ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 19-85

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

Francis A. Pearman, II
Stanford University

Matthew Springer
University of North Carolina, Chapel Hill

Mark Lipsey
Vanderbilt University

Mark Lachowicz
Vanderbilt University

Walker Swain
University of Georgia

Dale Farran
Vanderbilt University

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 that lasted through at least 3rd-grade. 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.

VERSION: July 2019

Suggested citation: Pearman, Francis, Matthew Springer, Mark Lipsey, Mark Lachowicz, Dale Farran, and Walker Swain. (2019). Teachers, Schools, and Pre-K Effect Persistence: An Examination of the Sustaining Environment Hypothesis. (EdWorkingPaper: 19-85). Retrieved from Annenberg Institute at Brown University: http://www.edworkingpapers.com/ai19-85

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

Francis A. Pearman, II
Stanford University

Matthew Springer
University of North Carolina, Chapel Hill

Mark Lipsey
Vanderbilt University

Mark Lachowicz
Vanderbilt University

Walker Swain
University of Georgia

Dale Farran
Vanderbilt University

Working Paper
May 2019

Corresponding author
Francis A. Pearman, II, Stanford Graduate School of Education, CERAS 524, 485 Lasuen Mall, Stanford, CA 94305 (apearman@stanford.edu)

Keywords: PreK, academic achievement, sustaining environments, teacher quality, school quality

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 that lasted through at least 3rd-grade. 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.

Introduction

Short-term boosts in children's language, literacy, and math skills that result from attending prekindergarten classrooms (PreK) often diminish soon after preschool ends (Yoshikawa, Weiland, & Brooks-Gunn, 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, Gormley, & Adelstein, 2015; Lipsey, Farran, & Durkin, 2018; Puma et al., 2012; Puma, Bell, Cook, & Heid, 2010). For example, a recent meta-analysis of existing preschool effectiveness research, which was based on over 60 evaluations of high-quality 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 percent in the year following the intervention, and again by 50 percent one to two years later (Bailey, Duncan, Odgers, & Yu, 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 are potentially maintained over time. This literature has focused, in part, on the skills and capacities learned during the PreK year and their consequences for later learning (Heckman, 2006). However, 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). Prior research has examined whether PreK effects are maintained if children go on to have high quality teachers during elementary school (Swain, Springer, & Hofer, 2015), or whether the elementary schools they later attend are of high quality themselves (Currie & Thomas, 2006; Lee & Loeb, 2008). Overall, this research has indicated that both teacher quality and school quality may play a role in determining the extent of PreK effect persistence, 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, combined with detailed teacher and school information to examine whether the persistence of PreK effects is influenced by subsequent teacher and school quality. Though economically disadvantaged children who attend public PreK programs like TN-VPK often attend lower quality elementary schools (Currie & Thomas, 2006), there is evidence that some children may nevertheless have exposure to highly effective teachers in such schools (Sass, Hannaway, Xu, Figlio, & Feng, 2012). Likewise, even if children are able to attend higher quality schools after PreK, children may not necessarily have good teachers in these schools. 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 enter.

1. Conceptual Framework and Prior Research

Whereas some early childhood research explains PreK fadeout in terms of the nature and extent of the learning that does (or does not) take 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). That is, 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 disproportionally on the learning needs of struggling children who were less likely to have attended 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 Curry and Thomas (2002), both studying fadeout in the Head Start Impact Study, 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).

More direct evidence in favor of the sustaining environments perspective is found in several recent studies. In Tennessee, Swain, Springer, and Hofer (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 year subsequent to PreK was associated with more persistent PreK effects. Moreover, the relationship 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. Ansari and Pianta (2018) used data from National Institute of Child Health and Human Development Study of Early Child Care and arrived at similar conclusions. In particular, they 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 mediated 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 that found that kindergarten teachers often teach material related to the knowledge children already possess at kindergarten entry (Bassok, Latham, & Rorem, 2016; Claessens, Engel, & Curran, 2013; Engel, Claessens, & Finch, 2012; Gervasoni & Perry, 2015; Magnuson, Ruhm, & Waldfogel, 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 absence of evidence has been noted at both the school and classroom level. 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 of Kindergarten (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. Likewise, Bassok et al. (2015), using ECLS-K data, 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) a common pattern of PreK effects on achievement-related outcomes is one of short-term beneficial impacts that fade over time, often quickly; and (b) the literature is largely unsettled regarding the systematic components responsible for pre-k effect persistence (or fadeout). Our work builds from this prior literature by leveraging a recently conducted state-wide PreK experiment in Tennessee, TN-VPK, and examining the extent to which the effects of PreK on 3rd grade achievement test performance is moderated by the quality of the teachers and schools that children subsequently experience after preschool. Measuring academic performance in 3rd grade is ideal for the current study because prior TN-VPK research found that positive pre-k effects at kindergarten entry faded and in some instances reversed by 3rd grade.

2. 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, which houses student test score data linked with specific teachers and schools throughout our study period.

2.1. Analytic Sample

The full TN-VPK experiment included a final analytic sample of 2,990 students spread across two cohorts who were offered admission to the state PreK program from randomly-ordered applicant lists at oversubscribed sites (see Lipsey et al., 2018 for details). The first cohort entered PreK in 2009-2010, and the second cohort entered PreK in 2010-2011. In this study, we focus on the second cohort of children. 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 (2011-12 school year). Thus, we lack kindergarten teacher evaluation records for students in the first experimental cohort recruited in 2009 and 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% percent of students had missing test scores in third grade. 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 robustness checks that multiply impute missing values on all covariates and outcomes (see Figures A.1-A.4 in the Appendix). It should 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.

Additionally, because the primary focus of this study is the relationship 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. Of the remaining 806 students, 507 were part of the TNVPK treatment group and 299 were part of the control condition. Of the treatment status students, 84% attended VPK for a minimum of 20 days and 16% did not. Of the 299 control students, 75% did not attend VPK and 25% crossed over, i.e., they attended PreK for a minimum of 20 days at a TN-VPK site even though they were not randomly lotteried into the treatment condition. Finally, we treated the 7 children who attended VPK for less than 20 days as non-participants. 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. We refer 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.

2.2. Primary Measures of Interest

Achievement. The primary 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 raw scale scores from the 3rd grade mathematics and English Language Arts (ELA) examinations.[1] 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.

Teacher quality. The primary 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.[2]Because students in grades K through 2 do not take standardized 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 calculate 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

---

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

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

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 denotes 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 less than 5 percent of the analytic sample was taught by a highly effective teachers 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 effect 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 two 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.

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 primary measure 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 their demographics and prior achievement history. 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 percent 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 percentage points below the expected growth rate, and, on the high end, performance growth was 7.9 percentage points above the expected growth rate, with the performance growth at the average school exceeding the expected growth rate by 1.2 percentage points. To aid interpretation, we standardize this measure to have a mean of zero and standard deviation of one.

3. Analytic Strategy

Our overarching research question is: Do later teachers and schools provide sustaining environments for PreK effects? 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 assess later (see Table 1). 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.) We begin by conducting a moderated multiple regression analysis that takes the following form:

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

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; 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 individual residual clustered at the school level. In addition to the vector of control variables, specified by z'_i, all models use propensity score based weighting to adjust for any differences in baseline characteristics across PreK participants and non-participants (Austin, 2011).[3] The coefficients of interest are the two- and three-way interactions amongst VPK participation, teacher effectiveness, and school quality.[4]

4. Results

4.1. 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 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

[3] Propensity score weights were based on inverse probability of treatment weighting (IPTW), which 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. (See Austin [2011] for a detailed description of the application of inverse probability of treatment weights).

[4] Another analytic strategy would be to use a two-staged least square (2SLS) approach. In particular, we could 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, Dekkers, Vandenbroucke, & Le Cessie, 2014), which is exacerbated in the current study because of the inclusion of three-way interactions. Indeed, a 2SLS approach produced unstable and implausible estimates that are not reported.

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 propensity score weights, i.e., there were no statistically significant differences on these and other baseline characteristics after the inclusion of propensity score 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 on value added scores from those of children who did not attend VPK.

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 percent of students in each condition were never enrolled in a highly effective teacher's classroom from K to 3rd grade, which contrasts with the fact that only 3% of Tennessee students overall 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 around 1 in 5 students in the analytic sample had such exposure.

Table 2 turns attention to the joint distribution of teacher and school quality across the analytic sample. For illustrative purposes, 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 at least 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 that over 20% of children were never exposed to a highly effective teacher between kindergarten and third grade as noted in the previous paragraph, 3 out of 4 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.

4.2. 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 1 standard deviation increase in value added. Column 2 in each table includes the same covariates as Columns 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 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. 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 general insight about the unique explanatory power of VPK, on average, versus having a high quality elementary school experience in terms of teacher and school quality, the current study is most interested whether PreK effects were more or less likely to persist depending on the quality of children's subsequent learning environments.

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 explained differences in 3rd grade ELA or math achievement between VPK participants and non-participants.

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. We find that 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 imprecisely estimated and fell short of statistical significance ($p = 0.154$ and $p = 0.075$, respectively). This imprecision may have been 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 answers 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 went on to attend high quality schools with highly effective teachers (ELA: $\beta = 0.15$, $p = 0.040$; math: $\beta = 0.17$, $p = 0.016$).

To provide some intuition for what these three-way interactions mean, Figures 1 and 2 plot the marginal effect of VPK on 3rd grade math and ELA achievement, respectively, across levels of school quality (-2 SD to +2 SD) for students with zero, one, two, and three (or four) highly effective teachers. The solid line refers to the point estimate and the dotted lines refer to the 95 percent confidence interval. Point estimates are considered statistically significant wherever the confidence interval excludes zero.

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 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 1 standard deviation above the mean ($\beta = 0.27$, p =

0.022). 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 = 0.003). 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 1 standard deviation above the mean (β = 0.24, p = 0.030), a difference that similarly grows if the number of highly effective teachers and the school value-added increases.

Notably, Figures 1 and 2 also 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 poor schools with good teachers were 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 score 1 standard deviation below the mean (math: β = -0.30, p = 0.003; ELA: β = -0.26, p = 0.028; ELA).

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 outperformed non-participants by the largest margin if these children attended high quality schools and had highly effective teacher *in the two years immediately following VPK* (β = 0.38, p = 0.072). No evidence was found that having highly effective teachers 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 two years preceding statewide assessments* (β = 0.57, p = 0.012). No evidence was found that having highly effective teachers in kindergarten and 1st grade moderated the observed differences in 3rd 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. 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 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 emerge 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.

4.3. Robustness Checks

Figures A.1-A.4 in the Appendix illustrate the robustness of our primary results to alternative approaches to handling missing data. These figures plot the marginal effect of VPK on

3rd grade achievement across levels of school quality for students with different numbers of highly effective teachers based on datasets wherein missing values were multiply imputed. Figures A.1 and A.2 provide results for 3rd grade ELA and math achievement, respectively, from a dataset wherein missing values on baseline characteristics and teacher and school quality measures were imputed. Figures A.3 and A.4 provide results for 3rd grade ELA and math achievement, respectively, based on a dataset wherein missing values on all variables, including outcomes measures, were imputed. Overall, Figures A.1-A.4 provide evidence that substantive conclusions are robust to how missing data were handle. As illustrated in Figures A.1-A.4, we observe statistically meaningful differences in math and ELA in favor of VPK participants when children have at least two highly effective teachers and attend schools with value added scores at least 1 SD above the mean, regardless of whether we imputed baseline covariates, teacher and school quality measures, or children's achievement in 3rd grade.

5. 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., 2012, 2010). In light of these findings, scholars have shifted increasing focus to understanding the source of this apparent test score fade out. 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 with highly effective teachers.

These results have a number of implications for theory and research about early childhood education. First, this research provides new perspective on the sustaining environments thesis. In particular, 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. Indeed, this finding may reconcile some of the previous debates about the role subsequent learning environments play in the persistence of PreK effects. In particular, prior research has been mixed regarding whether teachers, classrooms, and 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 perspective has focused primarily on the moderating capacity of, say, 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 were 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.

It is also notable that this study found that non-participants outperformed VPK participants in 3rd grade in both math and ELA if these children went on to attend low quality schools with at least two highly effective teachers. This counterintuitive pattern could have arisen if high quality teachers in low quality schools were adept at tailoring instruction toward higher-needs students. This targeted instruction could function to disadvantage VPK participants in terms of their potential learning gains. If this disadvantage fosters discouragement or disengagement on the part of VPK participants, one could reasonably expect a reversal of effects over time wherein non-participants go on to outperform VPK participants on standardized assessments of achievement in later grades.

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, Meador, Christopher, Nesbitt, & Bilbrey, 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 (defined as schools with gain scores at least 1 standard deviation above the mean) 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 TNVPK 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, given that our findings about the overexposure of children to low-quality schooling environments after PreK align with those from previous research (e.g., Lee & Loeb, 1995; Curry & Thomas, 2002), our findings should be 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, Swain, and Rodriguez, 2016; Swain, Rodriguez, and Springer, 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.

6. Limitations

Although this study extends prior research about the persistence (or fadeout) of PreK effects by highlighting a key interaction between teacher and school quality, this study is not without several limitations. First, given that we were unable to randomly assign students to schools and teachers of varying quality, 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 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, we found no evidence of observed differences in teacher and school quality between VPK participants and non-participants). Second, our measure of teacher effectiveness, which was based on classroom observation ratings, is admittedly a crude approximation of a teacher's ability to promote test score growth because correlations between classroom observation ratings, on which our measure was based, and test-score value-added measures of teacher effectiveness are not well-established. Third, this study was based on a subsample of the larger TNVPK study for whom teacher observation score were available in kindergarten. Finally, children must qualify for free- and reduced-price lunch services to enroll in TNVPK, thus the results of our study may not generalize to more socioeconomically-advantaged students.

7. 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 fadeout, 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. Implementing scaled up PreK programs without attending to the subsequent school experiences of the participants may not produce the long term positive effects in reducing the achievement gap that many advocates hope for.

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), 1934–5747. 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*. https://doi.org/10.1177/2332858415616358

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*. https://doi.org/10.1016/j.jclinepi.2014.05.019

Claessens, A., Engel, M., & Curran, F. C. (2013). Academic Content, Student Learning, and the Persistence of Preschool Effects. *American Educational Research Journal*. https://doi.org/10.3102/0002831213513634

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. (2006). School Quality and the Longer-Term Effects of Head Start. *The Journal of Human Resources*. https://doi.org/10.2307/146372

Danielson, C. (2013). *The framework for teaching: Evaluation instrument.* Retrieved from http://www.danielsongroup.org/userfiles/files/downloads/%0A2013EvaluationInstrument.pdf.

Engel, M., Claessens, A., & Finch, M. A. (2012). Teaching Students What They Already Know? The (Mis)Alignment Between Mathematics Instructional Content and Student Knowledge in Kindergarten. *Educational Evaluation and Policy Analysis*. https://doi.org/10.3102/0162373712461850

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. C., Meador, D., Christopher, C., Nesbitt, K. T., & Bilbrey, 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

Farran, D., Christopher, C., Pion, G., Henley, K., Low, T., Norvell, J., … Tunstel, A. (2018). *Kindergarten follow-up report: Full report*. Nashville, TN.

Gervasoni, A., & Perry, B. (2015). Children's Mathematical Knowledge Prior to Starting School and Implications for Transition. In *Mathematics and Transition to School. Early Mathematics Learning and Development* (pp. 47–64). Singapore: Springer.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*, 1900–1902. https://doi.org/10.1126/science.1130121

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 Preschool Fadeout? *Journal of Research on Educational Effectiveness*. https://doi.org/10.1080/19345747.2018.1441347

Lee, V. E., & Loeb, S. (2008). Where Do Head Start Attendees End Up? One Reason Why Preschool Effects Fade Out. *Educational Evaluation and Policy Analysis*. 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*. 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. Retrieved from http://files.eric.ed.gov/fulltext/ED507845.pdf

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., … Downer, J. (2012). *Third Grade Follow-up to the Head Start Impact Study: Final Report*. Washington, DC. Retrieved from http://files.eric.ed.gov/fulltext/ED539264.pdf

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., and Rodriguez, L.A. (2016). Effective Teacher Retention Bonuses: Evidence from Tennessee. Educational Evaluation and Policy Analysis, 38(2), 199-221.

Swain, W. A., Springer, M. G., & Hofer, K. G. (2015). Early Grade Teacher Effectiveness and Pre-K Effect Persistence: Evidence From Tennessee. *AERA Open*. https://doi.org/10.1177/2332858415612751

Swain, W.A., Rodriguez, L.A., and Springer, M.G. (2019). Selective Retention Boinuses for Highly Effective Teachers in High Poverty Schools: Evidence from Tennessee. *Economic of Education Review*, 68, 148-160.

Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (2016). When does preschool matter? *Future of Children*.

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	0.02
			(0.24)	(0.24)
Female	0.49	0.51	0.04	0.05
			(0.03)	(0.04)
White	0.66	0.45	0.09*	-0.03
			(0.04)	(0.04)
Black	0.22	0.23	0.04	0.04
			(0.03)	(0.03)
Hispanic	0.10	0.31	-0.15***	-0.02
			(0.04)	(0.04)
English Primary Language	0.89	0.68	0.14***	-0.00
			(0.04)	(0.04)
Teacher Characteristics				
% 0 HE Teachers	0.03	0.22	0.02	0.02
			(0.04)	(0.04)
% 1 HE Teachers	0.16	0.33	-0.04	-0.03
			(0.04)	(0.04)
% 2 HE Teachers	0.30	0.27	-0.02	-0.02
			(0.04)	(0.04)
% 3 or 4 HE Teachers	0.51	0.17	0.04	0.03
			(0.04)	(0.04)
School Characteristics				
Value-Added	1.40	0.98	0.33	0.23
			(0.20)	(0.20)
n =	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. Columns 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<.01, ***p<.001 for two-tailed tests of significance.

TABLE 2: JOINT DISTRIBUTION OF TEACHER AND SCHOOL QUALITY

	n *row* *cell*	School Quality Low	 Moderate	 High	total
Number of Highly Effective Teachers	0	46 0.24 0.06	123 0.65 0.15	19 0.10 0.02	188
	1	78 0.31 0.10	149 0.60 0.18	21 0.08 0.03	248
	2	45 0.21 0.06	134 0.64 0.17	32 0.15 0.04	211
	3 or 4	30 0.19 0.04	85 0.53 0.11	44 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 at least 1 standard deviation above the mean gain score in the analytic sample. 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.

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	-0.01	0.01	0.00
	(0.07)	(0.13)	(0.07)	(0.14)
# HE Teachers	0.04	0.03	0.16	0.04
	(0.04)	(0.07)	(0.16)	(0.06)
School Quality	-0.01	-0.01	-0.07	0.07
	(0.04)	(0.04)	(0.07)	(0.10)
HE Teachers* School Quality				-0.14*
				(0.06)
VPK* HE Teachers		0.01		-0.01
		(0.08)		(0.07)
VPK* School Quality			0.11	-0.05
			(0.08)	(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	-0.09	-0.02	-0.08
	(0.07)	(0.13)	(0.07)	(0.12)
# HE Teachers	0.05	0.02	0.11	0.03
	(0.05)	(0.07)	(0.18)	(0.06)
School Quality	0.04	0.04	-0.03	0.16†
	(0.05)	(0.05)	(0.07)	(0.09)
HE Teachers* School Quality				-0.17***
				(0.05)
VPK* HE Teachers		0.05		0.03
		(0.08)		(0.07)
VPK* School Quality			0.13†	-0.05
			(0.07)	(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<.01, ***p<.001 for two-tailed tests of significance.

TABLE 4: PREK EFFECT MODERATION BY EXPOSURE TO A HIGH QUALITY TEACHER DURING KINDERGARTEN AND 1ST GRADE OR DURING 2ND AND 3RD GRADE

	First two years with HE teacher		Last two years with HE teacher	
	(1)	(2)	(3)	(4)
Panel A. 3rd Grade ELA Achievement				
VPK	0.02	0.03	0.02	0.03
	(0.08)	(0.08)	(0.08)	(0.08)
Exposure to HE	-0.05	-0.14	0.21	0.26+
Teacher Both Years	(0.25)	(0.20)	(0.15)	(0.15)
School Quality	-0.01	-0.01	-0.02	-0.07
	(0.04)	(0.07)	(0.04)	(0.07)
HE Teachers* School Quality		-0.50**		-0.09
		(0.19)		(0.19)
VPK* HE Teachers	-0.05	0.04	-0.10	-0.17
	(0.26)	(0.21)	(0.22)	(0.22)
VPK* School Quality		0.09		0.11
		(0.09)		(0.09)
VPK* HE Teachers* School Quality		0.38†		0.16
		(0.21)		(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	0.01	-0.04	-0.04
	(0.08)	(0.08)	(0.08)	(0.08)
Exposure to HE	-0.06	-0.13	0.24	0.43*
Teacher Both Years	(0.23)	(0.21)	(0.18)	(0.18)
School Quality	0.05	0.02	0.03	0.01
	(0.05)	(0.07)	(0.05)	(0.07)
HE Teachers* School Quality		-0.38*		-0.61**
		(0.18)		(0.19)
VPK* HE Teachers	-0.05	0.03	0.15	-0.05
	(0.24)	(0.22)	(0.25)	(0.24)
VPK* School Quality		0.13		0.10
		(0.08)		(0.07)
VPK* HE Teachers* School Quality		0.23		0.57*
		(0.23)		(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, ***p<.001 for two-tailed tests of significance.

Figure 1:

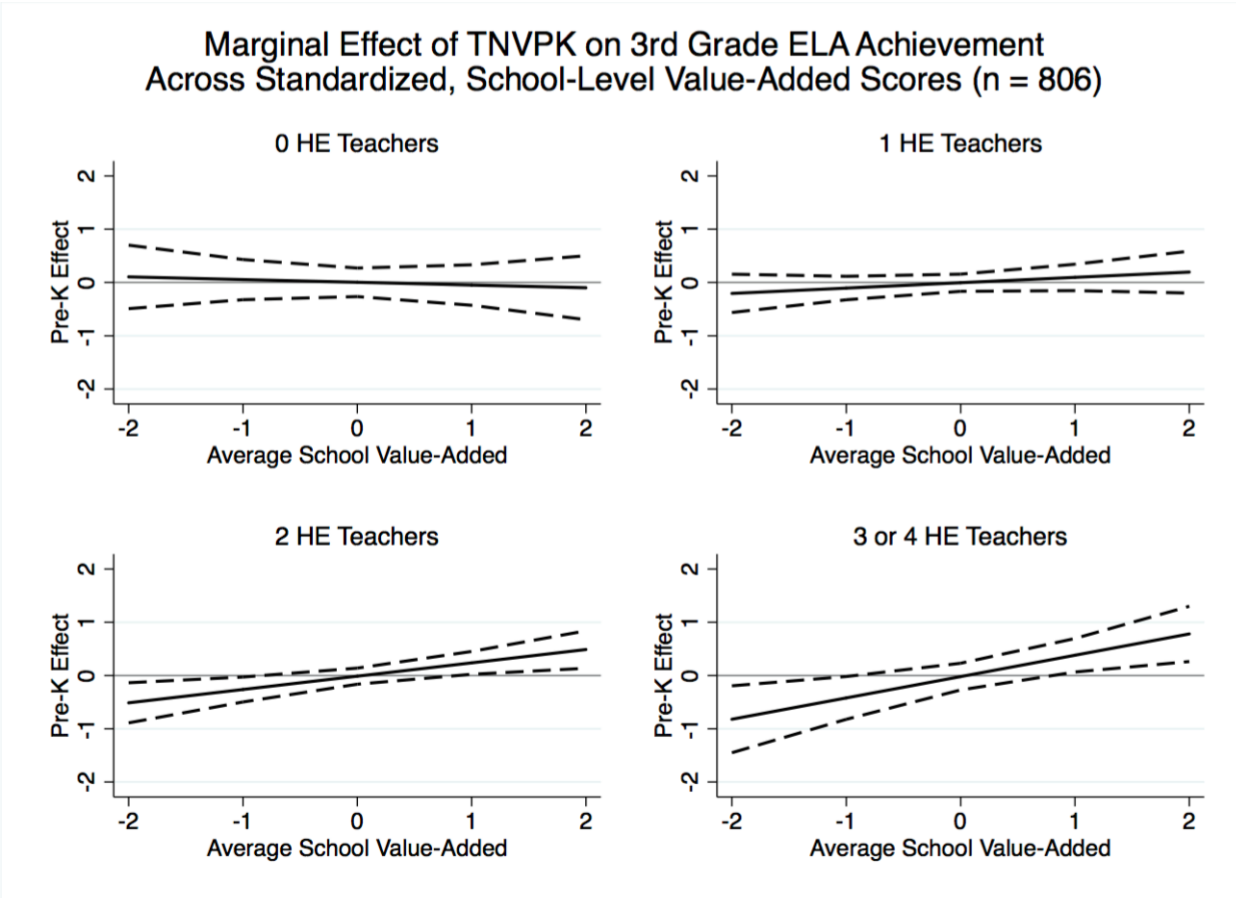
Marginal Effect of TNVPK on 3rd Grade ELA Achievement
Across Standardized, School-Level Value-Added Scores (n = 806)
0 HE Teachers
1 HE Teachers
2 HE Teachers
3 or 4 HE Teachers
Pre-K Effect
Average School Value-Added
-2
-1
0
1
2

Figure 2:

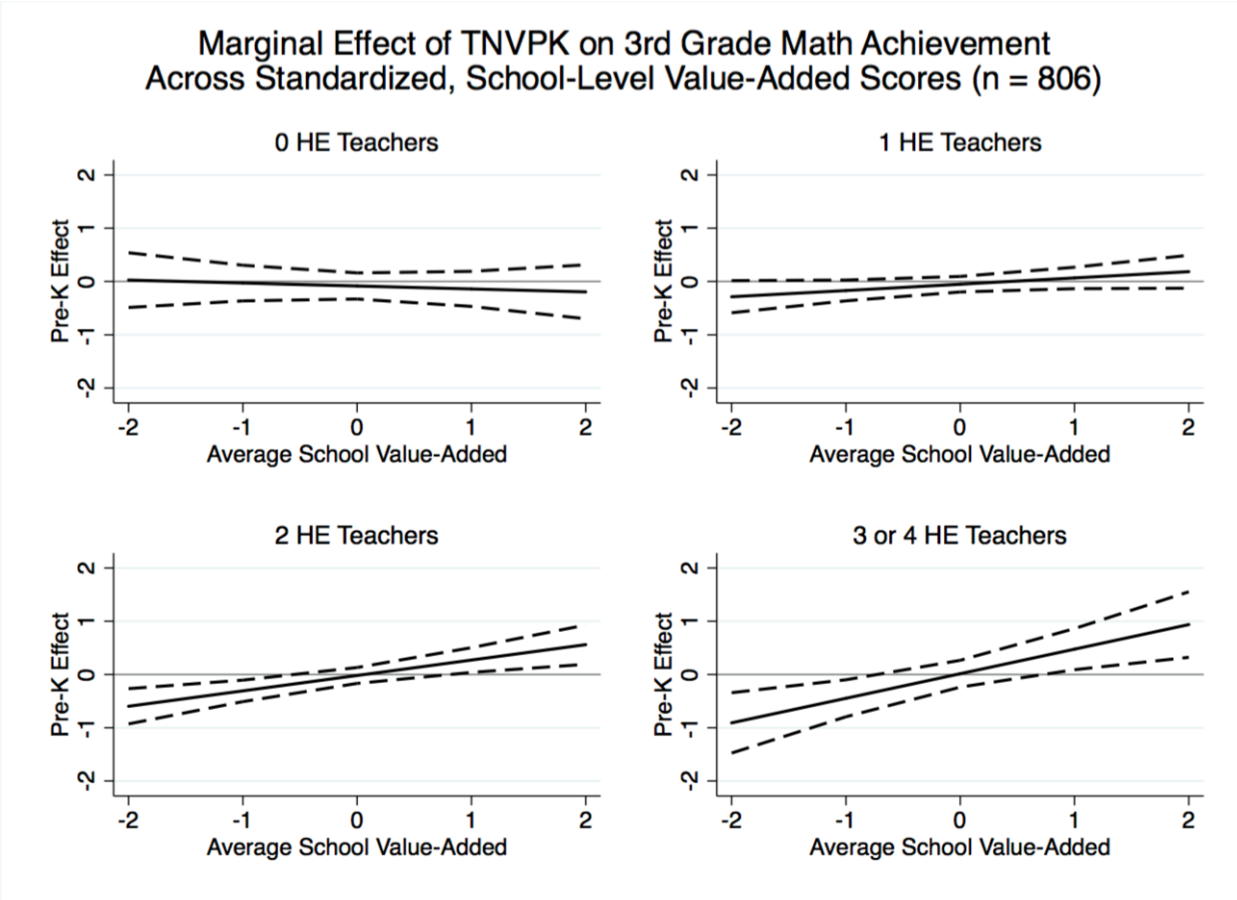
Marginal Effect of TNVPK on 3rd Grade Math Achievement
Across Standardized, School-Level Value-Added Scores (n = 806)
0 HE Teachers
1 HE Teachers
2 HE Teachers
3 or 4 HE Teachers
Pre-K Effect
Average School Value-Added
-2
-1
0
1
2

Figure 3:

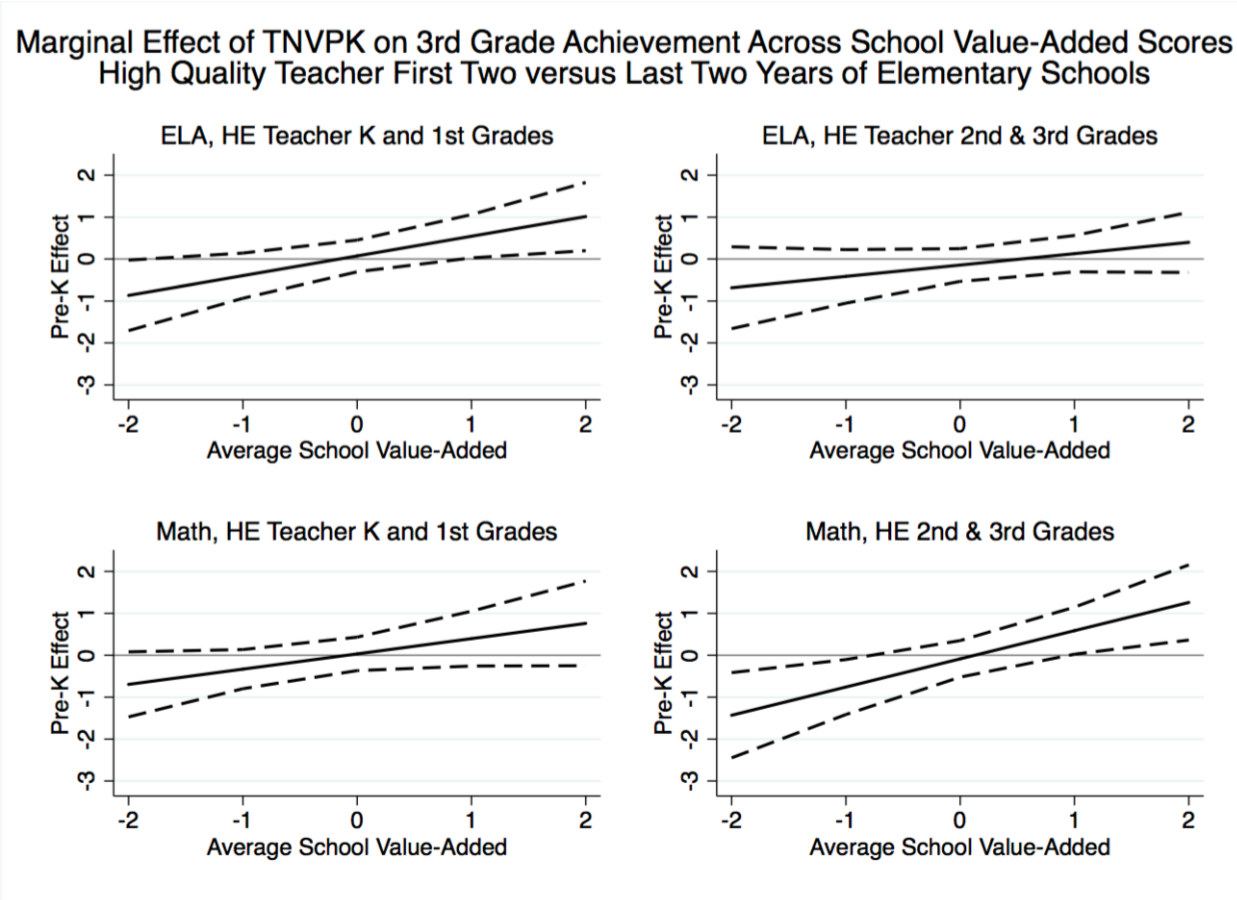
Marginal Effect of TNVPK on 3rd Grade Achievement Across School Value-Added Scores
High Quality Teacher First Two versus Last Two Years of Elementary Schools
ELA, HE Teacher K and 1st Grades
ELA, HE Teacher 2nd & 3rd Grades
Math, HE Teacher K and 1st Grades
Math, HE 2nd & 3rd Grades
Pre-K Effect
Average School Value-Added
-3
-2
-1
0
1
2

Figure A.1:

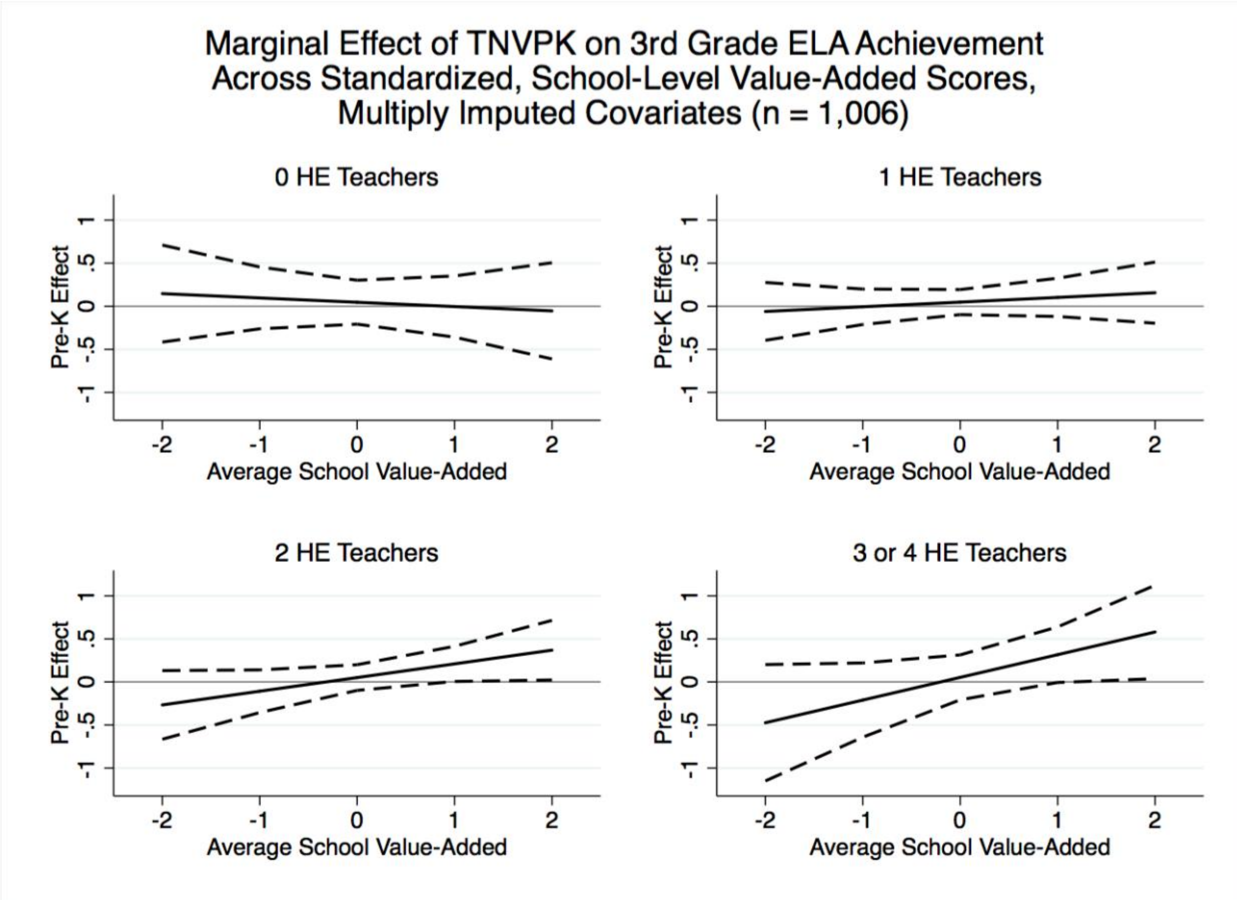
Marginal Effect of TNVPK on 3rd Grade ELA Achievement
Across Standardized, School-Level Value-Added Scores,
Multiply Imputed Covariates (n = 1,006)
0 HE Teachers
1 HE Teachers
2 HE Teachers
3 or 4 HE Teachers
Pre-K Effect
Average School Value-Added
-1
-.5
0
.5
1
-2
-1
0
1
2

Figure A.2:

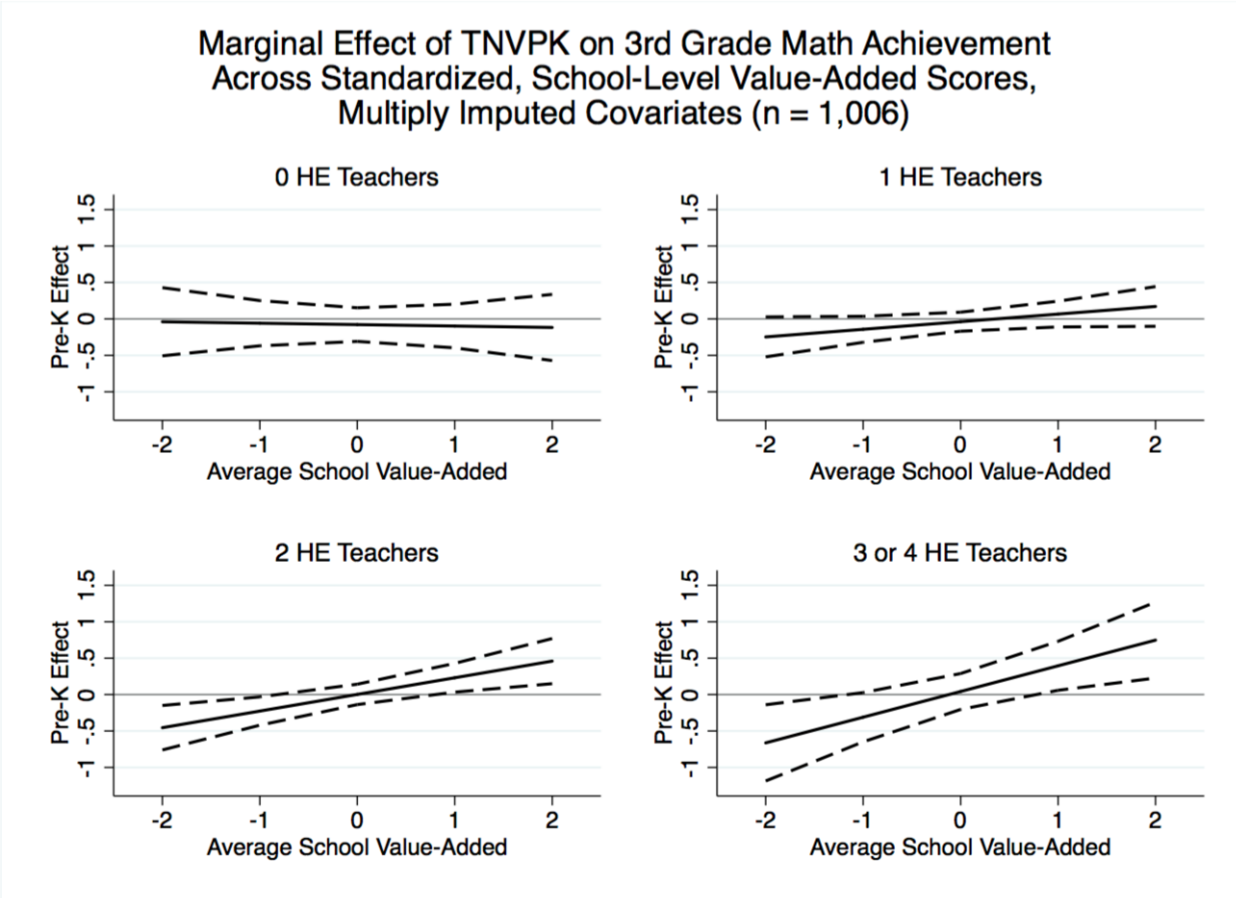
Marginal Effect of TNVPK on 3rd Grade Math Achievement
Across Standardized, School-Level Value-Added Scores,
Multiply Imputed Covariates (n = 1,006)
0 HE Teachers
1 HE Teachers
2 HE Teachers
3 or 4 HE Teachers
Pre-K Effect
-1 -.5 0 .5 1 1.5
-2 -1 0 1 2
Average School Value-Added

Figure A.3:

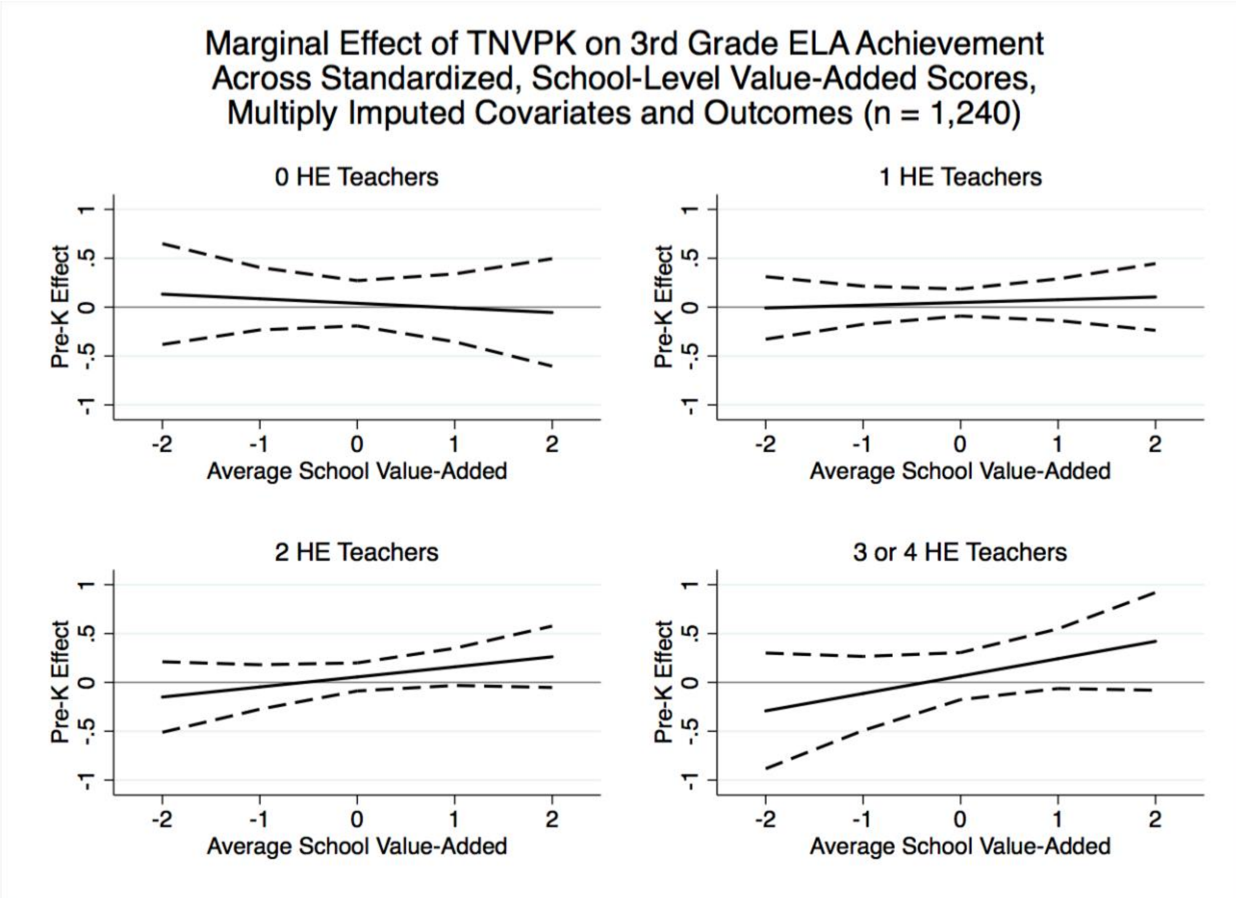
Marginal Effect of TNVPK on 3rd Grade ELA Achievement
Across Standardized, School-Level Value-Added Scores,
Multiply Imputed Covariates and Outcomes (n = 1,240)
0 HE Teachers
1 HE Teachers
2 HE Teachers
3 or 4 HE Teachers
Pre-K Effect
Average School Value-Added
-1
-.5
0
.5
1
-2
-1
0
1
2

Figure A.4:

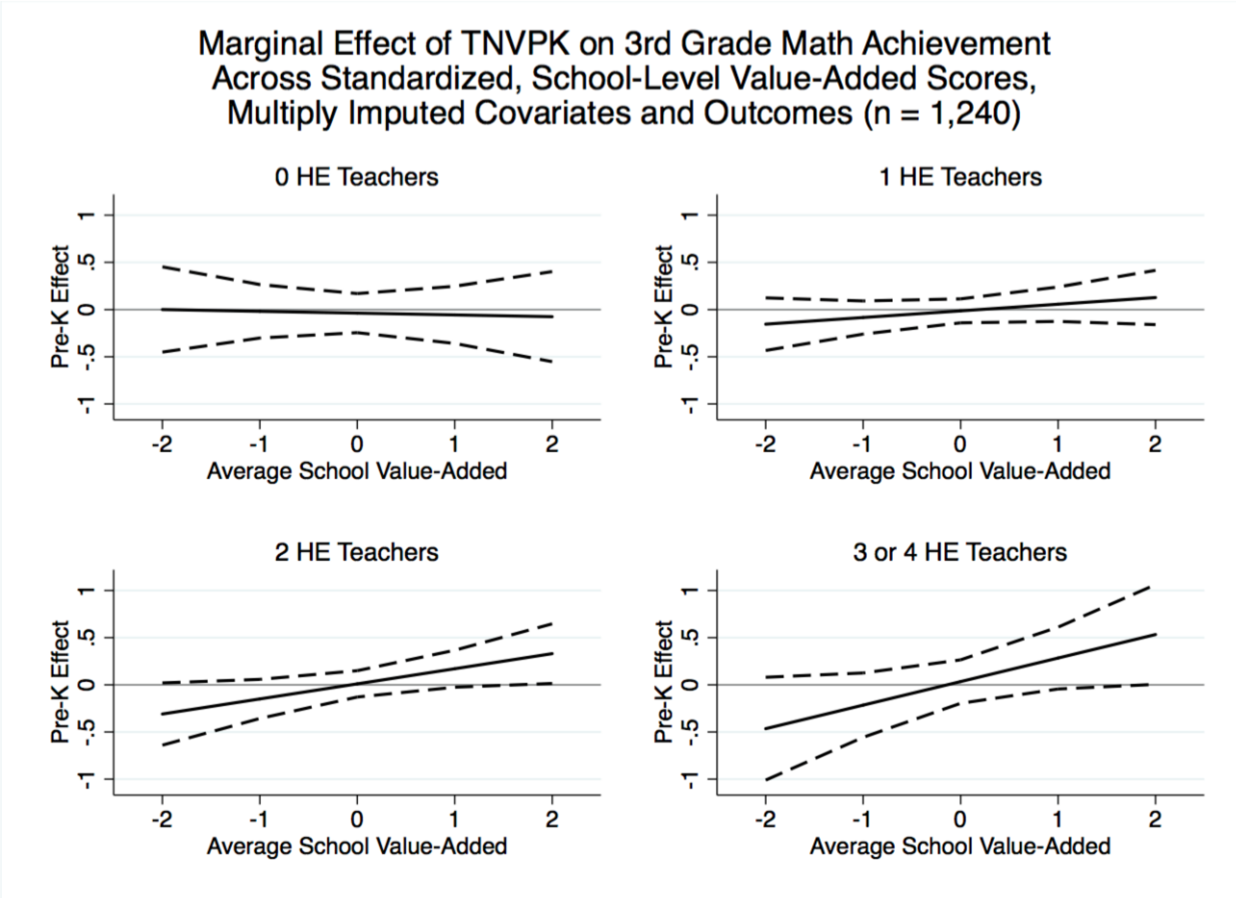
Marginal Effect of TNVPK on 3rd Grade Math Achievement
Across Standardized, School-Level Value-Added Scores,
Multiply Imputed Covariates and Outcomes (n = 1,240)
0 HE Teachers
1 HE Teachers
2 HE Teachers
3 or 4 HE Teachers
Pre-K Effect
-1
-.5
0
.5
1
-2
-1
0
1
2
Average School Value-Added

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)
n =	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. †p<.10, *p<.05, **p<.01, ***p<.001 for two-tailed tests of significance.