

Teachers, Schools, and Pre-K Effect Persistence: An Examination of the Sustaining Environment Hypothesis

Francis A. Pearman II, Matthew Springer, Mark Lipsey, Mark Lachowicz,
Walker Swain & Dale Farran

To cite this article: Francis A. Pearman II, Matthew Springer, Mark Lipsey, Mark Lachowicz, Walker Swain & Dale Farran (2020): Teachers, Schools, and Pre-K Effect Persistence: An Examination of the Sustaining Environment Hypothesis, Journal of Research on Educational Effectiveness, DOI: [10.1080/19345747.2020.1749740](https://doi.org/10.1080/19345747.2020.1749740)

To link to this article: <https://doi.org/10.1080/19345747.2020.1749740>



Published online: 28 May 2020.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Teachers, Schools, and Pre-K Effect Persistence: An Examination of the Sustaining Environment Hypothesis

Francis A. Pearman II^a, Matthew Springer^b, Mark Lipsey^c, Mark Lachowicz^c, Walker Swain^d and Dale Farran^c

^aStanford Graduate School of Education, Stanford University, Stanford, California, USA; ^bSchool of Education, University of North Carolina, Chapel Hill, North Carolina, USA; ^cPeabody College, Vanderbilt University, Nashville, Tennessee, USA; ^dCollege of Education, University of Georgia, Athens, Georgia, USA

ABSTRACT

The sustaining environments thesis hypothesizes that PreK effects are more likely to persist into later grades if children experience high-quality learning environments in the years subsequent to PreK. This study tests this hypothesis using data from a statewide PreK randomized experiment in Tennessee that found positive effects at the end of PreK that did not persist past kindergarten. These data were combined with teacher observation and school-level value-added scores from Tennessee's formal evaluation system to determine whether positive effects of PreK persisted for the subgroup of students exposed to higher-quality learning environments between kindergarten and 3rd grade. Neither exposure to highly effective teachers nor attending a high-quality school was sufficient by itself to explain differences in achievement between PreK participants and non-participants in 3rd-grade. However, this study found evidence that having both was associated with a sustained advantage for PreK participants in both math and ELA. Notably, however, very few children were exposed to high-quality learning environments after PreK, suggesting that maximizing PreK investments may require attending to the quality of learning environments during PreK and beyond.

ARTICLE HISTORY

Received 15 June 2019
Revised 30 January 2020
Accepted 5 March 2020

KEYWORDS

PreK; academic achievement; sustaining environments; teacher quality; school quality

Introduction

Short-term boosts in children's language, literacy, and math skills that result from attending prekindergarten (PreK) often diminish soon after preschool ends (Yoshikawa et al., 2016). This pattern, commonly known as *fadeout*, has been noted in preschool effectiveness literature dating back to the 1960s, and has since been documented in high fidelity studies of preschool effectiveness at the district, state, and national levels (Hill et al., 2015; Lipsey et al., 2018; Puma et al., 2010, 2012). For example, a recent meta-analysis of existing preschool effectiveness research based on over 60 evaluations of early childhood interventions published between 1960 and 2007 found that the average end-of-program-year impact of preschool on cognitive skills dropped by more than 50% in

CONTACT Francis A. Pearman II  apearman@stanford.edu  Stanford Graduate School of Education, Stanford University, CERAS 524, 485 Lasuen Mall, Stanford, CA 94305, USA.

© 2020 Taylor & Francis Group, LLC

the year following the intervention, and again by 50% 1–2 years later (Bailey et al., 2017).

In light of these findings, conceptual and empirical research has attempted to shed light on the context and processes through which benefits of early childhood education investments might potentially be maintained over time. Moreover, an emerging line of research has begun exploring the role subsequent learning environments can play in maintaining PreK effects beyond kindergarten entry—that is, the role of what many scholars have come to call *sustaining environments* (Bailey et al., 2017). For instance, prior research has indicated that both teacher quality and school quality may play a role in determining the extent of PreK effect persistence by providing learning environments that can adequately built off learning gains experienced during PreK (Currie & Thomas, 2000; Lee & Loeb, 1995; Swain et al., 2015), but as described in the next section, the evidence is mixed and features a variety of measures of school and teacher quality and approaches to estimating PreK effects.

The current study builds on the emerging literature on sustaining environments by combining data from a recently conducted randomized controlled trial of a statewide PreK program, the Tennessee Voluntary PreK (TN-VPK) program, with detailed teacher and school information to examine whether the persistence of PreK effects is influenced by subsequent exposure to two of the more often-studied contextual determinants of student achievement: teacher and school quality. The underlying argument of this study is that a single measure of teacher *or* school quality may miss important variation *within* each condition relevant for the persistence or fadeout of PreK effects. For instance, though economically disadvantaged children who attend public PreK programs like TN-VPK often attend lower quality elementary schools (Currie & Thomas, 2000), there is evidence that some children may nevertheless have exposure to highly effective teachers in such schools (Sass et al., 2012). Likewise, even if children are able to attend higher quality schools after PreK, children may not necessarily have good teachers in these schools, which is especially the case with respect to economically disadvantaged students who have increased odds of being assigned a special education designation and are more likely to be tracked into less rigorous classes (Blair & Scott, 2002; Kalogrides & Loeb, 2013). Therefore, it is relevant to consider the independent effects of high quality schools and teachers separately as well as the combined effects of exposure to both. More specifically, we ask:

- Is the association between PreK participation and 3rd grade achievement conditional on the number of teachers rated as highly effective that children have between PreK and 3rd grade or the timing of their exposure to such teachers?
- Is the association between PreK participation and 3rd grade achievement conditional on the quality of schools that children attend between PreK and 3rd grade?
- Is the association between PreK participation and 3rd grade achievement stronger with exposure to both higher quality schools and highly effective teachers?

Examining these questions about PreK persistence and fadeout in the context of the TN-VPK experiment provides an ideal opportunity given that Lipsey et al. (2018) found

large, positive effects of TN-VPK on student achievement at the end of the PreK year that disappeared, or in some instances, turned negative by the time students completed the third grade. In addition, Tennessee has robust teacher evaluation and school accountability systems. We are therefore able to draw on classroom observation scores of teaching and school-level growth data over time to measure the quality of the subsequent learning environments that children encounter.

Conceptual Framework and Prior Research

A near universal finding across recent evaluations of PreK programs is one of positive effects on achievement at the end of PreK that fade relatively quickly over time (Hill et al., 2015; Lipsey et al., 2018; Puma et al., 2010, 2012). Whereas some early childhood research explains PreK fadeout in terms of the nature and extent of the learning that takes place in preschool classrooms, the *sustaining environments* perspective emphasizes subsequent learning experiences after PreK as critical determinants of whether early learning advantages brought about by PreK persist as children progress through formal schooling (Bailey et al., 2017). The sustaining environments perspective holds that for early childhood interventions to be deemed successful subsequent learning environments must, at the very least, maintain the learning advantages brought about by attending preschool. Said otherwise, for initial gains from attending PreK to be persistently apparent, PreK children must go on to attend elementary schools where they are able to continue to learn at the same or higher rates as children who did not attend PreK. To the extent that subsequent learning environments are of lower quality or that subsequent teachers focus their efforts disproportionately on the learning needs of struggling children who did not attend PreK, convergence or fadeout of PreK effects becomes more likely.

The growing body of quasi-experimental literature examining the sustaining environments thesis has arrived at mixed conclusions. On the one hand, several studies have found evidence in favor of the view that subsequent learning environments matter for PreK effect persistence. For instance, Lee and Loeb (1995) and Currie and Thomas (2000), both studying fadeout in Head Start participants, indirectly referenced the sustaining environments thesis by noting that former Head Start students went on to attend schools of significantly lower quality than their peers who did not attend Head Start. Lee and Loeb concluded: “No matter how beneficial the Head Start experience was initially for its participants, such benefits are likely to be undermined if these students are thereafter exposed to lower quality schooling” (p. 3).

Direct evidence in favor of the sustaining environments perspective is found in several recent studies. In Tennessee, Swain et al. (2015) reported on the role teacher effectiveness plays in determining the extent to which PreK effects persist into kindergarten. Results indicated a modest, positive interaction between teacher quality and PreK exposure on cognitive measures such that higher teacher quality in kindergarten and first grade was associated with sustained advantage for PreK participants during these years. Moreover, the relations between teacher quality and PreK participation appeared to be particularly important for students who showed early cognitive deficits and language

barriers prior to PreK enrollment. Likewise, Ansari and Pianta (2018) using data from the National Institute of Child Health and Human Development Study of Early Child Care found that the benefits of high quality childcare on math and literacy persisted until age 15 for those children who went on to experience high quality classroom environments in elementary grades. In contrast, they found no evidence that the effects of high quality childcare persisted for children who subsequently attended lower quality classrooms in elementary grades.

These findings are somewhat echoed by Jenkins and colleagues (2018) who found that targeted professional supports for elementary grade teachers designed to promote continuity and avoid repetition between grade levels moderated fadeout from preschool. In particular, this study found that the persistence of PreK effects was linked to whether there were coordinated alignments in curriculum between preschool and later grades. This finding is particularly insightful in light of several recent studies demonstrating that kindergarten teachers often teach material related to the knowledge children already possess at kindergarten entry (Bassok et al., 2016; Claessens et al., 2014; Engel et al., 2013; Gervasoni & Perry, 2015; Magnuson et al., 2007). This repetition means that coordinated efforts and professional supports to build upon children's prior knowledge may be essential to ensure PreK effects persist into and beyond kindergarten.

In contrast, other research has found little to no evidence that subsequent learning environments matter for the persistence of PreK effects, and this contrary evidence has been reported at both the school and classroom levels. For instance, in the same study noted above, Jenkins et al. (2018) found no consistent evidence that school-level characteristics such as poverty or proficiency rates moderated the persistence of PreK effects. Likewise, Claessens et al. (2014) used classroom-level data from the Early Childhood Longitudinal Studies-Kindergarten Class of 2011–2012 (ECLS-K) and found no evidence that either the level of instruction or type of instruction in kindergarten classrooms moderated the persistence of PreK effects. Engel et al. (2013), also using data from the ECLS-K, found that this pattern of rapid fadeout was in part attributable to kindergarten teachers repeating instruction covered during preschool. However, using the same dataset, Bassok et al. (2015) found no meaningful differences in the rate of fadeout based on a range of kindergarten features, including class size, co-location with pre-k classroom, peer attendance, transition practices, and time devoted to reading instruction.

In sum, prior research indicates that a common pattern of PreK effects on achievement-related outcomes is one of short-term beneficial impacts that fade over time, often quickly; and the literature is largely unsettled regarding the systematic components responsible for PreK effect persistence. Our work builds from this prior literature by leveraging a recently conducted state-wide PreK experiment in Tennessee, the Tennessee Voluntary Prekindergarten Program (TN-VPK), and examining the extent to which the effects of PreK on 3rd grade achievement are moderated by the quality of the teachers and schools that children subsequently experience after PreK. Measuring academic performance in 3rd grade is ideal for the current study because prior TN-VPK research found that positive PreK effects at kindergarten entry faded and in some instances reversed by 3rd grade (Lipsey et al., 2018).

Data and Sample

The data in this study come from two primary sources. First, student information and related data were collected by researchers at the Peabody Research Institute (PRI) as part of the TN-VPK study. Second, teacher evaluation and school performance records along with supplemental student, teacher, and school information were collected by the Tennessee Department of Education (TNDOE) and processed for research purposes by the Tennessee Education Research Alliance (TERA), which houses student test score data linked with specific teachers and schools throughout our study period.

Analytic Sample

This study is situated in a larger evaluation of the TN-VPK program, conducted by Vanderbilt University's PRI in collaboration with TNDOE. This larger study evaluated the impact of TN-VPK across two cohorts of children that entered PreK during the 2009–2010 and 2010–2011 school years respectively. At the start of each school year, a group of oversubscribed PreK centers agreed to randomly assign applicants either to a treatment condition that was offered admission or to a control condition that was denied admission. These efforts yielded a total of 2,990 students in 79 schools in 29 school districts across Tennessee (see Lipsey et al., 2018 for details). In this study, we focus on the second cohort of children who entered PreK in 2010–2011. The second cohort permits access to measures of teacher effectiveness for students in each school year because Tennessee's teacher evaluation system was introduced during this cohort's kindergarten year (2011–2012 school year). Thus, we lack kindergarten teacher evaluation records for students in the first cohort recruited in 2009 who were already in 1st grade when the teacher evaluation policy was introduced.

Of the 1,240 children in the analytic sample, 434 were missing data on one or more variables. Two percent of students were missing school level value added scores, 25% of students were missing a valid teacher observation score in at least one school year between kindergarten and 3rd grade, and 18% of students had missing test scores in third grade.¹ Although there are numerous potential approaches for handling these missing data, the approach adopted in this study is to report results from a complete case analysis in the primary results section based on the 806 students who did not have any missing data followed by a robustness check that multiply imputes missing values on covariates (see [Tables 1 and 2](#)).

Of the remaining 806 students, 507 had been randomly assigned to the TN-VPK treatment group and 299 had been randomly assigned to the control condition. However, there existed noncompliance in treatment assignment. In particular, of the 507 students assigned to VPK, 16% did not attend VPK and 84% complied. Of the 299 control students, 75% did not attend VPK and 25% crossed over, i.e., attended PreK for at least 20 days at a TN-VPK site even though they were not randomly assigned into the treatment condition. Given that the primary focus of this study is the relationship

¹Table C.1 in the [Appendix](#) provides balance tests of baseline characteristics across students with and without valid teacher observation scores. This table indicates black students were disproportionately likely to have missing teacher observation scores.

Table 1. Balance tests on student, teacher, and schools characteristics.

	All TN students (1)	Control group mean (2)	Exp. versus control (unadjusted) (3)	Exp. versus control (adjusted) (4)
Student characteristics				
Age (in months)	...	53.34	0.09 (0.24)	0.02 (0.24)
Female	0.49	0.51	0.04 (0.03)	0.05 (0.04)
White	0.66	0.45	0.09* (0.04)	−0.03 (0.04)
Black	0.22	0.23	0.04 (0.03)	0.04 (0.03)
Hispanic	0.10	0.31	−0.15*** (0.04)	−0.02 (0.04)
English primary language	0.89	0.68	0.14*** (0.04)	−0.00 (0.04)
Teacher characteristics				
% 0 HE teachers	0.03	0.22	0.02 (0.04)	0.02 (0.04)
% 1 HE teachers	0.16	0.33	−0.04 (0.04)	−0.03 (0.04)
% 2 HE teachers	0.30	0.27	−0.02 (0.04)	−0.02 (0.04)
% 3 or 4 HE teachers	0.51	0.17	0.04 (0.04)	0.03 (0.04)
School characteristics				
Value-added	1.40	0.98	0.33 (0.20)	0.23 (0.20)
<i>n</i> =	52,817	315		491

Note: This table presents balance tests of equivalency for baseline characteristics. “HE teachers” refers to “Highly Effective” teachers. Column (1) reports means for all Tennessee elementary students. Column (2) reports means for PreK non-participants. Column (3) reports unadjusted differences between PreK participants and PreK non-participants, which are estimated using OLS regressions of each characteristic on a binary indicator for Pre-K participation. Column (4) reports adjusted differences based on OLS regressions that include propensity score weights. Age in months was unavailable for all elementary children in Tennessee. Standard errors are reported in parenthesis and are clustered at the school level. * $p < .05$, *** $p < .001$ for two-tailed tests of significance.

between PreK participation and subsequent teacher effectiveness and school quality within the context of the sustaining environment hypothesis, we use non-participants irrespective of their experimental assignment as the counterfactual group. In other words, we are interested in the conditions under which the effects of PreK *participation* persist over time.² (We report results in [Tables 1](#) and [2](#) that use an intent-to-treat indicator of assignment.) Finally, we treated the 7 children who attended VPK for less than 20 days as non-participants, corresponding to the standard used by the VPK program for what constitutes VPK enrollment. That is, our analysis compares those who attended a VPK program for a minimum of 20 days and those who did not attend or attended fewer than 20 days of a VPK program.³ We refer

²One option would be to use the intent-to-treat indicator and interactions between the intent-to-treat indicator and subsequent learning environments as instruments for PreK participation and the interaction between PreK participation and subsequent learning environments. However, prior research has shown that instrumental variable methods produce more bias than OLS estimates when using limited sample sizes (Boef et al., 2014), which is exacerbated in the current study because of the inclusion of three-way interactions. Indeed, a two-staged least squares approach produced unstable and implausible estimates that are not reported.

³It should also be noted that some children were retained and had not reached 3rd grade by the time achievement data were collected. However, as indicated by [Table A.1](#) in the [Appendix](#), there was no statistically significant difference in retention across VPK participants and non-participants. Therefore, results based on children who were not retained should not be biased by any retention differences between PreK participants and non-participants.

Table 2. Joint distribution of teacher and school quality.

	<i>n</i> Row Cell	School quality			Total
		Low	Moderate	High	
Number of highly effective teachers		46	123	19	
	0	0.24 0.06 78	0.65 0.15 149	0.10 0.02 21	188
	1	0.31 0.10 45	0.60 0.18 134	0.08 0.03 32	248
	2	0.21 0.06 30	0.64 0.17 85	0.15 0.04 44	211
	3 or 4	0.19 0.04	0.53 0.11	0.28 0.05	159
	Total	199	491	116	806

Note: Low quality schools are defined as those with gain scores that did not meet state growth standards. High quality schools are defined as those with gain scores 3 points above the mean gain score in the analytic sample, which corresponds to roughly 1 standard deviation above average. Moderate quality schools fall between these two thresholds. Teacher quality was based on teacher observation scores from Tennessee's statewide educator evaluation system. Observation scores ranged between 0 and 5. Highly effective teachers were defined as those with observation scores of 4 or above.

to these two groups of students as VPK participants or non-participants. (Results are robust to the inclusion into the treatment group of those children who attended VPK for less than 20 days.) The final analytic sample included 491 VPK participants and 315 non-participants who attended 218 elementary schools in Tennessee.

Primary Measures of Interest

Achievement

The outcome variables of interest are generated from the Tennessee Comprehensive Assessment Program (TCAP), a series of standardized assessments administered to students in grades 3 through 8. We use the scale scores from the 3rd grade mathematics and English Language Arts (ELA) examinations.⁴ Values in the analytic sample range between 628 and 900 with a mean of 760.8 in mathematics and between 600 and 868 with a mean of 748.5 in ELA. To facilitate interpretation, we standardized scores by subject to have a mean of 0 and standard deviation of one across the entire sample.

Teacher Quality

Our teacher effectiveness measure is calculated using data collected as part of the statewide educator evaluation system. In Tennessee, annual evaluations differentiate teacher performance based on a composite teacher effectiveness rating score that uses individual and school-level student growth scores and achievement data as well as classroom observations of teachers.⁵ Because students in grades K through 2 do not take standardized

⁴We do not have access to student test scores in 4th grade due to failed implementation of the state's online testing program in that school year.

⁵As of July 2011, the Tennessee State Board of Education approved four teacher evaluation models—the Tennessee Educator Acceleration Model (TEAM); Project Coach; Teacher Effectiveness Measure (TEM); and Teacher Instructional Growth for Effectiveness and Results (TIGER). Although implementation is quite different from one model to the next, the evaluation models all follow the requirements set forth by Tennessee's Teacher Effectiveness Advisory Committee

assessments that contribute to a teacher's overall performance evaluation rating, our measure focuses exclusively on the classroom observation component of the evaluation system, which is an adaptation of the Charlotte-Danielson rubric (Danielson, 2013) and assesses teachers multiple times per year in the areas of instruction, planning, and environment.

To create the teacher effectiveness measure, we calculated a teacher's average observation rating by year. While most elementary students had only one primary teacher per school year, in instances where more than one teacher and rating existed, we averaged across those teachers based on the number of school days in which a student was enrolled in each teacher's classroom. These scores ranged between 0 and 5, where values equal to or less than 1 denote that a teacher performed significantly below expectation and 4 or greater means the teacher performed significantly above expectation. We define teachers with observation scores of 4 or above as *highly effective*.

We use a teacher's average observation score to create variables that capture overall exposure and timing of exposure to highly effective teachers. Overall exposure is calculated as the number of times that a student was assigned to a highly effective teacher from kindergarten to 3rd grade. These values range from zero to four, where zero means a student was never enrolled in a classroom taught by a teacher rated as highly effective and 4 denotes a student was taught by a highly effective teacher in every year from K to 3rd grade. However, because fewer than 5% of the analytic sample was taught by a highly effective teacher in every year from K to 3rd grade, we lump together students that have 3 or 4 years of exposure. For the timing of exposure, we first created a variable for whether a student was assigned to a highly effective teacher in at least the last two school years (i.e., 2nd and 3rd grades) irrespective of prior exposure. Here, we are trying to capture the possibility that having a highly effective teacher in grades closest to when a student takes the math and ELA assessments will lessen the chances of fadeout. We also create a variable that denotes whether a student had a highly effective teacher in the first 2 years after PreK (i.e., kindergarten and 1st grade) irrespective of later exposure. This variable was used to represent the possibility that the knowledge- and skill-acquisition most needed to prevent fadeout may occur in years immediately following PreK participation.

A potential concern of our observed measure of teacher quality is whether teachers with higher observed quality measures actually produce better student outcomes or, more simply, have higher value-added scores. However, Doan (2019) found classroom observation scores in Tennessee not only capture teacher impacts on students' K-12, post-secondary, and labor market outcomes but also that the effects of observation scores alone are at least comparable to, if not larger than, effects of teacher value-added scores on various student outcomes. Using longitudinal administrative K-12, post-secondary, and labor market data from 2006–2007 to 2017–2018, Doan reported that a one standard deviation increase in observation scores is expected to result in a 0.089 standard deviation increase in test scores, a 0.145 decrease in student absence, and a 0.019 decrease in student suspensions. The findings are similarly positive for high school graduation, post-secondary enrollment, and

and adopted by the State Board of Education, and have the same goals—to monitor teacher performance and encourage teacher development. More than 80% of teachers across Tennessee used TEAM as their evaluation model, while TEM is the second most frequently used (11%), followed by Project COACH (5%) and TIGER (2%). In our analytic sample, only a small number of teachers were evaluated under models other than TEAM.

degree completion. In short, observation scores capture important teacher impacts on student success and provide a useful measure for examining teacher quality in the context of the sustaining environments hypothesis.

School Quality

For school quality, our interest is assessing the extent to which children's broader schooling environment, beyond the classroom in which they are a part, might facilitate learning and achievement. Our measure of school quality is a school-level value added score as calculated by the Tennessee Value-Added Assessment System (SAS 2017). This measure captures the average relative progress that schools make on state assessments compared to the state's growth standard, which represents the minimum amount of progress a school's student population is expected to make each year. According to this measure, school quality is indexed by the extent to which student performance in a given school is better than expected given the demographics and prior achievement history of those students. This means, for instance, that a high-poverty school with overall low test scores could still have strong positive value-added scores because the students make greater gains than expected given their circumstances. Notably, this is quite different from common metrics of school quality, such as the percent of students who score proficient or advanced on state assessments, which are oftentimes not relative but absolute measures of performance that closely approximate the socioeconomic composition of a school.

Because fewer than 5% of students in our sample changed schools from K to 3rd grade, we take the average value-added score across children's kindergarten through 3rd grade years. This value-added measure of school quality ranges between -7.6 and 7.9 with a mean of 1.2 in the analytic sample, meaning that, on the low end, school performance growth was 7.6% points below the expected growth rate, and, on the high end, performance growth was 7.9% points above the expected growth rate, with the performance growth at the average school exceeding the expected growth rate by 1.2% points. To aid interpretation, we standardize this measure to have a mean of zero and standard deviation of one across the schools in the sample.

Analytic Strategy

Our overarching research question is: do later higher quality teachers and schools provide sustaining environments for the positive achievement effects found for VPK participants relative to nonparticipants at the end of the PreK year? To inform this question, we focus on 2-way interactions of PreK participation and teacher effectiveness and PreK participation and school quality, and the 3-way interaction of PreK participation, teacher effectiveness, and school quality as predictors of student test scores in math and ELA. In an ideal scenario we would derive this effect estimate with sequential randomization with students randomly assigned to PreK and then randomly assigned to teachers and schools with different levels of effectiveness and quality. In lieu of sequential randomization, we assume equal exposure of PreK participants and non-participants to subsequent quality schools and effective teachers; that is, we assume that PreK itself does not affect such exposure net of relevant covariates. This is a testable assumption that we assessed (see Table 1). We begin by conducting a moderated multiple regression analysis that

takes the following form:

$$Y_i = \beta_0 + \beta_1 VPK_i + \beta_2 SQ_i + \beta_3 TE_i + \beta_4 (VPK_i \times SQ_i) + \beta_5 (VPK_i \times TE_i) \\ + \beta_6 (SQ_i \times TE_i) + \beta_7 (VPK_i \times SQ_i \times TE_i) + z'_i + \varepsilon_i$$

where Y_i is a standardized measure of TCAP math or ELA scores in third grade for student i ; VPK_i is dummy variable coded 1 if student participated in VPK and zero otherwise; SQ_i is average school quality for student i between kindergarten and 3rd grade; TE_i is an indicator of the number of highly effective teachers student i had between kindergarten and 3rd grade; z'_i is a vector of student baseline characteristics (age, sex, race/ethnicity, and primary language); and ε_i is the individual residual clustered at the school level (substantive conclusions are qualitatively similar if cluster robust standard errors are used; see [Tables 1](#) and [2](#)). In addition to the vector of control variables specified by z'_i , all models use inverse probability of treatment weighting (IPTW) to adjust for any differences in baseline characteristics across PreK participants and non-participants (Austin, 2011).⁶ The coefficients of interest are the two- and three-way interactions among VPK participation, teacher effectiveness, and school quality.⁷ To be clear, the treatment-control contrasts of interest are being treated as a quasi-experiment with associated procedures to adjust for baseline differences that may result in bias. (We also conduct robustness checks based on alternative approaches for handling missing data.)

Results

Descriptive Summary of Teacher and School Quality by VPK Participants and Non-Participants

We first present descriptive information on baseline student characteristics and subsequent teacher and school quality for VPK participants and non-participants. Column (2) in [Table 1](#) displays means for the control group. Column (3) reports unadjusted differences

⁶Inverse probability of treatment weights (IPTW) were derived from the following equation:

$$w_i = \frac{Z_i}{e_i} + \frac{(1-Z_i)}{1-e_i},$$

where Z_i is an indicator variable for whether child i was a VPK participant, and where e_i denotes the probability that child i was a VPK participant, calculated from a logistic regression of PreK participation on age, race, gender, and whether children's primary language was English (these variables included all the covariates used in the primary analysis). Each child's weight is therefore equal to the inverse of the probability of receiving the treatment that the child actually received. Of note, estimates based on stabilized IPTWs, which reduce the weight of either PreK participants with low propensity scores or PreK non-participants with high propensity scores, produced virtually identical point estimates (see [Tables 1](#) and [2](#)). (Stabilized IPTWs are calculated by multiplying the IPTWs by the marginal probability of receiving the treatment actually received. See Austin (2011) for a detailed description of the application of IPTWs and stabilized IPTWs). As a point of reference, [Figure B.1](#) in the [Appendix](#) shows that the distribution and range of propensity scores were similar across treatment conditions and that there existed no threats to or violations of the common support assumption.

⁷We do not include fixed effects for randomization site because there is limited variation in the number of schools attended by children from the same randomization list, especially in rural areas in Tennessee. In particular, only 40% of the variation in school quality is accounted for by students on the same randomization list, meaning that the majority of information about school quality would be lost in an analysis that included fixed effects for randomization site. Consequently, the inclusion of randomization pool fixed effects would reduce our ability to identify whether PreK effects differed based on the quality of children's subsequent learning environments and are not included in our primary analysis. [Tables 4](#) and [5](#) report results from robustness checks that include randomization pool fixed effects. Of note, [Figure C.1](#) in the [Appendix](#) provides little evidence that the relation between randomization list and subsequent school quality varied by baseline student characteristics.

between PreK participants and PreK non-participants on observable characteristics. Columns (4) reports adjusted differences based on propensity score weighting. For comparison, Column (1) shows means for all children in elementary schools in Tennessee.

Column (3) indicates that VPK participants relative to non-participants were comprised of students who were, on average, more likely to be White, more likely to speak English as their primary language, and less likely to be Hispanic. However, Column (4) indicates that these baseline differences on student-level characteristics were effectively balanced after the inclusion of inverse probability of treatment weights.

Moreover, PreK participants and non-participants had similar exposure to numbers of highly effective teachers between kindergarten and 3rd grade in both unadjusted and adjusted comparisons. Finally, we found no evidence in unadjusted or adjusted comparisons that the schools of students who attended VPK differed in quality from those of children who did not attend VPK. Taken together, there is no evidence based on observable comparisons of teacher and school quality that PreK participants actively sought out and enrolled in higher quality learning environments in the years after PreK more so than PreK non-participants. In other words, there is no evidence that differential selection of treatment groups into teachers and schools of varying quality in the years after PreK biased our estimates.

Of note is the distribution for the total number of highly effective teachers a student encounters from K to 3rd grade. Although we do not find any significant differences between VPK participants and non-participants, there is considerable variation in the totals and the distribution is skewed such that very few VPK participants and non-participants had access to multiple highly effective teachers. Indeed, more than 20% of students in each condition were never enrolled in a highly effective teacher's classroom from K to 3rd grade, which contrasts with only 3% of Tennessee students overall who never had exposure to a highly effective teacher from K to 3rd grade. Moreover, while more than 50% of elementary school students across Tennessee were exposed to a highly effective teacher in at least three of four years between K to 3rd grade, only 20% of students in the analytic sample had such exposure.

Table 2 presents the joint distribution of teacher and school quality across the analytic sample. We define low-quality schools as those with gain scores that did not meet state growth standards while we define high-quality schools as those with gain scores 3 points above average (which corresponds to roughly 1 standard deviation above average). Moderate-quality schools were defined as those in between. As indicated by Table 2, there was substantial variation in the types of learning environments that VPK participants and non-participants experienced between kindergarten and 3rd grade even among children who had similar levels of exposure to teacher versus school quality. For instance, despite the fact that over 20% of children were never exposed to a highly effective teacher between kindergarten and third grade, three out of four of these children nevertheless attended a school that met or exceeded state growth standards. Similarly, 20% of children who had three or four highly effective teachers between kindergarten and third grade nonetheless attended a school that did not meet state growth standards. These patterns of variation underscore the importance of modeling not only the independent effects of high quality schools and highly effective teachers but also the combined effects of exposure to both.

Achievement Outcomes

Number of Highly Effective Teachers and School Quality

Panels A and B of Table 3 present results for 3rd grade test scores for ELA and mathematics, respectively. Column (1) in each panel presents estimates of the covariate-adjusted differences between VPK participants and non-participants on 3rd grade achievement as well as the incremental difference in achievement associated with having an additional highly effective teacher and attending a school with a one standard deviation increase in value added. Column (2) in each table includes the same covariates as Column (1) but adds an interaction between VPK enrollment and exposure to an additional highly effective teacher. Column (3) in each table replaces the interaction between VPK enrollment and teacher effectiveness with an interaction between VPK enrollment and school quality. Column (4) presents estimates from our fully-specified

Table 3. TN-VPK effect moderation by number of high quality teachers and average school quality from kindergarten through 3rd grade.

	Main effect (1)	Teacher interaction (2)	School interaction (3)	3-Way interaction (4)
Panel A. 3rd grade ELA achievement				
VPK	0.01 (0.07)	−0.01 (0.13)	0.01 (0.07)	0.00 (0.14)
# HE teachers	0.04 (0.04)	0.03 (0.07)	0.16 (0.16)	0.04 (0.06)
School quality	−0.01 (0.04)	−0.01 (0.04)	−0.07 (0.07)	0.07 (0.10)
HE teachers * School quality				−0.14* (0.06)
VPK * HE teachers		0.01 (0.08)		−0.01 (0.07)
VPK * School quality			0.11 (0.08)	−0.05 (0.14)
VPK * HE teachers * School quality				0.15* (0.07)
R ²	0.05	0.05	0.05	0.06
n =	806	806	806	806
Panel B. 3rd grade math achievement				
VPK	−0.02 (0.07)	−0.09 (0.13)	−0.02 (0.07)	−0.08 (0.12)
# HE teachers	0.05 (0.05)	0.02 (0.07)	0.11 (0.18)	0.03 (0.06)
School quality	0.04 (0.05)	0.04 (0.05)	−0.03 (0.07)	0.16† (0.09)
HE teachers * School quality				−0.17*** (0.05)
VPK * HE teachers		0.05 (0.08)		0.03 (0.07)
VPK * School quality			0.13† (0.07)	−0.05 (0.11)
VPK * HE teachers * School quality				0.17* (0.07)
R ²	0.03	0.03	0.03	0.05
n =	806	806	806	806

Note: This table provides coefficient estimates from an OLS regression of children’s 3rd grade achievement on an indicator for VPK enrollment and interactions between VPK enrollment and quality measures at the teacher and school level during children’s elementary grades. All models controlled for children’s age, race, gender, and primary language. All estimates used propensity score weighting. Standard errors were clustered at the school level. †p < .10, *p < .05, ***p < .001 for two-tailed tests of significance.

model that includes a three-way interaction between VPK, teacher effectiveness, and school quality, as well as the low-order, two-way interaction terms.

Model 1 in each panel provides no evidence that 3rd grade achievement in either ELA or math differed between VPK participants and non-participants.⁸ Moreover, there is no evidence that having an additional highly effective teacher or attending a high quality school, independent of whether a child attended VPK, was associated with differences in 3rd grade achievement in either ELA or Math. Although these main effect estimates provide insight about the magnitude of the average treatment effect of attending VPK versus being exposed to a high quality elementary school experience as measured by teacher and school quality, the current study is most interested in whether PreK effects were more or less likely to persist depending on the quality of a children's subsequent learning environment.

The first research question we sought to answer was whether the association between VPK participation and 3rd grade achievement was conditional on the number of highly effective teachers that children had from kindergarten to 3rd grade. As indicated in Column (2) of Panels A and B, there is no evidence that the number of high effective teachers children had between kindergarten and 3rd grade were related to 3rd grade ELA or math achievement scores.

Column (3) turns attention to the question of whether the association between VPK participation and 3rd grade achievement was conditional on the quality of the schools children attended between kindergarten and 3rd grade. VPK participants scored higher than non-participants in both mathematics and ELA if they went on to attend higher quality schools ($\beta = 0.11$ and $\beta = 0.13$, respectively). However, these estimates were only marginally significant in ELA ($p = .075$) and not statistically different from zero in math ($p = .154$). This imprecision may be due to the fact that these conditional associations do not account for the distribution of high quality teachers within schools of a given quality. In other words, children attending similar schools in terms of quality may have varied in terms of the numbers of highly effective teachers they had, a subtlety not accounted for in Model 3.

Model 4 in Panels A and B does account for these differences. Model 4 addresses the research question of whether the associations between PreK participation and 3rd grade achievement were conditional not only on the quality of the schools children attended from kindergarten to 3rd grade but also on the number of highly effective teachers that students had. This question concerns potential three-way interactions between VPK exposure, school quality, and teacher effectiveness. Our estimates reveal that VPK participants scored highest relative to non-participants if children subsequently attended high quality schools and had highly effective teachers (ELA: $\beta = 0.15$, $p = .040$; math: $\beta = 0.17$, $p = .016$).

To provide some intuition for what these three-way interactions mean, [Figures 1](#) and [2](#) plot the marginal effect of VPK on 3rd grade ELA and math achievement, respectively, across levels of school quality (-2 SD to $+2$ SD) for students with zero,

⁸Notably, Lipsey et al. (2018) found that TN-VPK non-participants outperformed participants in 3rd grade math achievement by a statistically significant margin. That our estimate differs from Lipsey et al.'s may be due to the smaller sample size on which our estimate is based (our analytic sample is a subsample of the one used in Lipsey et al.'s study), as well as the fact that our study controls for subsequent learning environments whereas Lipsey et al. did not.

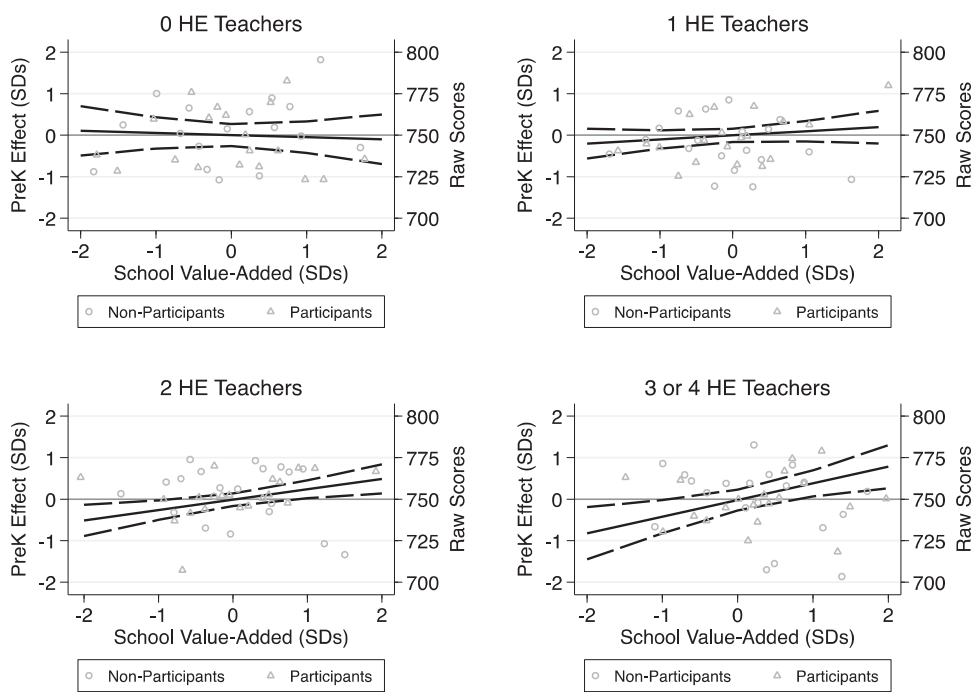


Figure 1. Marginal effect of TN-VPK on 3rd grade ELA achievement across. Standardized school-level value added scores ($n = 806$). *Note:* The solid black line in each panel plots the marginal effect of PreK on 3rd grade ELA achievement (left y-axis) across levels of school value-added (x-axis) for students with zero, one, two, and three or four highly effective teachers. The dotted lines correspond to a 95% confidence interval. To provide an understanding of the distribution of the underlying achievement scores on which these marginal effects were based, the background of each panel provides a binned scatter plot of the conditional association between 3rd grade ELA achievement (right y-axis) and school-level value added scores (x-axis) for students with zero, one, two, and three or four highly effective teachers. The gray triangles refer to achievement scores for PreK participants; the gray circles refer to achievement scores for non-participants. These points were constructed by regressing achievement scores on the full set of baseline child-level covariates (race, age, primary language, and gender), regressing school-level value added scores on the same baseline child-level outcomes, then plotting the relationship between the residuals from each of these regressions for children who had zero, one, two, and three or four highly effective teachers. We constructed 20 equal size bins of the residuals for each regression, and, in each bin, plotted the mean of the residuals from each regression. Intuitively, the average marginal effect at each point along the x-axis can be thought of as the difference in achievement between the average score of participants and non-participants at each x-value after imposing a linearity assumption.

one, two, and three (or four) highly effective teachers. The solid line plots the point estimate and the dotted lines refer to the 95% confidence interval. Point estimates are considered statistically significant wherever the confidence interval excludes zero. In the background of each figure is a binned scatterplot of the conditional association between 3rd grade achievement and school-level value added scores for students with different numbers of highly effective teachers. (The plotted marginal effect and the binned scatterplots have the same x-axis but different y-axes.) These binned scatterplots provide an understanding of the distribution of scores for VPK participants and non-participants across varying conditions of teacher and school quality. The gray triangles refer to

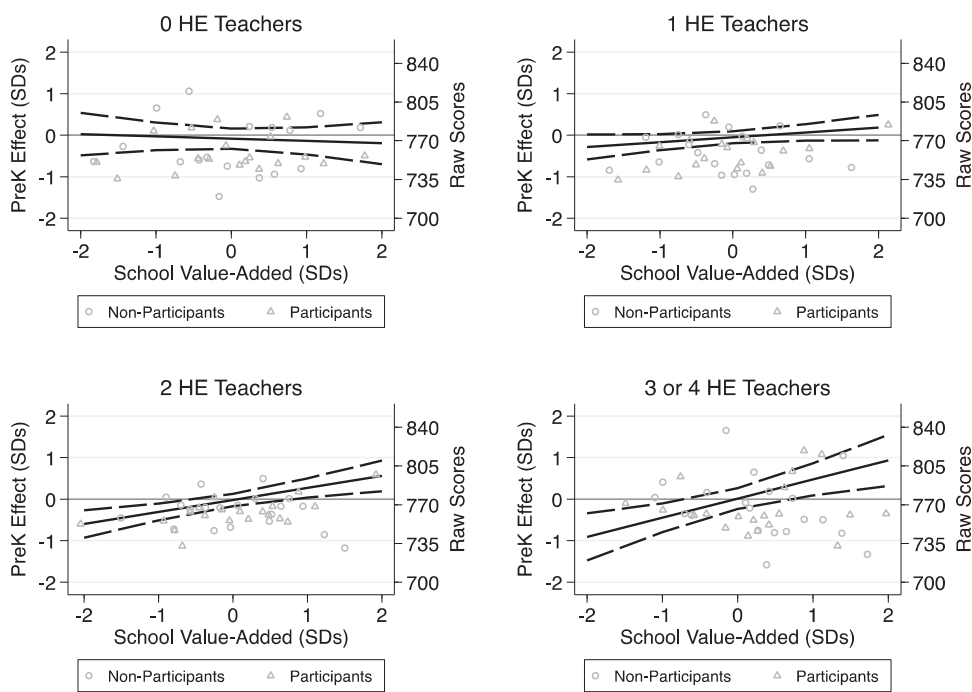


Figure 2. Marginal effect of TN-VPK on 3rd grade math achievement across. Standardized school-level value added scores ($n = 806$). *Note:* The solid black line in each panel plots the marginal effect of PreK on 3rd grade math achievement (left y-axis) across levels of school value-added (x-axis) for students with zero, one, two, and three or four highly effective teachers. The dotted lines correspond to a 95% confidence interval. To provide an understanding of the distribution of the underlying achievement scores on which these marginal effects were based, the background of each panel provides a binned scatter plot of the conditional association between 3rd grade math achievement (right y-axis) and school-level value added scores (x-axis) for students with zero, one, two, and three or four highly effective teachers. The gray triangles refer to achievement scores for PreK participants; the gray circles refer to achievement scores for non-participants. These points were constructed by regressing achievement scores on the full set of baseline child-level covariates (race, age, primary language, and gender), regressing school-level value added scores on the same baseline child-level outcomes, then plotting the relationship between the residuals from each of these regressions for children who had zero, one, two, and three or four highly effective teachers. We constructed 20 equal size bins of the residuals for each regression, and, in each bin, plotted the mean of the residuals from each regression. Intuitively, the average marginal effect at each point along the x-axis can be thought of as the difference in achievement between the average score of participants and non-participants at each x-value after imposing a linearity assumption.

achievement scores for PreK participants. The gray circles refer to achievement scores for non-participants.⁹ Intuitively, the average marginal effect at each point along the x-axis can be thought of as the difference in achievement between the average score of VPK participants versus the average score of VPK non-participants at each x-value after imposing an assumption of linearity.

⁹These points were constructed by regressing achievement scores on the full set of baseline child-level covariates (race, age, primary language, and gender), regressing school-level value added scores on the same baseline child-level outcomes, then plotting the relationship between the residuals from each of these regressions for children who had zero, 1, 2, and 3 or 4 highly effective teachers. We constructed 20 equal size bins of the residuals for each regression and, in each bin, plotted the mean of the residuals from each regression.

As shown in the top left panel of each Figure, there is no evidence that school quality moderates the estimated difference between VPK participants and non-participants when children had zero highly effective teachers between kindergarten and third grade. However, there is an increasing divergence in achievement between treatment groups in favor of VPK participants as school quality increases for those children who went on to have highly effective teachers in subsequent years. This divergence is strong enough that VPK participants outperformed non-participants in 3rd grade achievement by a statistically significant margin if they went on to have both highly effective teachers and high quality learning environments after VPK.

For instance, the estimated difference between VPK participants and non-participants in 3rd grade math achievement is positive for children with at least two highly effective teachers in schools with above average quality; however, this estimate becomes statistically significant when students attend schools with value-added scores that are one standard deviation above the mean ($\beta = 0.27, p = .022$). This difference intensifies in even more enriching learning contexts. For instance, the estimated difference between VPK participants and non-participants who had three highly effective teachers and attend schools with value-added scores that are two standard deviations above the mean is 0.93 standard deviations ($p = .003$). Virtually the same pattern holds for ELA: The estimated difference between VPK participants and non-participants in 3rd grade ELA becomes statistically significant when students have at least two teachers and attend schools with value-added scores that are one standard deviation above the mean ($\beta = 0.24, p = .030$), a difference that similarly grows if the number of highly effective teachers increases along with school value-added scores.

Notably, the marginal effects in [Figures 1 and 2](#) provide evidence that non-participants out-performed participants in 3rd grade math and ELA achievement when these children went on to attend low-quality schools with at least two highly effective teachers. That is, students in low quality schools with good teachers were estimated to be better off if they did not attend VPK than if they attended VPK. These patterns emerge as statistically significant for math and ELA in schools with value added scores that were 1 standard deviation below the mean (math: $\beta = -0.30, p = .003$; ELA: $\beta = -0.26, p = .028$; ELA). However, the binned scatterplots in the background reveal that there were very few binned observations that experienced multiple highly effective teachers in low quality schools, suggesting that this negative effect may be negligible in practice.

Timing of Exposure to Highly Effective Teachers and School Quality

A related research question concerns the timing of exposure to high quality teachers. A summary of these results is provided in [Table 4](#). Columns (1) and (2) refer to having a highly effective teacher during kindergarten and 1st grade. Columns (3) and (4) refer to having a highly effective teacher during 2nd and 3rd grade, irrespective of exposure during other years in both cases. Columns (1) and (3) provide estimates based on two-way interactions between VPK enrollment and the timing indicator. Columns (2) and (4) provide estimates based on three-way interactions between VPK enrollment, the timing indicator, and school quality. Panel A provides results for 3rd grade ELA achievement. Panel B provides results based on 3rd grade math achievement.

Overall, [Table 4](#) shows that the moderating capacity of the timing of exposure depends on the subject. For 3rd grade ELA achievement, VPK participants

Table 4. PreK effect moderation by exposure to a high quality teacher during kindergarten and 1st grade or during 2nd and 3rd grade.

	First 2 years		Last 2 years	
	With HE teacher		With HE teacher	
	(1)	(2)	(3)	(4)
Panel A. 3rd grade ELA achievement				
VPK	0.02 (0.08)	0.03 (0.08)	0.02 (0.08)	0.03 (0.08)
Exposure to HE	−0.05 (0.25)	−0.14 (0.20)	0.21 (0.15)	0.26† (0.15)
Teacher both years				
School quality	−0.01 (0.04)	−0.01 (0.07)	−0.02 (0.04)	−0.07 (0.07)
HE teachers * School quality		−0.50** (0.19)		−0.09 (0.19)
VPK * HE teachers	−0.05 (0.26)	0.04 (0.21)	−0.10 (0.22)	−0.17 (0.22)
VPK * School quality		0.09 (0.09)		0.11 (0.09)
VPK * HE teachers * School quality		0.38† (0.21)		0.16 (0.22)
R^2	0.05	0.06	0.05	0.05
$n =$	806	806	806	806
Panel B. 3rd grade math achievement				
VPK	−0.00 (0.08)	0.01 (0.08)	−0.04 (0.08)	−0.04 (0.08)
Exposure to HE	−0.06 (0.23)	−0.13 (0.21)	0.24 (0.18)	0.43* (0.18)
Teacher both years				
School quality	0.05 (0.05)	0.02 (0.07)	0.03 (0.05)	0.01 (0.07)
HE teachers * School quality		−0.38* (0.18)		−0.61** (0.19)
VPK * HE teachers	−0.05 (0.24)	0.03 (0.22)	0.15 (0.25)	−0.05 (0.24)
VPK * School quality		0.13 (0.08)		0.10 (0.07)
VPK * HE teachers * School quality		0.23 (0.23)		0.57* (0.23)
R^2	0.03	0.04	0.04	0.06
$n =$	806	806	806	806

Note: This table provides coefficient estimates from an OLS regression of children's 3rd grade achievement on an indicator for VPK enrollment and interactions between VPK enrollment and quality measures at the teacher and school level during children's elementary grades. All models controlled for children's age, race, gender, and primary language. All estimates used propensity score weighting. Standard errors were clustered at the school level. † $p < .10$, * $p < .05$, ** $p < .01$ for two-tailed tests of significance.

outperformed non-participants by the largest margin if these children attended high quality schools and had highly effective teacher *in the 2 years immediately following VPK* ($\beta = 0.38$, $p = .072$). No evidence was found that having highly effective teachers only in 2nd and 3rd grades moderated the observed differences in 3rd grade ELA achievement between VPK participants and non-participants across levels of school quality.

For 3rd grade math achievement, the pattern of timing is just the opposite. VPK participants outperformed non-participants by the largest margin if these children attended high quality schools and had highly effective teachers *in the 2 years preceding statewide assessments* ($\beta = 0.57$, $p = .012$). No evidence was found that having highly effective teachers only in kindergarten and 1st grade moderated the observed differences in 3rd

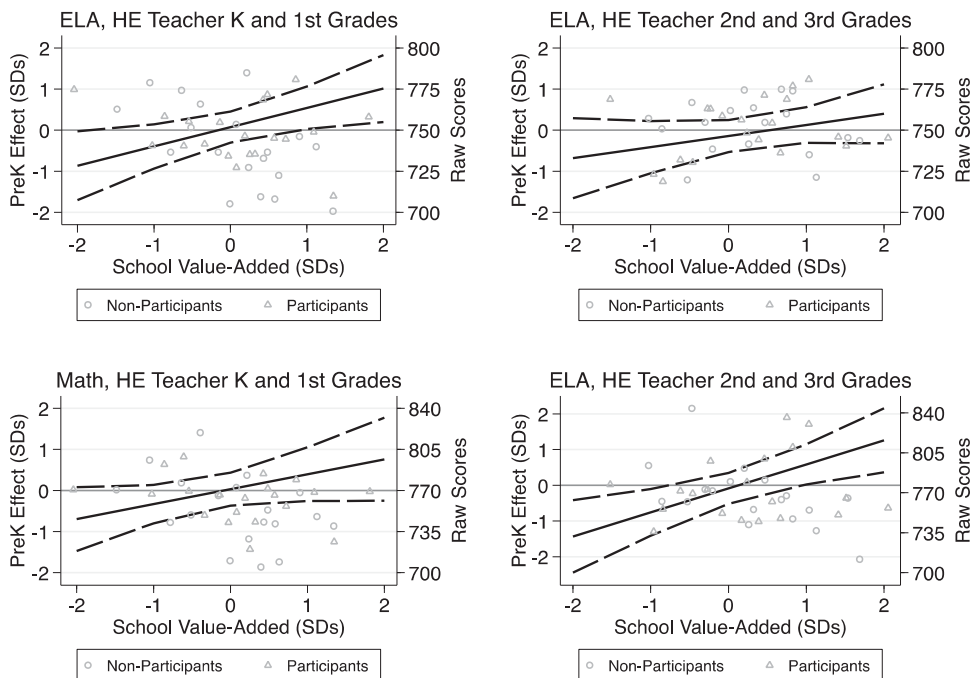


Figure 3. Marginal Effect of TN-VPK on 3rd grade achievement across school value-added scores for children with at least one high quality teacher in kindergarten and 1st grade versus 2nd and 3rd grades. *Note:* The top two panels refer to ELA achievement. The bottom two panels refer to math achievement. The solid black line in each panel plots the marginal effect of PreK on 3rd grade achievement (left y-axis) across levels of school value-added (x-axis) for students with a highly effective teacher during kindergarten and 1st grade versus 2nd and 3rd grade. The dotted lines correspond to a 95% confidence interval. To provide an understanding of the distribution of the underlying achievement scores on which these marginal effects were based, the background of each panel provides a binned scatter plot of the conditional association between 3rd grade achievement (right y-axis) and school-level value added scores (x-axis) for students with a highly effective teacher during kindergarten and 1st grade versus 2nd and 3rd grade. The gray triangles refer to achievement scores for PreK participants; the gray circles refer to achievement scores for non-participants. These points were constructed by regressing achievement scores on the full set of baseline child-level covariates (race, age, primary language, and gender), regressing school-level value added scores on the same baseline child-level outcomes, then plotting the relationship between the residuals from each of these regressions for children who had a highly effective teacher during kindergarten and 1st grade versus 2nd and 3rd grade. We constructed 20 equal size bins of the residuals for each regression, and, in each bin, plotted the mean of the residuals from each regression. Intuitively, the average marginal effect at each point along the x-axis can be thought of as the difference in achievement between the average score of participants and non-participants at each x-value after imposing a linearity assumption.

grade math achievement between VPK participants and non-participants across levels of school quality.

For illustrative purposes, Figure 3 plots these three-way interactions based on the timing of exposure variable and show the specific conditions under which VPK participants outperformed non-participants along with binned scatterplots as described above. The general pattern for ELA achievement reveals significant differences in favor of VPK participants when they have two highly effective teachers in kindergarten and 1st grade and are enrolled in schooling environments that contribute meaningfully to their education

Table 5. Robustness checks of PreK effect moderation by number of high quality teachers and average school quality from kindergarten through 3rd grade: 3rd grade reading achievement.

	Reported Results (1)	ITT Indicator (2)	Randomization Pool FEs (3)	Bootstrapped SEs (4)	Exclusion of IPTWs (5)	Stabilized IPTWs (6)	Imputed Covariates (7)
VPK	0.00 (0.14)	−0.10 (0.13)	0.09 (0.15)	0.00 (0.14)	0.02 (0.13)	0.00 (0.14)	0.05 (0.13)
# HE teachers	0.04 (0.06)	0.01 (0.06)	0.12* (0.05)	0.04 (0.06)	0.05 (0.06)	0.04 (0.06)	0.03 (0.06)
School quality	0.07 (0.10)	0.03 (0.09)	0.01 (0.10)	0.07 (0.11)	0.03 (0.09)	0.08 (0.10)	0.06 (0.09)
HE teachers School quality	−0.14* (0.06)	−0.04 (0.06)	−0.10† (0.05)	−0.01 (0.08)	−0.12* (0.06)	−0.14* (0.06)	−0.09 (0.06)
VPK * HE teachers	−0.01 (0.07)	0.06 (0.07)	−0.05 (0.08)	−0.05 (0.14)	−0.01 (0.07)	−0.01 (0.07)	0.00 (0.08)
VPK * School quality	−0.05 (0.14)	0.04 (0.14)	0.00 (0.14)	−0.14* (0.06)	−0.01 (0.13)	−0.05 (0.14)	−0.05 (0.13)
VPK * HE teachers * School quality	0.15* (0.07)	−0.01 (0.07)	0.11 (0.07)	0.15† (0.08)	0.13† (0.07)	0.15* (0.07)	0.10 (0.07)
R ²	0.06	0.05	0.18	0.06	0.06	0.06	0.05
n =	806	806	806	806	806	806	1,006

Note: This table provides coefficient estimates from a series of robustness checks. All models controlled for children's age, race, gender, and primary language. Column (1) presents the reported estimates described in the main text. Column (2) replaces the indicator for whether a child enrolled in PreK with an indicator of whether a child was randomly assigned to attend PreK. Column (3) includes randomization pool fixed effects. Column (4) uses bootstrapped standard errors clustered at the school level (Cameron & Miller, 2015). Column (5) excludes inverse probability of treatment weights. Column (6) replaces inverse probability of treatment weights with stabilized inverse probability of treatment weights that reduce the influence of observations with very high or very low propensity scores. Column (7) reports results after imputing missing values on an baseline covariates. † $p < .10$, * $p < .05$ for two-tailed tests of significance.

in terms of school-level value-added scores, as illustrated in the top left plot. For math achievement, significant differences between VPK participants and non-participants emerged for children who had highly effective teachers in 2nd and 3rd grade and who were enrolled in high quality schooling environments in terms of value added scores, as illustrated in the bottom right plot.

Robustness Checks

Tables 5 and 6 provide a series of robustness checks. In general, substantive conclusions about the joint, moderating capacity of high quality teachers and schools with respect to children's 3rd grade reading achievement were robust to the exclusion of inverse probability of treatment weights, the use of stabilized inverse probability of treatment weights, and the use of cluster robust standard errors as opposed to clustering standard errors at the school level. The magnitude of the three-way interaction diminishes somewhat but remains positive when including randomization pool fixed effects as well as when imputing baseline covariates, although these estimates are no longer statistically different from zero at conventional levels. These patterns of robustness were generally similar with respect to 3rd grade math achievement with the addition that significant three-way interactions were also observed when including randomization pool fixed effects and when baseline covariates were imputed. Of note, we found no evidence of significant three-way interactions when using an intent-to-treat indicator of treatment assignment rather than a treatment-on-treated indicator of treatment assignment. We

Table 6. Robustness checks of PreK effect moderation by number of high quality teachers and average school quality from kindergarten through 3rd grade: 3rd grade math achievement.

	Reported Results (1)	ITT Indicator (2)	Randomization Pool FEs (3)	Bootstrapped SEs (4)	Exclusion of IPTWs (5)	Stabilized IPTWs (6)	Imputed Covariates (7)
VPK	−0.08 (0.12)	−0.18 (0.12)	0.02 (0.13)	−0.07 (0.12)	−0.07 (0.12)	−0.08 (0.12)	−0.08 (0.12)
# HE teachers	0.03 (0.06)	0.01 (0.06)	0.11* (0.06)	0.04 (0.06)	0.04 (0.06)	0.04 (0.06)	0.02 (0.06)
School quality	0.16† (0.09)	0.11 (0.09)	0.11 (0.10)	0.13 (0.11)	0.13 (0.09)	0.16† (0.09)	0.10 (0.08)
HE teachers * School quality	−0.17*** (0.05)	−0.09† (0.05)	−0.12* (0.05)	0.02 (0.07)	−0.16** (0.05)	−0.17*** (0.05)	−0.12* (0.05)
VPK * HE teachers	0.03 (0.07)	0.09 (0.07)	−0.01 (0.08)	−0.03 (0.13)	0.02 (0.07)	0.03 (0.07)	0.04 (0.07)
VPK * School quality	−0.05 (0.11)	0.03 (0.13)	−0.06 (0.12)	−0.16* (0.06)	−0.03 (0.12)	−0.05 (0.11)	−0.02 (0.10)
VPK * HE teachers * School quality	0.17* (0.07)	0.04 (0.07)	0.14† (0.08)	0.16† (0.08)	0.16* (0.07)	0.17* (0.07)	0.12* (0.06)
R ²	0.05	0.04	0.16	0.05	0.05	0.05	0.04
n =	806	806	806	806	806	806	1,006

Note: This table provides coefficient estimates from a series of robustness checks. All models controlled for children's age, race, gender, and primary language. Column (1) presents the reported estimates described in the main text. Column (2) replaces the indicator for whether a child enrolled in PreK with an indicator of whether a child was randomly assigned to attend PreK. Column (3) includes randomization pool fixed effects. Column (4) uses bootstrapped standard errors clustered at the school level (Cameron & Miller, 2015). Column (5) excludes inverse probability of treatment weights. Column (6) replaces inverse probability of treatment weights with stabilized inverse probability of treatment weights that reduce the influence of observations with very high or very low propensity scores. Column (7) reports results after imputing missing values on an baseline covariates. † $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$ for two-tailed tests of significance.

attribute this pattern to the noncompliance rates noted earlier that degraded the intent-to-treat indicator as a representation of actual participation in the PreK program. However, it remains possible that parents who crossed over on treatment status may have changed the relationship between unobserved determinants of student achievement and school or teacher quality, and thus is still the primary threat to the internal validity of this study.

Discussion

Recent studies have found that the test score benefits of PreK participation fade relatively quickly once participants and non-participants progress into elementary school (Hill et al., 2015; Lipsey et al., 2018; Puma et al., 2010, 2012). In light of these findings, scholars have devoted increasing attention to understanding the source of the decreased effects of PreK participation on test scores. In particular, there is growing interest into whether a high quality subsequent learning environment might serve as what has been termed a sustaining environment (Bailey et al., 2017). In this study, we examined the intersection of two quality indicators of subsequent learning environments: school quality as measured by the average value-added scores of the schools that children attended between kindergarten and 3rd grade, and exposure to high quality teaching as measured by the number of highly effective teachers students had during these years. We find that the academic advantage of VPK participants versus non-participants at kindergarten

entry was most likely to persist until 3rd grade among those students who went on to attend high quality schooling environments and were taught by highly effective teachers. These findings were generally robust to a variety of alternative specifications.

These results have a number of implications for theory and research about early childhood education. First, our finding about the joint moderating capacity of highly effective teachers teaching in high quality schools suggests that supporting early gains from PreK may require exposure to both as opposed to either. This finding may reconcile some of the previous debates about the role subsequent learning environments play in the persistence of PreK effects. Prior research has been mixed regarding whether teachers, classrooms, and/or schools moderate the effects of PreK during the elementary grades. One reason for these incongruent results could be that prior research into the sustaining environments hypothesis has focused primarily on the moderating capacity of either teacher quality or school quality without considering whether the moderating effect of one depended on the other. Indeed, our study found no consistent evidence that either the number of highly effective teachers to which children were exposed or the quality of the schools children attended alone was adequate to explain differential achievement between PreK participants and non-participants in 3rd grade. It was only among the subgroup of children who had multiple highly effective teachers and who attended high quality schools wherein PreK participants were found to outperform their non-participant peers in 3rd grade. Future research into PreK effect persistence would do well to consider how quality interacts and is arrayed at different levels of children's schooling experience—from the teacher to the school itself.

One interesting finding from our study was that we found no evidence of significant main effects for either teacher or school quality. On face value, this finding appears to stand in contrast to a robust literature on the unique academic benefits associated with exposure to high quality teachers and high quality schools (e.g., Davis & Warner, 2018; Harris & Sass, 2011; Rice, 2003; Rivkin et al., 2005). However, there are at least two plausible explanations for these null effects. First, despite that school-level growth rates are an adequate measure of the amount of learning that takes place in a school, recent evidence has shown that school-level growth rates—which comprised our measure of school quality—do not correlate with achievement levels because baseline achievement levels vary considerably across schools (Reardon, 2019). That is, a school with high levels of achievement growth does not necessarily have high levels of achievement (and vice versa). Second, it is possible that the academic needs of low-income children, who generally trail behind their peers academically, may require accommodations that are not captured by conventional measures of teacher quality or by school-level growth rates. In fact, results from the current study suggest that exposure to PreK may be a prerequisite for low-income children benefiting from having a high quality teacher or attending an elementary school wherein a great deal of learning takes place.

In addition to our finding that having a highly effective teacher was associated with increased performance among VPK participants relative to non-participants so long as these teachers taught in high-quality schooling environments, we also found that the timing associated with having a highly effective teacher, in terms of its moderating capacity, differed by subject. Attending a high-quality school and having a highly effective teacher immediately after VPK, in kindergarten and 1st grade, was most beneficial for ELA achievement, while attending a high-quality school and having a highly effective teacher in 2nd and 3rd grade was most beneficial for math achievement.

One speculative explanation for this pattern may have to do with the timing of when ELA versus math is conventionally emphasized in elementary classrooms. In particular, prior research has indicated that preschools often place disproportionate emphasis on literacy instruction relative to math instruction (Farran et al., 2017). Moreover, there is evidence from Tennessee that this early focus on literacy (and deemphasis on math) may persist into the early elementary grades (Farran et al., 2018). Therefore, it is possible that advantages associated with having a high-quality teacher in kindergarten and first grade may have been restricted to ELA because ELA is what these teachers primarily focused on.

In any case, these findings about the timing of exposure to highly effective teachers should be understood in light of recent research in Tennessee that has found that the most effective teachers in the elementary grades are often pushed to teach in the later grades of elementary school, presumably because these are the years during which state assessments occur (Doan & Rogers, 2019). In fact, this research has shown that teachers in the upper elementary grades are more likely to be reassigned to teach in lower elementary grades if these teachers receive low scores on effectiveness ratings. The findings of the current study suggest that such a pattern of teacher assignment that places less emphasis on the quality of teachers during the earliest grades may hinder expected benefits associated with investments in preschool in terms of ELA achievement but may be less consequential for math achievement.

Finally, our findings that VPK participants outperformed non-participants only if they went on to attend high-quality schools with a succession of highly effective teachers should be understood within the context of how many children in the analytic sample actually experienced these types of high-quality learning environments. As indicated in Table 2, only 12% of children in the analytic sample (a) attended a high-quality school between kindergarten and 3rd grade, and (b) had one or more highly effective teachers during these years. This contrast with over 40% of children in the analytic sample that either attended a school that did not meet state growth standards or had zero highly effective teachers between kindergarten and 3rd grade.

Indeed, these patterns provide some understanding of why previous TN-VPK research found null effects, on average, on 3rd-grade achievement (see Lipsey et al., 2018). In particular, as this study points out, very few low-income children in Tennessee experienced learning conditions that we would reasonably expect to sustain early advantages associated with VPK participation. Moreover, our findings about the overexposure of children to low-quality schooling environments after PreK align with those from previous research (e.g., Currie & Thomas, 2000; Lee & Loeb, 1995) and are both encouraging and sobering—encouraging that high-quality learning environments after PreK can possibly sustain PreK effects but sobering that business as usual results in so few low-income children being exposed to such conditions. In other words, it is promising that having highly effective teachers and attending a high-quality school may provide a sustaining environment for PreK effects, but this promising finding is tempered by the fact that very few low-income children who qualified for VPK actually experienced learning conditions in subsequent years that would reasonably approximate a sustaining environment.

One potential strategy for counteracting the inequitable distribution of high quality teachers among schools within districts is paying high quality teachers a premium for teaching in high poverty schools. Prior research has shown that retention and

recruitment bonuses for highly effective teachers not only increase student learning in high poverty schools but also increase the likelihood that highly effective teachers teach in high poverty schools (Springer et al., 2016; Swain et al., 2019). If recruitment and retention bonuses operate as intended, these interventions would function, in part, to promote a sustaining environment by increasing the number of highly effective teachers to which low-income students are exposed in the years subsequent to PreK. It should be noted, however, that this strategy may be insufficient to produce a high-quality school, which likely depends on other factors as well.

Limitations

Although this study extends prior research about the persistence of PreK effects by highlighting a key interaction between teacher and school quality, this study is not without limitations. First, given that we were unable to randomly assign students to schools and teachers of varying quality, and given the limited number of covariates available to us, we were unable to establish causation regarding whether attending a high quality school or having a highly effective teacher caused achievement to persist into 3rd grade. It is possible, therefore, that higher achieving students who benefited from PreK and who would otherwise outperform their non-participant peers may somehow have selected into high performing schools or into classrooms of highly effective teachers. (Notably, however, we found no evidence of observed differences in teacher and school quality between VPK participants and non-participants, suggesting that bias due to this pattern of selection was unlikely.) Second, this study was based on a subsample of the larger TN-VPK study for whom teacher observation scores were available in kindergarten, thus raising concerns about the generalizability of this study. Indeed, Table C.1 in the Appendix shows that Black students were more likely than not to have missing teacher observation data. Finally, children must qualify for free- and reduced-price lunch services to enroll in TN-VPK, thus the results of our study may not generalize to more socioeconomically-advantaged students.

Conclusion

This study provides new evidence about the persistence of PreK effects. Despite finding no evidence that having a high quality teacher or attending a high quality school was sufficient by itself to explain differences in achievement between PreK participants and non-participants in 3rd grade, this study found evidence that having both was associated with persistent gains from PreK in both math and ELA that lasted into at least 3rd grade. It is important to acknowledge, however, that very few students actually experienced these facilitative conditions in either group. These findings highlight the importance of understanding the contextual nature of subsequent learning environments. Specifically, this study suggests that quality should be understood as arrayed at multiple levels and potentially interacting in policy relevant ways. Combining PreK exposure with highly effective teachers in subsequent years may be insufficient to eliminate fade-out, but pairing high quality teachers with a broader schooling environment that fosters learning, collaboration, and creativity may provide an adequate context for sustaining early advantages associated with PreK participation.

Funding

This work was supported by the Institute of Education Sciences [R305E090009], National Institute of Child Health and Human Development [R01HD079461-01], and National Institutes of Health [3R01HD079461 - 01W1].

References

- Ansari, A., & Pianta, R. C. (2018). Variation in the long-term benefits of child care: The role of classroom quality in elementary school. *Developmental Psychology*, 54(10), 1854–1867. <https://doi.org/10.1037/dev0000513>
- Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424. <https://doi.org/10.1080/00273171.2011.568786>
- Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, 10(1), 7–5747. <https://doi.org/10.1080/19345747.2016.1232459>
- Bassok, D., Gibbs, C., & Latham, S. (2015). *Do the benefits of early childhood interventions systematically fade? Exploring variation in the association between preschool participation and early school outcomes* (No. EFPolicyWorks Working Paper Series (36)). Charlottesville, VA.
- Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open*, 2(1), 233285841561635. <https://doi.org/10.1177/2332858415616358>
- Blair, C., & Scott, K. G. (2002). Proportion of LD placements associated with low socioeconomic status: Evidence for a gradient? *The Journal of Special Education*, 36(1), 14–22. <https://doi.org/10.1177/00224669020360010201>
- Boef, A. G. C., Dekkers, O. M., Vandenbroucke, J. P., & Le Cessie, S. (2014). Sample size importantly limits the usefulness of instrumental variable methods, depending on instrument strength and level of confounding. *Journal of Clinical Epidemiology*, 67(11), 1258–1264. <https://doi.org/10.1016/j.jclinepi.2014.05.019>
- Cameron, C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372. <https://doi.org/10.3368/jhr.50.2.317>
- Claessens, A., Engel, M., & Curran, F. C. (2014). Academic content, student learning, and the persistence of preschool effects. *American Educational Research Journal*, 51(2), 403–434. <https://doi.org/10.3102/0002831213513634>
- Currie, J., & Thomas, D. (2000). School quality and the longer-term effects of head start. *The Journal of Human Resources*, 35(4), 755. <https://doi.org/10.2307/146372>
- Danielson, C. (2013). *The framework for teaching: Evaluation instrument*. <http://www.daniel-songroup.org/userfiles/files/downloads/%0A2013EvaluationInstrument.pdf>
- Davis, J. R., & Warner, N. (2018). Schools matter: The positive relationship between New York City High Schools' student academic progress and school climate. *Urban Education*, 53(8), 959–980. <https://doi.org/10.1177/0042085915613544>
- Doan, S. (2019). *What do classroom observations tell us about student success? Capturing the impacts of teachers using at-scale classroom observation scores*. [Doctoral dissertation, Vanderbilt University].
- Doan, S., & Rogers, L. K. (2019). *Exploring teacher effectiveness in early grades: A research brief on strengthening Tennessee's educator labor markets*. Tennessee Education Research Alliance.
- Engel, M., Claessens, A., & Finch, M. A. (2013). Teaching students what they already know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. *Educational Evaluation and Policy Analysis*, 35(2), 157–178. <https://doi.org/10.3102/0162373712461850>
- Farran, D., Christopher, C., Pion, G., Henley, K., Lowe, T., Norvell, J., Houseman, A., Kasul, R., & Tunstel, A. (2018). *Kindergarten follow-up report: Full report*. https://cdn.vanderbilt.edu/vu-my/wpcontent/uploads/sites/1479/2016/09/14123013/K-Followup_OverallReport_FINAL.pdf

- Farran, D. C., Meador, D., Christopher, C., Nesbitt, K. T., & Billbre, L. E. (2017). Data-driven improvement in prekindergarten classrooms: Report from a partnership in an urban district. *Child Development*, 88(5), 1466–1479. <https://doi.org/10.1111/cdev.12906>
- Gervasoni, A., & Perry, B. (2015). Children's mathematical knowledge prior to starting school and implications for transition. In B. Perry, A. MacDonald, & A. Gervasoni (Eds.), *Mathematics and transition to school. Early mathematics learning and development* (pp. 47–64). Springer.
- Harris, D. N., & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. *Journal of Public Economics*, 95(7–8), 798–812. <https://doi.org/10.1016/j.jpubeco.2010.11.009>
- Hill, C. J., Gormley, W. T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly*, 32, 60–79. <https://doi.org/10.1016/j.ecresq.2014.12.005>
- Jenkins, J. M., Watts, T. W., Magnuson, K., Gershoff, E. T., Clements, D. H., Sarama, J., & Duncan, G. J. (2018). Do High-Quality kindergarten and first-grade classrooms mitigate pre-school fadeout? *Journal of Research on Educational Effectiveness*, 11(3), 339–374. <https://doi.org/10.1080/19345747.2018.1441347>
- Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, 42(6), 304–316. <https://doi.org/10.3102/0013189X13495087>
- Lee, V. E., & Loeb, S. (1995). Where do head start attendees end up? One reason why preschool effects fade out. *Educational Evaluation and Policy Analysis*, 17(1), 62–82. <https://doi.org/10.3102/01623737017001062>
- Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, 45, 155–176. <https://doi.org/10.1016/j.ecresq.2018.03.005>
- Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, 26(1), 33–51. <https://doi.org/10.1016/j.econedurev.2005.09.008>
- Puma, M., Bell, S., Cook, R., & Heid, C. (2010). *Head Start impact study final report*. Washington, DC. <http://files.eric.ed.gov/fulltext/ED507845.pdf>
- Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). *Third grade follow-up to the head start impact study: Final report*. Washington, DC. <http://files.eric.ed.gov/fulltext/ED539264.pdf>
- Reardon. (2019). Educational opportunity in early and middle childhood: Using full population administrative data to study variation by place and age. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 5(2), 40. <https://doi.org/10.7758/rsf.2019.5.2.03>
- Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Economic Policy Institute.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458. <https://doi.org/10.1111/j.1468-0262.2005.00584.x>
- Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, 72(2–3), 104–122. <https://doi.org/10.1016/j.jue.2012.04.004>
- Springer, M. G., Swain, W. A., & Rodriguez, L. A. (2016). Effective teacher retention bonuses: Evidence from Tennessee. *Educational Evaluation and Policy Analysis*, 38(2), 199–221. <https://doi.org/10.3102/0162373715609687>
- Swain, W. A., Rodriguez, L. A., & Springer, M. G. (2019). Selective retention bonuses for highly effective teachers in high poverty schools: Evidence from Tennessee. *Economics of Education Review*, 68, 148–160. <https://doi.org/10.1016/j.econedurev.2018.12.008>
- Swain, W. A., Springer, M. G., & Hofer, K. G. (2015). Early grade teacher effectiveness and pre-K effect persistence: Evidence from Tennessee. *AERA Open*, 1(4), 233285841561275. <https://doi.org/10.1177/2332858415612751>
- Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *Future of Children*. <https://files.eric.ed.gov/fulltext/EJ1118535.pdf>

Appendix

Table A.1. Balance tests on retention through 3rd grade.

	Control group mean (1)	Exp. versus control (2)
Retention through 3rd grade	0.09	0.02 (0.02)
<i>n</i> =	348	567

Note: This table presents balance tests of equivalency for retention through third grade. Column (2) reports the difference between the PreK participants and PreK non-participants, which is estimated using a weighted OLS regression of retention on a binary indicator for Pre-K participation. The sample includes all Cohort 2 children with non-missing data on baseline characteristics. The sample sizes reported here are larger than those of the primary analytic sample because the primary analytic sample is restricted to children with valid TCAP scores in 3rd grade. Children in these retention models may or may not have valid TCAP scores in 3rd grade. By definition, children with a retention were not grade eligible for the TCAP during the wave in which achievement data were collected. Standard errors are reported in parenthesis and are clustered at the school level. Estimates are weighted using inverse probability of treatment weighting.

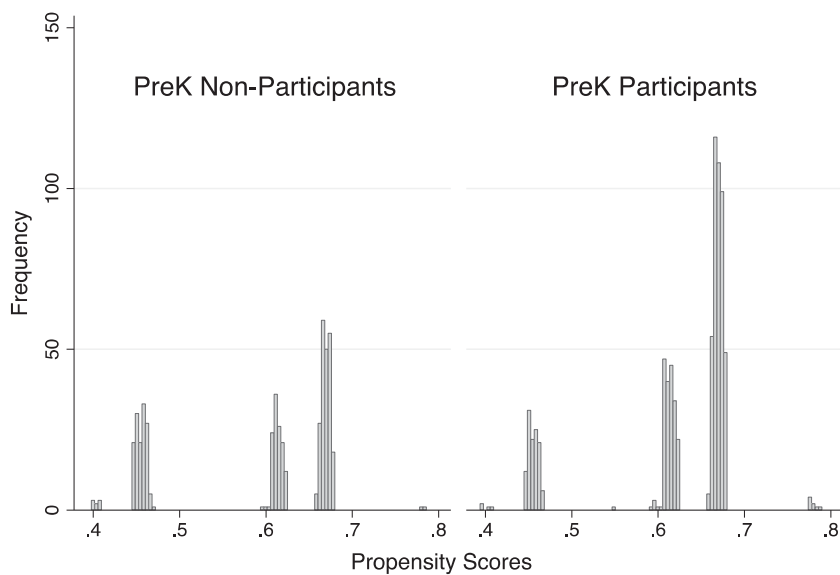


Figure B.1. Distribution of propensity scores across treatment conditions.

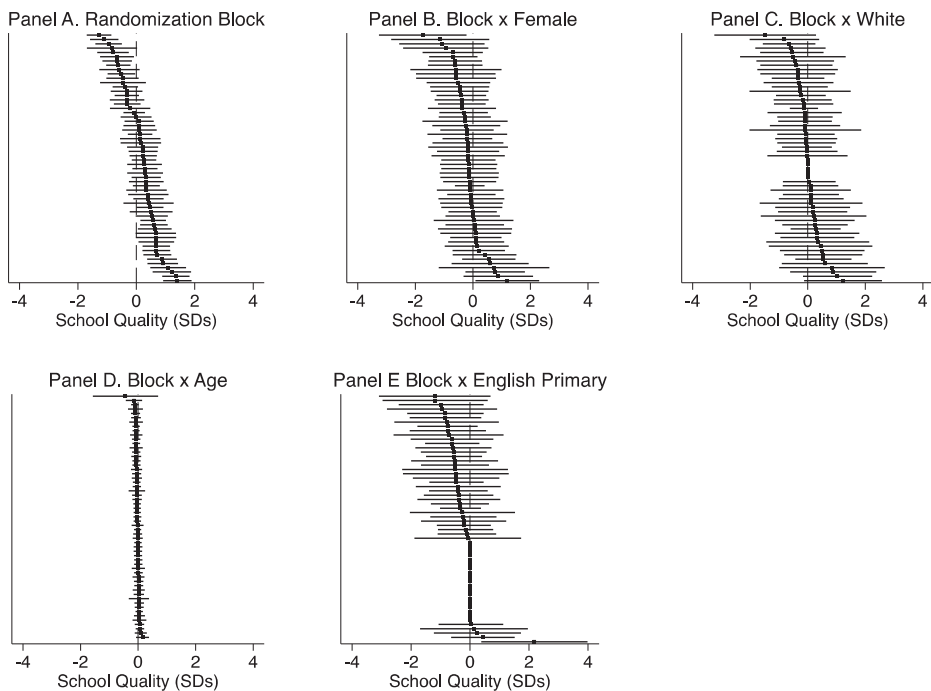


Figure C.1. Variation in school quality across randomization blocks. *Note:* This figure provides coefficient estimates and standard errors for indicators of randomization block (Panel A) and interactions between randomization block and the following baseline student-level characteristics: female (Panel B), white (Panel C), age (Panel D), and English as primary language (Panel E). Randomization blocks with no variation in the baseline characteristic specified in the panel title are displayed as zero coefficients with no confidence interval. Coefficients from each panel are generated from separate regressions.

Table C.1. Balance tests of baseline characteristics across students with and without teacher observation data.

	With teacher data (1)	Without teacher data Unadjusted (2)	Without teacher teacher Adjusted (3)
Age	53.20	−0.16 (0.21)	−0.19 (0.21)
Female	0.52	−0.02 (0.03)	−0.00 (0.03)
White	0.53	−0.06 (0.03)	−0.06 (0.04)
Black	0.24	0.07* (0.03)	0.08* (0.04)
Hispanic	0.21	−0.02 (0.03)	−0.02 (0.03)
English primary language	0.78	−0.03 (0.03)	−0.03 (0.03)
<i>n</i> =	923	317	

Note: Columns (1) reports means for students with valid teacher observation scores. Column (2) reports unadjusted differences between students with and without teacher observation scores, which are estimated using OLS regressions of each characteristic on a binary indicator equal to 1 if students had missing teacher observation scores. Column (3) reports adjusted differences based on OLS regressions that include propensity score weights. Standard errors are reported in parenthesis and are clustered at the school level. **p* < .05 for two-tailed tests of significance.