PK classroom quality scores explained more variance in both K and F reading scores than did the same-year measures of quality for both grades. Moreover, the amount of variance explained by PK classroom quality measures showed an increasing trend from K to S. This can be considered as supporting evidence for the lasting effect of PK classroom quality through second grade, in contrast to expectations of fadeout effect. Nevertheless, it should be noted that the direction of the effect was negative for some domain-specific quality scores.

In summary, the conditional model results provided statistical evidence on the predictive power of early years' classroom qualities for later years' student outcomes. As per our third research question, we found that the CO quality of PK classrooms was positively related to students' reading outcomes measured at K, F, and S. Moreover, CO quality at later grades was also the only positive and significant predictor of within-year student outcomes. Also, PK classroom quality scores explained a considerable amount of variance in K, F, and S reading scores. As a result, the quality of PK classrooms in terms of CO seemed to be important for later years' student achievements in reading. Nevertheless, our results also contained some occasional negative effects of quality for the ES and IS domains.

TABLE 5 Amount of random-effect variance explained by classroom quality.

Model level	Random effect	Domain quality scores as predictors	
		Reading	Math
PK	η_K^{PK}	12.3%	-
	η_F^{PK}	18.2%	
	η_S^{PK}	25.4%	
K	η_K^{K}	10.8%	2.4%
F	η_F^{F}	14.5%	6.8%
S	η_S^{S}	-	5.7%

PK: prekindergarten, K: kindergarten, F: first grade, S: second grade.

4 Discussion

This study aimed to examine the effect of PK to S classroom quality on students' K, F, and S reading and mathematics scores. We analyzed a large and population-representative dataset with rigorous models by accounting for the dependency among observations that were cross-classified with multiple classrooms over four grades. Prior to examining the effect of classroom quality on student outcomes, we explored within-year and carryover effects from early years' classrooms.

Carryover classroom effects from PK to later grades were significant only for reading. In other words, we were able to show that students' later-year reading scores showed meaningful magnitudes of variation based on which PK classrooms they were in. Interestingly, however, the carryover classroom effects from K to F and K to S showed a clear pattern of decreasing variance (i.e., fadeout) both for reading and mathematics. Thus, students' F and S scores did not considerably vary based on which K classrooms they were in. Note that these statements do not entail any effect of classroom quality yet and only points to the existence of meaningful variations based on PK classrooms.

In conclusion, our results underlined that the classroom effect of PK does not fadeout, at least for reading, until second grade. This supports the idea that a student's PK classroom continues to impact that student for at least three years thereafter, following similar results found by Vandell et al. (2010). These results also confirm the large body of research advocating for the importance of PK educational opportunities for all students and have important implications for educational policy regarding the importance of early childhood education.

4.1 Quality of classroom organization

After identifying the magnitudes and significance of classroom-level variation in student outcomes, we employed the classroom-level quality measures as potential predictors of such variation. Our expectation was to see that classroom quality scores in three domains are positively related to student outcomes. Based on the domain-specific quality effects, CO was a positive, significant predictor of student achievement across multiple grades and sometimes in both reading and mathematics. The CO domain rates teacher-student interactions on three dimensions: instructional learning formats, productivity, and behavior management. According to the CLASS manual (Teachstone, 2022), high quality CO helps students develop important self-regulatory and executive functioning skills which should help them later in their academic careers.

Our findings specifically highlight an important relationship between quality CO in the PK year and student reading outcomes in later years. This association between high quality CO at PK and later-years' reading outcomes has important practical significance that reiterates previous similar findings from the field. Ponitz et al. (2009) found higher literacy gains for K students in classrooms with higher CO ratings than those in classrooms with lower CO ratings. Likewise, Maier et al. (2012) found that students demonstrated stronger writing skills, phonological and alphabet knowledge, and listening comprehension when they participated in a Head Start program rated highly on CO. Higher CO ratings have also been associated with increased student engagement (Rimm-Kaufman et al., 2009), later academic growth (Curby et al., 2009; Dominguez et al., 2010), and future levels of working memory capacity (Hamre et al., 2014).

One potential reason for finding a significant positive relationship between PK CO and reading outcomes, but not IS or ES, might be the high proportion of students with LEP in our sample. Hindman and Wasik (2013) found that bilingual students in classrooms with higher CLASS scores made greater gains in both Spanish and English language development. Additionally, based on the results of Langeloo et al. (2019), consistency of classroom routines, use of clear examples, and adaptation of lessons (specific practices suggested by CLASS) are all beneficial practices that specifically support the needs of dual-language learners.

Furthermore, our sample was primarily comprised of students from economically disadvantaged backgrounds. Previous research has demonstrated the positive impact of PK quality (particularly elements of quality related to CO) on student outcomes for similar samples of students (Reynolds et al., 2001, 2010; Campbell et al., 2002; Schweinhart et al., 2005; Pungello et al., 2010; Reynolds and Ou, 2011; Duncan and Magnuson, 2013; Ansari, 2018). Schmerse (2020) also found that students' socioeconomic status moderates the relationship between PK quality and persistence, with larger impacts for children from economically disadvantaged households. These results, in conjunction with the results from the present study, provide additional evidence regarding the strong relationship between CO quality in PK and student achievement for this population of students over multiple years.

4.2 Quality of emotional support and instructional support

Interesting results emerged for both the ES and IS quality domains and their relationship with students' reading and mathematics scores. At K, ES was a negative, significant predictor of both K reading and mathematics scores. Similarly, IS at PK was a negative, significant predictor of reading scores measured at K. Potential explanations for these negative relationships can be twofold. First, it is possible that increased IS was provided to students already struggling with pre-literacy skills, and subsequently their reading scores in later grades were lower. Also, as previously stated, our sample was largely LEP students and came from almost exclusively economically disadvantaged backgrounds. This also might have contributed to these findings of negative effects for IS considering that classrooms with mostly LEP students scored higher in IS compared to other classrooms with more balanced or lower LEP student concentration. Second, these negative relationships might have been related to the specific thresholds of the domains to be met in order to observe a meaningful relationship between CLASS scores and student outcomes. These levels of quality within CLASS, or thresholds, define the minimum level at which a positive association between quality and child outcomes is consistently observed (Burchinal et al., 2010). We examined the correlations between K reading and mathematics scores with K ES and IS scores higher than the recommended thresholds (5 for ES and 3.2 for IS). Correlations with trimmed quality scores were small but still negative (varied between -0.102 and -0.016). Despite being a simplistic attempt, we were not able to find evidence neither to support the hypothesis of minimum quality thresholds nor to explain the significant yet negative relationships between student outcomes and ES & IS scores.

5 Limitations and future research

This investigation could have been strengthened by considering other factors related to quality such as school-level conditions, adult-teacher ratios, teacher experience, and classroom environment. Additionally, the conditional models only used domain-specific classroom-level quality scores as predictors and did not include any student-level covariates in the model. Importantly, results from the current study are not evidence for the sustained environments hypothesis, since we did not test the effects of sustained/additive quality from PK to S. Also, our analyses did not include any students who did not attend PK, so there was no examination of differential effects for attending PK or not.

It is also important to note that the data we analyzed do not represent a nationally normative sample of children, though it is representative of one large urban school district, and a sample that is of great interest to researchers and policymakers. Specifically, majority of the students had limited English proficiency and more than 80% of the students were economically disadvantaged. These two sample characteristics, strongly represent the actual body of the public early childhood education students in the region. Nevertheless, replication with other samples of children is necessary before drawing generalized conclusions from these findings related to impacts of quality early childhood education.

The findings of this study point to several opportunities for further research. First, further investigation into cumulative effects of the early grades' quality is warranted, as this study found that grade-specific associations between quality and outcomes were inconsistent. Second, the significant but negative effects of instructional and emotional support qualities on student outcomes at specific grades lead to questions about how these domains of quality are operationalized and measured by CLASS. Further, both the consistent positive effect of CO as a domain measure, combined with the counterintuitive finding that ES and IS had some significant negative within-year and carryover effects, call for further investigation.

Relevant to the operationalization of the CLASS tool, one potential reason for observing inconsistent impacts of domains might be the large correlations between the domain scores. This study relied on the three-factor structure recommended by Hamre et al. (2013) and used the domain scores as predictors in a linear model. In future studies, the same cross-classified models can be fitted by assuming different factor structures as reported in Li et al. (2020) including a single or a bi-factor measurement model for the CLASS scores. For a more precise model, mentioned factor models can be integrated to the cross-classified model rather than using the composite scores as the measures of the outcomes.

6 Conclusion

The results of this study add to the parallel bodies of literature on both the carryover or fadeout of contextual PK effects and the impact of early-year classrooms' quality on student outcomes across years. The novel methods applied in this study allowed for a cross-classified hierarchical structure that most accurately represents the complexity of longitudinal education research while also accounting for the carryover of random classroom effects. Ultimately, this study furthers the arguments that quality in early childhood is a complex construct that requires ongoing examination and refinement.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: data analyzed in this study cannot be shared with third parties as per the data sharing agreement between the school district and the research institution. Requests to access these datasets should be directed to ykara@smu.edu.

Ethics statement

The studies involving humans were approved by Southern Methodist University IRB and Dallas ISD RRB. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because this study used deidentified secondary data.

Author contributions

YK: Methodology, Visualization, Writing – original draft, Writing – review & editing. DF: Conceptualization, Data curation, Resources, Writing – original draft, Writing – review & editing. AW: Conceptualization, Resources, Writing – original draft, Writing – review & editing. AK: Methodology, Software, Writing – review & editing. JH: Writing – original draft. SB: Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The research reported here was supported, in whole or in part, by the Institute of Education Sciences, U.S. Department of Education, through grant R305H180015 to Southern Methodist University. The opinions expressed are those of the authors and do not represent the views of the Institute of the U.S. Department of Education.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ansari, A. (2018). The persistence of preschool effects from early childhood through adolescence. *J. Educ. Psychol.* 110, 952–973. doi: 10.1037/edu0000255

Ansari, A., and Pianta, R. C. (2018). Variation in the long-term benefits of child care: the role of classroom quality in elementary school. *Dev. Psychol.* 54, 1854–1867. doi: 10.1037/dev0000513

Bailey, D., Duncan, G. J., Odgers, C. L., and Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *J. Res. Educ. Effect.* 10, 7–39. doi: 10.1080/19345747.2016.1232459

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *Futur. Child.* 5, 25–50. doi: 10.2307/1602366

Bassok, D., Gibbs, C. R., and Latham, S. (2019). Preschool and children's outcomes in elementary school: have patterns changed nationwide between 1998 and 2010? *Child Dev.* 90, 1875–1897. doi: 10.1111/cdev.13067

Belfield, C. R., Nores, M., Barnett, S., and Schweinhart, L. (2006). The high/scope Perry preschool program: cost-benefit analysis using data from the age-40 follow-up. *J. Hum. Resour.* XLI, 162–190. doi: 10.3368/jhr.XLI.1.162

Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction. *Appl. Dev. Sci.* 12, 140–153. doi: 10.1080/10888690802199418

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

Campbell, F. A., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., et al. (2014). Early childhood investments substantially boost adult health. *Science* 343, 1478–1485. doi: 10.1126/science.1248429

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., and Miller-Johnson, S. (2002). Early childhood education: young adult outcomes from the abecedarian project. *Appl. Dev. Sci.* 6, 42–57. doi: 10.1207/S1532480XADS0601_05

Cash, A. H., Ansari, A., Grimm, K. J., and Pianta, R. C. (2019). Power of two: the impact of 2 years of high quality teacher child interactions. *Early Educ. Dev.* 30, 60–81. doi: 10.1080/10409289.2018.1535153

Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal, M., et al. (2009). The relations of observed pre-k classroom quality profiles to children's achievement and social competence. *Early Educ. Dev.* 20, 346–372. doi: 10.1080/10409280802581284

Data Recognition Corporation (2024a). Available at: SUPERA brochure. Available at: https://terranova3.com/PDFs/SUPERA_Brochure.pdf

Data Recognition Corporation (2024b). TerraNova technical quality brochure. Available at: https://terranova3.com/PDFs/TerraNova_Technical_Quality_Brochure.pdf

Dominguez, X., Vitiello, V. E., Maier, M. F., and Greenfield, D. B. (2010). A longitudinal examination of young children's learning behavior: child-level and classroom-level predictors of change throughout the preschool year. *Sch. Psychol. Rev.* 39, 29–47. doi: 10.1080/02796015.2010.12087788

Duncan, G. J., and Magnuson, K. (2013). Investing in preschool programs. *J. Econ. Perspect.* 27, 109–132. doi: 10.1257/jep.27.2.109

Durkin, K., Lipsey, M. W., Farran, D. C., and Wiesen, S. E. (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Dev. Psychol.* 58, 470–484. doi: 10.1037/dev0001301

Engel, M., Claessens, A., and Finch, M. A. (2013). Teaching students what they already know? The (mis) alignment between mathematics instructional content and student knowledge in kindergarten. *Educ. Eval. Policy Anal.* 35, 157–178. doi: 10.3102/0162373712461850

Fuller, B., Bein, E., Bridges, M., Kim, Y., and Rabe-Hesketh, S. (2017). Do academic preschools yield stronger benefits? Cognitive emphasis, dosage, and early learning. *J. Appl. Dev. Psychol.* 52, 1–11. doi: 10.1016/j.appdev.2017.05.001

Gormley, W. T., Gayer, T., Phillips, D., and Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Dev. Psychol.* 41, 872–884. doi: 10.1037/0012-1649.41.6.872

Hamre, B., Hatfield, B., Pianta, R., and Jamil, F. (2014). Evidence for general and domain-specific elements of teacher–child interactions: associations with preschool Children's development. *Child Dev.* 85, 1257–1274. doi: 10.1111/cdev.12184

Hamre, B. K., and Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Dev.* 76, 949–967. doi: 10.1111/j.1467-8624.2005.00889.x

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., et al. (2013). Teaching through interactions: testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *Elem. Sch. J.* 113, 461–487. doi: 10.1086/669616

Hill, C. J., Gormley, W. T., and Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Child. Res. Q.* 32, 60–79. doi: 10.1016/j.ecresq.2014.12.005

Hindman, A. H., and Wasik, B. A. (2013). Vocabulary learning in head start: nature and extent of classroom instruction and its contributions to children's learning. *J. Sch. Psychol.* 51, 387–405. doi: 10.1016/j.jsp.2013.01.001

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, R., Clifford, R., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Child. Res. Q.* 23, 27–50. doi: 10.1016/j.ecresq.2007.05.002

Kwok, O., Lai, M. H., Tong, F., Lara-Alecio, R., Irby, B., Yoon, M., et al. (2018). Analyzing complex longitudinal data in educational research: a demonstration with project English language and literacy acquisition (ELLA) data using xxM. *Front. Psychol.* 9, 1–12. doi: 10.3389/fpsyg.2018.00790

Langeloo, A., Mascareño Lara, M., Deunk, M. I., Klitzing, N., and Strijbos, J.-W. (2019). A systematic review of teacher–child interactions with multilingual young children. *Rev. Educ. Res.* 89, 536–568. doi: 10.3102/0034654319855619

LaParo, K. M., Pianta, R. C., and Stuhlman, M. (2004). The classroom assessment scoring system: findings from the prekindergarten year. *Elem. Sch. J.* 104, 409–426. doi: 10.1086/499760

Lee, V. E., and Loeb, S. (1995). Where do head start attendees end up? One reason why preschool effects fade out. *Educ. Eval. Policy Anal.* 17, 62–82. doi: 10.3102/01623737017001062

Li, H., Liu, J., and Hunter, C. V. (2020). A Meta-analysis of the factor structure of the classroom assessment scoring system (CLASS). *J. Experi. Edu.* 88, 265–287. doi: 10.1080/00220973.2018.1551184

Lipsey, M. W., Farran, D. C., and Durkin, K. (2018). Effects of the Tennessee prekindergarten program on children's achievement and behavior through third grade. *Early Child. Res. Q.* 45, 155–176. doi: 10.1016/j.ecresq.2018.03.005

Loeb, S., Bridges, M., Bassok, D., Fuller, B., and Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Econ. Educ. Rev.* 26, 52–66. doi: 10.1016/j.econedurev.2005.11.005

Magnuson, K. A., Ruhm, C., and Waldfogel, J. (2007). The persistence of preschool effects: do subsequent classroom experiences matter? *Early Child. Res. Q.* 22, 18–38. doi: 10.1016/j.ecresq.2006.10.002

Maier, M. F., Vitiello, V. E., and Greenfield, D. B. (2012). A multilevel model of child- and classroom-level psychosocial factors that support language and literacy resilience of children in head start. *Early Child. Res. Q.* 27, 104–114. doi: 10.1016/j.ecresq.2011.06.002

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., et al. (2008). Measures of classroom quality in pre-kindergarten and children's development of academic, language, and social skills. *Child Dev.* 79, 732–749. doi: 10.1111/j.1467-8624.2008.01154.x

Mehta, P. (2013). xxM: N-level structural equation modeling for dependent data (R package version 0.6.0). Available at: https://xxm.times.uh.edu/

NICHD Early Child Care Research Network (2002). Child-care structure process outcome: direct and indirection effects of child-care quality on young children's development. *Psychol. Sci.* 13, 199–206. doi: 10.1111/1467-9280.00438

Pearman, F. A., Springer, M. P., Lipsey, M., Lachowicz, M., Swain, W., and Farran, D. (2019). Teachers, schools, and pre-k effect persistence: an examination of the sustaining environments hypothesis. *J. Res. Educ. Effect.* 13, 547–573. doi: 10.1080/19345747.2020.1749740

Petscher, Y., and Schatschneider, C. (2019). Using n-level structural equation models for causal modeling in fully nested, partially nested, and cross-classified randomized controlled trials. *Educ. Psychol. Meas.* 79, 1075–1102. doi: 10.1177/0013164419840071

Phillipsen, L. C., Burchinal, M. R., Howes, C., and Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Child. Res. Q.* 12, 281–303. doi: 10.1016/S0885-2006(97)90004-1

Pianta, R. C., Howes, C., Burchinal, M., Byrant, D., Clifford, R., Early, D., et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: do they predict observed classroom quality and child-teacher interactions? *Appl. Dev. Sci.* 9, 144–159. doi: 10.1207/s1532480xads0903_2

Pianta, R.C., LaParo, K., and Hamre, B. (2008). *The classroom assessment scoring system (CLASS), Charlottesville*, Teachstone Training, VA.

Pianta, R. C., and Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *Sch. Psychol. Rev.* 33, 444–458. doi: 10.1080/02796015.2004.12086261

Ponitz, C. C., McClelland, M. M., Matthews, J. S., and Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Dev. Psychol.* 45, 605–619. doi: 10.1037/a0015365

Puma, M., Bell, S., Cook, R., and Heid, C. (2010). Head start impact study final report. Washington, DC. Available at: http://files.eric.ed.gov/fulltext/ED507845.pdf

Pungello, E. P., Kainz, K., Burchinal, M., Wasik, B. H., Sparling, J. J., Ramey, C. T., et al. (2010). Early educational intervention, early cumulative risk, and the early home environment as predictors of young adult outcomes within a high-risk sample. *Child Dev.* 81, 410–426. doi: 10.1111/j.1467-8624.2009.01403.x

R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Reynolds, A. J., and Ou, S. (2011). Paths of effects from preschool to adult well-being: a confirmatory analysis of the child-parent center program. *Child Dev.* 82, 555–582. doi: 10.1111/j.1467-8624.2010.01562.x

Reynolds, A. J., Temple, J. A., and Ou, S. (2010). Preschool education, educational attainment, and crime prevention: contributions of cognitive and non-cognitive skills. *Child Youth Serv. Rev.* 32, 1054–1063. doi: 10.1016/j.childyouth.2009.10.019

Reynolds, A. J., Temple, J. A., Robertson, D. L., and Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest: a 15-year follow-up of low-income children in public schools. *JAMA* 285, 2339–2346. doi: 10.1001/jama.285.18.2339

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K. J., Nathanson, L., and Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Dev. Psychol.* 45, 958–972. doi: 10.1037/a0015861

Schmerse, D. (2020). Preschool quality effects on learning behavior and later achievement in Germany: moderation by socioeconomic status. *Child Dev.* 91, 2237–2254. doi: 10.1111/cdev.13357

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., and Nores, M. (2005). Lifetime effects: The HighScope Perry preschool study through age 40. HighScope Press. Available at: http://highscope.org/file/Research/PerryProject/specialsummary_rev2011_02_2.pdf

Schweinhart, L. J., and Weikart, D. P. (1993). Success by empowerment: the high/scope Perry preschool study through age 27. *Young Child.* 49, 54–58. Available at: https://www.jstor.org/stable/42725546

Teachstone (2022). *Pre-k-3rd reference manual: Classroom assessment scoring system. 2nd* Edn. Charlottesville, Virginia: Teachstone, Inc.

Texas Education Agency. (2024). Texas public prekindergarten programs and enrollment ages 3 and 4. Available at: https://www.texaseducationinfo.org/PickList_Vbox.aspx?Page=New+Reports&ReportName=tpeir_pk_enroll_funds_dist&PickList=District&SubList=School+Year&Title=Texas+Public+Prekindergarten+Programs+-+District+and+School&Graph=N&from=Home%2fIndex

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., and Vandergrift, N.NICHD Early Child Care Research Network (2010). Do effects of early child care extend to age 15 years? Results from the NICHD study of early child care and youth development. *Child Dev.* 81, 737–756. doi: 10.1111/j.1467-8624.2010.01431.x

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

Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Tarullo, L., and Burchinal, M. (2010). Quality dosage, thresholds, and features in early childhood settings: A review of the literature (OPRE 2011–5). Available at: http://www.acf.hhs.gov/programs/opre/cc/q_dot/quality_review.pdf

Zaslow, M., Martinez-Beck, I., Tout, K., and Halle, T. (2011) in *Quality measurement in early childhood settings*. ed. H. Paul (Washington, D.C.: Brookes Publishing Co).