IRVINE NETWORK ON INTERVENTIONS IN DEVELOPMENT

# Do High Quality Kindergarten and First Grade Classrooms Mitigate Preschool Fadeout?

Working Paper
This version: March 2016

Jade M. Jenkins[1]
Tyler W. Watts[1]
Katherine Magnuson[2]
Elizabeth Gershoff[3]
Douglas Clements[4]
Julie Sarama[4]
Greg J. Duncan[1]

[1] School of Education, University of California, Irvine
[2] School of Social Work, University of Wisconsin, Madison
[3] Department of Human Development and Family Sciences, University of Texas at Austin
[4] College of Education, University of Denver

## Abstract

Prior research shows that short-term treatment effects from preschool may disappear, but little research has considered which environmental conditions might sustain the academic advantages from preschool into elementary school. Using two random-assignment studies, we investigate whether two features of the elementary school environment help to sustain preschool academic gains: 1) exposure to advanced and high-quality instruction in kindergarten and first grade; 2) professional supports to coordinate curricular instruction and transition. Across both studies, our measures of instruction did not moderate fadeout from preschool. However, results indicated that targeted teacher professional supports substantially mitigated fadeout between kindergarten and first grade. Future research should investigate whether aligned preschool-elementary school curricular approaches can sustain the benefits of preschool programs for low-income children.

## Introduction

Many studies document the benefits of early childhood education and formal preschool experiences on children's school readiness, with low-income and otherwise disadvantaged children benefitting significantly from these programs (Barnett, 2011; Camilli, Vargas, Ryan, & Barnett, 2010; Duncan & Magnuson, 2013; Reynolds, Temple, & Ou, 2010). However, some studies also show that these academic benefits do not persist through the early elementary school years; after a year or two of schooling most differences between preschoolers and the comparison groups have converged (Barnett, 1995; Currie, 2001; Puma, Bell, Cook, & Heid, 2010). The convergence, or "catch-up", in performance on tests of a variety of academic and cognitive skills is often termed "fadeout" because the preschool impact observed at program completion is undetectable at follow-up points. Indeed, most programmatic prevention and intervention efforts, academic and otherwise, demonstrate such a pattern of diminished program impacts over time (Bailey, Duncan, Odgers, & Yu, 2015). This is discouraging to policy makers and practitioners because promoting later school achievement is one of the core motivations for funding public preschool programs.

The magnitude of this fadeout has been quantified across studies. A recent meta-analytic study estimated a geometric decline in preschool effects on cognitive outcomes, with end-of-treatment impacts declining by half in the first year following the end of treatment and then by half again over the next two years (Li et al., 2016), confirming prior estimates (Aos, Miller, & Drake, 2006). Despite these patterns, fadeout in the early school years has not received much theoretical attention. One longstanding explanation is that low-income preschool graduates enter schools that are less safe, more disadvantaged, and that generally have poorer instructional quality than their higher income peers (Lee & Loeb, 1995; Stipek, 2004). An elaboration of this argument targets the content, focus, and pace of instruction. Specifically, that elementary school curricula and teaching focus on content that children already learned during preschool (Engel, Claessens, & Finch, 2013; Gervasoni & Perry, 2015), and that teachers do not differentiate instruction, which will reduce the likelihood of continued academic growth relative to the expected growth of comparison children who did not learn the content in preschool (Bennett, Desforges, Cockburn, & Wilkinson, 1984; Engel et al., 2013).

In this study, we investigated whether preschool participants' instructional experiences in elementary school moderate the persistence of preschool's effects using two different ways of conceptualizing early school instruction. We used two experimental studies of preschool interventions that included follow-up data on children's elementary school environments to examine whether the quality of instructional content or whether professional development supports for early grade teachers moderate the impacts of two well-known early education programs on children's cognitive skills: Head Start and Building Blocks Math Curriculum.

First using the Head Start Impact Study, we considered whether advanced instruction in kindergarten and first grade would moderate the extent to which Head Start benefits persist over time. We hypothesized that children who attended Head Start and then experienced more rigorous instructional content may have longer lasting program benefits compared with children who did not attend Head Start or who experienced less rigorous instruction. Second, we utilized an evaluation of Building Blocks, which included a treatment condition involving professional support during elementary school. Specifically, preschool teachers interacted with their kindergarten and first grade teacher counterparts to develop a smooth transition and alignment for classroom instruction from one grade to the next. We hypothesized that classrooms with these professional supports would produce longer lasting program benefits compared with classrooms that did not have such supports. Furthermore, we expected that observed classroom quality would partially mediate this benefit.

## Background

The U.S. has seen an unprecedented expansion of early childhood education (ECE) programs over the past 40 years, primarily in the form of state pre-kindergarten (pre-k) programs, an expansion made possible because policymaker and educators now view early childhood as a particularly opportune time for investment (Jenkins, 2014). Research from neuroscience, education, psychology, and economics compose much of the foundational literature for the importance of ECE programs and the role of early intervention increasing the chances of success across several key domains of adjustment later in life, particularly for children in poverty (Barnett, 2011; Duncan & Magnuson, 2013; Fox, Levitt, & Nelson, 2010; Hackman & Farah, 2009; Shonkoff & Meisels, 2000).

Early support for ECE programs came from the random assignment evaluations of two well-known interventions begun in the 1960s and 1970s: Perry Preschool and Abecedarian. These evaluations showed that intensive (and expensive) early education programs for disadvantaged children can improve cognitive and language abilities by .75 to 1.5 standard deviations (SD) at the end of treatment. In the long-term, Abecedarian participants had lower levels of grade retention and high school dropout, reduced placement in special education, and higher rates of attending a 4-year college and having a full-time job as a result of treatment (Barnett & Masse, 2007; Campbell et al., 2012; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Campbell et al., 2008). For Perry Preschool participants, the program produced lasting effects through age forty by increasing employment and earnings and by reducing the likelihood of adult poverty rates and arrest (Belfield, Nores, Barnett, & Schweinhart, 2006; Campbell et al., 2014; Schweinhart, 2005).

Yet looking more broadly at the larger literature of ECE studies provides a more complicated pattern of effects. First, what constitutes an ECE intervention varies from full-day educational programs to short, discrete curricular interventions focused on math or reading skills. Secondly, most studies show evidence of initial impacts on cognitive or early academic outcomes, but also suggest that these impacts diminish significantly over time with the comparison children catching up to preschool attendees. Indeed, evidence from a recent meta-analysis finds that preschool effects fade during the early elementary school years, such that the average effect is reduced to zero 10 years after program completion (Leak et al., 2013). The pattern of program impact fadeout for achievement and cognitive outcomes has been especially noted in more recent studies of large public programs including Head Start and several state pre-k programs (e.g, TN pre-k; Lipsey, Farran, & Hofer, 2015).

The Head Start Impact Study experiment (HSIS) was the first nationally representative random assignment evaluation of the federal preschool program for low-income children. In 2002, two cohorts of children were randomly assigned to receive Head Start services at sites across the country. The end-of-program-year effect sizes averaged .2 SD for both the age-3 and age-4 cohorts on early language and literacy skills, and a .15 SD effect size on early math skills for age-3 cohort participants (Puma et al., 2010). These results were about 50 percent larger when non-compliance with treatment assignment was taken into account (Ludwig & Phillips, 2008). Nonetheless, the modest short-term gains from Head Start were entirely gone by the study's follow-up periods in elementary school (kindergarten, first, and third grade; Puma et al., 2012).

Several evaluations of public preschool programs have also suggested a quick reduction in preschool program impacts after program completion. An experimental evaluation of the one-year state pre-k program in Tennessee found that the short-term gains in achievement made during the preschool year disappeared completely in elementary school (Lipsey et al., 2015; Lipsey, Hofer, Dong, Farran, & Bilbrey, 2013). Similarly, a recent non-experimental evaluation (using propensity scores) of Oklahoma's pre-k program indicated that the large end-of-treatment effects disappeared by third-grade, with effects remaining only for boys' math skills (Hillm, Gormley, & Adelstein, 2015).

Using other methods, some studies have found evidence of persistent preschool effects on a wider set of outcomes than just achievement skills. Exploiting variation in the timing of Texas pre-k implementation across districts, Andrews et al. (2012) showed improvements in third grade reading and math scores and reduced special education and retention rates. Using a similar strategy of exploiting variation in between-county program rollout and expenditures, Ladd and colleagues found that North Carolina's preschool investments improved aggregate test scores in third grade (Ladd, Muschkin, & Dodge, 2014; Muschkin, Ladd, & Dodge, 2015). Analyses of Head Start participants in other strong quasi-experimental studies found substantial long-run effects, with positive impacts on academic and health outcomes of .2-.3 SD (Currie & Thomas, 1995; Deming, 2009; Garces, Thomas, & Currie, 2002; Ludwig & Phillips, 2008).

In sum, the literature on preschool fadeout and persistence is equivocal, leaving researchers trying to reconcile the reduction of moderate short-term academic effects with the promotion of long-term benefits in other domains. Often discussed in the literature are children's subsequent developmental contexts, and their role in sustaining or erasing the impacts of preschool. Bailey and colleagues (2015) use the term "sustaining environments" to refer to the idea that the quality of a child's environments subsequent to the completion of a preschool intervention may be important for sustaining early skill advantages. For young children, their elementary school environments are central to cultivating and maintaining the academic benefits from preschool.

**School Environment**

One common explanation for differences in the persistence of the impact of early education programs focuses on the quality of children's subsequent school experiences. Because school finances are tied to local property taxes, and because disadvantaged children tend to live in property-poor neighborhoods, they are more likely to enter lower resourced schools with lower quality instruction (M. A. Clements, Reynolds, & Hickey, 2004; Crosnoe & Cooper, 2010; McLoyd, 1998; Pianta, Belsky, Houts, & Morrison, 2007; Stipek, 2004). Unfortunately, these are the very children who have been shown to benefit the most from consistently cognitively stimulating environments (Crosnoe et al., 2010), which suggests that the schools they attend may be missing opportunities to have positive impacts on these children. This pattern may also mean that disadvantaged children have fewer opportunities for maintaining the benefits of preschool across the early grades (Currie & Thomas, 2000; Lee & Loeb, 1995; Reynolds, Ou, & Topitzes, 2004; Zhai, Raver, & Jones, 2012).

**Classroom instruction.** Evidence of how school quality or classroom quality affects the persistence of preschool gains is somewhat mixed. The fadeout phenomenon refers to one group of children with an initial skill advantage learning less over time than another group of children. The argument about poor quality schools eroding preschool gains is premised on the fact that low quality schools with low expectations of children's skills and low quality instruction are only able to help non-preschool attendees, who are likely lower skilled at school entry and gain more over time than preschool attendees (i.e., "catch-up"). However, few scholars would seriously argue that low quality instruction is more beneficial for low-skilled children.

A serious limitation of the prior work on school and classroom factors moderating preschool effects is the limited attention given to the important features of subsequent classroom instructional experiences. It seems intuitive that enriched instructional experiences produce greater gains and academic learning among all children. In turn, high quality instruction would benefit the academic skills of preschool graduates by continuing their skilled growth at the same (or faster) pace than their non-attending peers (Barnett, 2011; McKey et al., 1985; Swain, Springer, & Hofer, 2015; Zigler & Styfco, 2004).

A useful way to categorize components of quality instruction is by how teachers teach (i.e., pedagogy, instructional activities), what topics they cover (i.e., content and difficulty level) and the extent to which it meets the needs of individual students. Some consensus exists regarding the content teachers should cover in literacy and math instruction during the early grades, yet considerable debate surrounds early childhood pedagogical choices. During kindergarten, children need to acquire skills that will prepare them to read, including understanding the structure and uses of print, achieving basic phonemic awareness, and developing the ability to recognize and write most of the alphabet (Griffin, Snow, & Burns, 1999; National Reading Panel, 2000). As children grasp reading independently during first grade, meaning-focused instruction and daily encouragement to read and enjoy books are important for reading comprehension growth. In mathematics, kindergarten instruction typically involves activities designed to boost early numeracy skills. These activities often center on building an understanding of counting and number cardinality (e.g., when counting to 5, the number "5" represents a quantity of objects). As children become more confident in their ability to connect numbers to magnitudes, they then begin to work on basic operations (i.e., addition and subtraction; CCSSO/NGA, 2010; D. H. Clements & Sarama, 2014; National Council of Teachers in Mathematics, 2006). During the early-grade years, children also learn concepts and procedures relevant to geometry and measurement, such as knowing the salient properties of shapes (e.g., a square has 4 equal sides; D. H. Clements & Sarama, 2014).

No consensus exists concerning what constitute the best pedagogical approaches in early literacy and math. Research does not support either entirely “student centered” or “teacher directed” instruction (National Mathematics Advisory Panel, 2008). High-quality classroom instruction involves balancing and blending both child-directed and teacher led methods, and recognizing the individual needs of students (Connor et al., 2009; Early et al., 2010; National Mathematics Advisory Panel, 2008). Inherent here is that teachers tailor their instruction and classroom activities based on children’s skills. This is the idea behind formative assessment methods, with which teachers can recognize where students are struggling or in need of a greater challenge and adjust instruction accordingly (Penuel & Shepard, In press).

Nevertheless, teachers may provide high quality instruction *without* individualization. In the preschool fadeout scenario, good instruction could involve the teacher going as fast or far as possible *as measured by the lowest skilled children*—in this case, those who did not attend preschool—and not individualizing instruction for the higher-skilled preschool participants. This good but not individualized instruction would benefit the students with lower skills, facilitating catch-up or convergence with preschool attendees. In the best case, a teacher’s instruction is both of high pedagogical quality and tailored to the needs of students. Efforts to individualize elementary school instruction to prevent fadeout or convergence of preschool effects should therefore align instruction and curriculum with the preschool curriculum.

***Alignment with pre-k instruction.*** A high quality instructional experience in elementary school requires alignment between what children learned in pre-k and what is taught in kindergarten and first grade. Developmental theory suggests that regular exposure to content that is both beyond a child's current skill level and still within their range of abilities is critical for children's intellectual development (Bronfenbrenner, 1989; Vygotsky, 1978). This means that when preschool attendees enter elementary school with the foundational early math and literacy skills learned during preschool (e.g., letter recognition, cardinality), they should be exposed to sequentially more challenging tasks and concepts as they progress through the early grades for continued cognitive development. Thus, aligned instruction should prevent preschool fadeout.

However, recent evidence indicates a general lack of alignment between children's knowledge and teachers' instruction from preschool and kindergarten (Abry, Latham, Bassok, & LoCasale-Crouch, 2015). Primary curricula assume students have limited prior knowledge and only include low-level content, which most teachers are required to follow rigidly. Thus, teachers remain unaware that some of their students have already mastered the material they intend to teach (Sarama & Clements, 2015). A number of studies have found that kindergarten teachers spend too much instructional time on content already mastered by preschool graduates and that this may reduce achievement gains and facilitate non-preschool children catching-up to preschool participants (Engel et al., 2013; 2014; Gervasoni & Perry, 2015; Magnuson, Ruhm, & Waldfogel, 2007). In a kindergarten classroom, some children will have mastered key math and literacy skills (i.e., preschool graduates), whereas children whose first exposure to formal schooling occurs in kindergarten likely have very basic skills. Given the wide range of children's abilities in a given classroom, a teacher may be forced to teach to the lowest skill student to pass minimal competency assessments, leaving preschool attendees without the appropriate challenge they need to maintain growth (D. H. Clements, Sarama, Wolfe, & Spitler, 2013; Sarama & Clements, 2015; Sarama, Clements, Wolfe, & Spitler, 2012). Perhaps not surprisingly, exposure to advanced content, or content not already mastered prior to kindergarten entry, appears to support achievement growth while exposure to content already mastered impedes growth (Engel et al., 2013; 2014). Furthermore, exposure to advanced content in mathematics and literacy can be beneficial for both preschool graduates and students who did not attend preschool (Claessens, Engel, & Curran, 2013).

Providing teachers with information about children's entering skills and differentiating instruction based on their skills seem to be an effective means of boosting classroom learning. The results of experiments conducted by Connor and colleagues demonstrate the benefit of literacy instruction that is explicitly matched to a child's literacy skills (2009). Using assessments of children's literacy at the beginning of the school year and prescribing individualized instruction based on a computer algorithm produced significant and positive "child x instruction" interactions for reading skills. Thus, instruction may be most beneficial when it is tailored to the skill level of preschool graduates.

With respect to early math learning, Clements and colleagues demonstrate the importance of communicating to early elementary school teachers the math content that children learned in preschool in an attempt to get teachers to differentiate instruction (2013). Their research team evaluated the scale-up of a preschool mathematics curriculum in state preschool programs serving low-income communities in New York and Massachusetts. They found that random assignment to the preschool mathematics intervention had a large impact on end-of-preschool mathematics scores, but this effect faded substantially by first grade. However, some students were assigned to another condition, known as "follow-through," that included both the preschool mathematics intervention along with additional professional development sessions for their kindergarten and first grade teachers. In these sessions, teachers were informed of the mathematics content the students had learned in preschool, with the hope of reducing repetition in kindergarten and first grade. Results showed that, when compared with children who only received the preschool intervention, students assigned to the follow-through condition had substantially less effect fade out at the end of first grade. Our research includes the follow-through condition as well.

In summary: (a) some preschool interventions produce short term achievement effects that fade over time; (b) theoretical explanations implicate low quality instruction in the early elementary years, but evidence using rough proxy measures of quality have not fully evaluated this hypothesis; and (c) a better way to think about the issue might be the ability of subsequent classrooms to further growth among preschool graduates who are relatively more skilled. Data suggests that instruction may be largely low quality or too basic and not sufficiently advanced or individualized for many of the children who have attended preschool.

**Present Study**

This study builds on prior work to more directly investigate whether subsequent classroom instruction moderates the persistence of preschool gains. Using data from two evaluations of prominent preschool interventions, namely the Head Start Impact Study (HSIS; Puma et al., 2010) and the scale-up of the Building Blocks preschool mathematics curriculum, called TRIAD (Technology-enhanced, Research-based, Instruction, Assessment, and professional Development; D. H. Clements & Sarama, 2008), we examined the extent to which the persistence of preschool program effects on children's cognitive skills depends upon the features of the kindergarten and first grade classroom they attend. We used two key instructional characteristics, exposure to literacy content (HSIS) and math content (TRIAD), to operationalize the quality of the elementary school learning environment in kindergarten and first grades.

The primary research questions for this study were:

1. Does the quality of academic instruction in kindergarten and first grade moderate the magnitude of preschool intervention effects on children's academic skills in kindergarten and first grade?
2. Does a professional development intervention for kindergarten and first grade teachers that provided techniques designed to build upon the preschool program moderate preschool intervention effects on children's academic skills in kindergarten and first grade?

Although previous investigations of both the HSIS and TRIAD studies have examined the studies' initial treatment impacts and their subsequent fadeout patterns (Clements et al., 2011; 2013; Puma et al., 2010), our study is the first to investigate whether measures of later classroom instructional quality moderate treatment impact fadeout in these two interventions.

## Data

Our study involves data from two preschool interventions—Head Start and the TRIAD preschool mathematics intervention. *Head Start* is a comprehensive child development program that provides children with preschool education, health screenings and examinations, and nutritious meals, in a full-day, center-based setting. The Head Start children in our sample participated in the program during their pre-kindergarten year at age 4 at different research sites across the country as a part of the Head Start Impact Study evaluation that began in 2002 (Puma et al., 2010).

TRIAD evaluated the scale-up of the *Building Blocks* preschool mathematics curriculum, which encourages the acquisition of conceptual and procedural knowledge in both numeracy and geometric/spatial reasoning through the emphasis of empirically-supported learning trajectories (see D. H. Clements & Sarama, 2008). The evaluation was designed to assess the long-term impacts of the *Building Blocks* curriculum in 42 public elementary schools operating state preschool programs serving low-income communities in Massachusetts and New York in 2006. Our use of two different random assignment studies allows us to test the robustness of our findings. However, the family, child, and classroom characteristics differ slightly between each dataset. We have aligned our variables so that they are as similar as possible across datasets, but we present our data, analyses, and results separately by study for ease of exposition.

**Preschool Intervention I: Head Start**

The Head Start sample comes from the Head Start Impact Study (HSIS) experimental dataset, which is a nationally representative sample of Head Start participants and a group of comparable non-participants. The full sample included newly entering 3- and 4-year old Head Start applicants in which treatment children were randomly assigned to receive a Head Start spot, and control group children were not offered a spot in Head Start. Control group parents either made other ECE arrangements or cared for their children at home. Baseline survey and child assessment data were collected by study investigators in the Fall of 2002; post-treatment child assessments were collected at the end of Head Start in Spring 2003 and during kindergarten and first grade in Spring 2004 and 2005. Our analyses use the 4-year-old cohort only so that the children in both the HSIS and TRIAD analyses received the preschool intervention during the same developmental period. The HSIS 4-year-old cohort was further limited in our study to children whose elementary school teachers responded to the study survey and children who had not left the study at the kindergarten and first grade waves (*n* = 1080, 54% of the original 4-year-old cohort).

Table 1 presents descriptive statistics, along with p-values for differences between treatment and control groups, for the analytic sample. The children and families were all very low income and had the following characteristics: 45% Hispanic, 39% White, and 15% African American; 42% of parents had less than a high school degree; 23% of parents were recent immigrants; 16% of parents were teenage mothers; and a majority (84%) of the families lived in an urban area. Information on children's elementary school experiences was collected from kindergarten and first grade teachers through a teacher survey in the springs of 2004 and 2005.

**Children's language and literacy skills**. Literacy skills were measured through direct child assessments with the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997) and Letter Word and Spelling standard scores from the Woodcock-Johnson Psycho-Educational Battery-Revised III (Woodcock, McGrew, & Mather, 2001). To mitigate chance findings, owing to multiple testing, we created a literacy assessment composite measure to use as the main dependent variable by standardizing all three measures to mean 0 and standard deviation of 1, averaging across the three, and then re-standardizing the measure.

**Classroom environment**. In the HSIS teacher survey, teachers were asked how many times in the past week their class did a given literacy activity. We coded each activity into basic or advanced for grade level based on consultations with early literacy experts (available in Appendix A). We converted each basic and advanced activity into times per month by taking the mean value of the answer category (e.g., never = 0; 1-2 times per week = 1.5), multiplied it by 4, and then standardized this measure to have a mean of 0 and standard deviation of 1, following Claessens, Engel, and Curran (2013). We operationalized instructional quality during the first grade year as a cumulative measure of quality from both kindergarten and first grade, averaging the measures across the two years. We did this to create a more stable measure of instruction based on both assessments of the instructional environment in early grades (i.e., two teacher surveys). Prior research indicates that teacher reports of classroom activities can be valid measures of the quantity of instruction (Herman, Klein, & Abedi, 2000). Note that the available measures of literacy instruction are relatively weak when compared with detailed studies of literacy instructional research (e.g., Connor et al., 2009).

**Child and family covariates.** The HSIS dataset includes a robust set of child and family covariates. Our analyses include child race, gender, literacy assessment prescores (at study baseline), and special needs status. Family covariates include mother's race, age, and education level, urbanity, martial status, and language spoken at home (Spanish or English).

**Additional classroom and school variables to test for moderation.** Also included in the HSIS are measures of early school experiences that may influence children's learning and preschool effect persistence in the post-treatment period. We incorporate the following variables into additional moderation analyses for the HSIS only (see Results, *Additional models*): attendance at full-day kindergarten, kindergarten class size, classroom-level proportion of children in poverty (as measured by free and reduced-price lunch), and school-level proportions of children in poverty and children who are proficient in reading and mathematics.

**Preschool Intervention II: TRIAD**

The TRIAD evaluation study sampled a cohort of low-income children who attended state-funded pre-k in public schools at 4 years of age (n = 1375; D. H. Clements, Sarama, Spitler, Lange, & Wolfe, 2011; D. H. Clements et al., 2013; Sarama et al., 2012). In this study, schools were randomly assigned to one of three conditions: 1) *Building Blocks* preschool curriculum; 2)

*Building Blocks* preschool curriculum with follow-through; or 3) control (pre-k as usual). Children in schools assigned to the two *Building Blocks* groups received the curriculum during preschool (age 4), and preschool teachers attended 13 study-administered professional development (PD) sessions across two consecutive preschool years. Kindergarten and first grade teachers in schools assigned to the *"Building Blocks* with follow-through" group received PD designed to help bridge the gaps between preschool, kindergarten, and first grade. These PD sessions introduced teachers to what their children had learning in the previous year(s) and to the Building Blocks learning trajectories, with the intent that they would use these to alter their instruction to build on what students had already learned. Study participants were enrolled at study schools at the beginning of the preschool year (2006-2007).

Table 1 presents descriptive statistics for the full sample, as well as *p*-values indicating whether baseline characteristics differed by treatment status. In the full sample, 35% of students were assigned to the preschool-only treatment group and 36% were assigned to the preschool treatment with follow-through group. The majority of students qualified for free or reduced price lunch (84%); 53% identified as African American and 22% as Hispanic. All baseline characteristics were balanced between the three study conditions.

The analytic sample consists of students who had non-missing classroom observations in kindergarten and first grade, and had complete data on achievement tests in preschool, kindergarten and first grade ($n$ = 876). Students included in the analytic sample were less likely to be male (6%) and more likely to be free or reduced price lunch (30%).

**Children's mathematics skills.** Math achievement was assessed at preschool entry, and at the end of the preschool, kindergarten and first grade years via the Research Based Early Mathematics Assessment (REMA; D. H. Clements, Sarama, & Liu, 2008; D. H. Clements, Sarama, & Wolfe, 2011). The REMA is designed to measure the mathematics achievement of children from ages 3 to 8 and is administered through two structured interviews. During the interviews, the examiner presents the child with a series of questions targeted at assessing competency in counting, operations, measurement, and geometry, among other topics, and the interviews are videotaped and subsequently coded for both correctness and strategy use. The codes are then converted to a Rasch-IRT scaled score. The assessment was extensively validated in multiple samples, and has been shown to have a high correlation (.89) with the *Applied Problems* subtest of the Woodcock Johnson. The REMA has strong internal reliability ($\alpha$= .94; D. H. Clements, Sarama, & Wolfe, 2011).

**Classroom environment.** In the TRIAD evaluation*,* teachers' instructional practices were evaluated via the Classroom Observation of Early Mathematics Environment and Teaching (COEMET; see D. H. Clements, Sarama, Spitler, et al., 2011). The COEMET is composed of 28 Likert-scaled items. Assessors, who were blind to treatment status, rated classrooms for teaching practices known to support early math development, such as the use of engaging small group activities and emphasizing cognitively demanding concepts and strategies. For the majority of the items, observers rated each item every time they observed a new mathematical activity. For the kindergarten year, we took the average of these 28 items across every activity observed, and then standardized the scores. As with the HSIS, our measure of cumulative kindergarten and first grade instructional quality is the standardized average of a child's kindergarten and first grade COEMET scores. We did this to create a more stable measure of instruction based on two observations of the instructional environment in early grades (i.e., both COEMET assessments). We also included the number of mathematical activities observed during each COEMET period in our analysis to capture of the amount of instructional time spent on mathematics.

**Child and family covariates.** Although we found no indication of treatment assignment imbalance on observable characteristics, we still included a set of family and child covariates in order to improve precision in our models. This set of characteristics included child gender, ethnicity, age at preschool entry, free or reduced price lunch status, special education status at preschool entry, baseline mathematics score, and whether designated limited English proficient. We also included a control for mother's education and a set of dummy variables to account for blocking group at random assignment.

[Table 1 about here]

### Comparing Classroom Instructional Constructs

Note that our measures of classroom instructional topics in the HSIS are literacy activities dichotomized into advanced and basic. Therefore they capture content level and do not capture the quality of pedagogical strategies used in children's kindergarten and first grade classrooms. Conversely, the measures available in the TRIAD study assess the quality of mathematics pedagogy, and do not indicate whether the classroom content is advanced or basic for grade level. The second instructional measure in TRIAD, overall quantity of math activities, measures quantity of instruction, and not content difficulty. Thus, the strength of one study was the weakness of the other. This provided us with a unique opportunity to examine several key dimensions of children's subsequent educational experiences that may help to explain persistence or fadeout. However, we recognize that the schools and classrooms in both datasets were serving primarily low-income children, and therefore may not provide population-representative variation in instructional quality.

## Analysis

Our key research questions focus on whether subsequent instructional experiences moderate the magnitude of preschool treatment effects in children's kindergarten and first grade year. This requires an approach that models both the programmatic preschool impact and an interaction between this programmatic impact and observed classroom characteristics. We employed multivariate regression models using interaction terms to test for this moderation in both studies.

Because preschool interventions were randomly assigned, regression can provide unbiased estimates of the offer of preschool attendance. A key methodological issue, however, is that only the preschool offer is randomly assigned; children's later classroom experiences are not and children and families may select into different types of classroom environments post-treatment. To explore the potential for such bias, we first tested for selection into subsequent environments in both the HSIS (see Appendix B) and the TRIAD study (see Appendix C). Here we regressed a number of kindergarten classroom and school characteristics (e.g., class size, teacher's education, school reading proficiency level) on children's treatment status, including indicators for random assignment block. These tests revealed no evidence of differential selection into classroom or school environments by preschool treatment status. Nevertheless, if children with better potential outcomes were more likely to experience high quality instructional environments in kindergarten and first grade, our estimates are likely to be upwardly biased and represent an upper bound of the true effect of the environmental moderation.

Note that the TRIAD study dataset did not include the large set of classroom and school characteristics available in the HSIS, limiting our investigation of differential selection with these data. However, both the treatment and control group children in the TRIAD study had selected into public school pre-k programs, unlike the HSIS which included varied control group conditions (e.g., center-based care, parent care, family child care home). In addition, the pre-k programs were located in the local public schools where children's kindergarten programs were also located. These features of the TRIAD study likely reduced the opportunities for children to differentially sort into public school kindergarten and first grade environments. Indeed, we found no evidence that children assigned to either treatment condition were more likely to remain in the same school than children in the control group.

In the HSIS models, we focused on literacy outcomes as the dependent variable, while in the TRIAD models we focused on mathematics. These were the domains with the largest treatment impact in each study respectively and thus provided the most potential variation to detect treatment effect moderation in kindergarten and first grade. In all models, we regressed achievement measures (taken at either end of preschool, kindergarten, or first grade) on treatment status, baseline assessment scores, indicators for random assignment block, and a set of child and family control variables, varying slightly between TRIAD and HSIS. We then added measures of classroom instruction as covariates to see how much of the treatment effect was explained by high-quality instructional practices. Finally, we added models in which treatment was interacted with classroom instruction. If high-quality instruction in kindergarten and first grade helps reduce fadeout, then these interactions should be positive and significant.

Although there was no evidence of differential selection into elementary school environments based on observed characteristics, one might be concerned that unobserved or unmeasured features of the kindergarten classroom may still bias our estimates of fadeout. One way to address this concern is to compare the outcomes of treatment and control children experiencing the same instructional environment in a classroom fixed effects model (i.e., within-classroom analysis). Because the HSIS sample included treatment and control children who attended the same kindergarten class, we were able to estimate a kindergarten classroom fixed effect model with the HSIS data but not with the TRIAD study. This model is included in the main results below.

**Additional Details for HSIS Analyses**

In the HSIS study, children in the control group who left the study by first grade had significantly lower outcome scores (Woodcock-Johnson Applied Problems, Letter-Word and Spelling subtests) at the end of the preschool year, on average, than children from the control group who left the study by first grade. The HSIS investigators created analytic weights for each wave to address differences in nonresponse, attrition, and for complex sampling. However, recent studies using the HSIS do not use these weights because they could not be replicated by other analysts using the data (Bitler, Hoynes, & Domina, 2014; Bloom & Weiland, 2015). To address attrition as well as imbalances between treatment and control children without using the provided weights, we followed the strategy from Bitler et al. (2014) in their analyses of the HSIS data and used inverse probability of treatment weights (IPTW) for all HSIS models. Another efficiency from using IPTW is that they also effectively control for the covariates shown in Table 1. Note that these weights do not adjust for kindergarten and first grade teacher nonresponse.

The HSIS was a very comprehensive study, and the dataset includes other characteristics about the kindergarten classroom environment, such as class size and proportion of children in poverty. Therefore we estimated additional models with the HSIS only that test for alternative hypotheses from prior research about the suppression or maintenance of treatment effects in elementary school (Chetty et al., 2011; Magnuson et al., 2007; Nye, Hedges, & Konstantopoulos, 2000; Stipek, 2004).

## Results

All HSIS results are presented in Table 2, and all TRIAD results are presented in Table 3. All variables except the treatment indicator and parents' education are scaled in standard deviation units to facilitate their interpretation as effect sizes. Analyses of outcomes based on teacher response status are available in Appendices D and E for HSIS and TRIAD, respectively. Analyses of individual literacy outcomes comprising the literacy composite for the HSIS are available in Appendix G.

### Preschool Intervention I: Head Start

**Kindergarten.** Table 2a presents the results for tests of moderation of impacts by kindergarten instruction on literacy and language skills for the HSIS sample, where the dependent variable is a composite of three literacy and language assessments. Consistent with prior analyses of these data, we find a modest effect of the Head Start offer on children's early literacy and language skills (effects size of .16, Model 1). However, by the end of kindergarten Head Start offer is no longer predicting children's language and literacy skill level (Model 2). When we add the instructional quality variables (basic and advanced literacy instruction) in Model 3, we find that, as expected, a higher frequency use of advanced literacy activities is associated with better skills. This is not true for basic literacy activities, which are negatively associated with children's skills. Although we would expect such a pattern of associations, we note that the cross-sectional nature of this analysis does not allow for causal interpretation. That is, it might reflect that teachers are more likely to implement more advanced instruction with children that enter their classroom with higher levels of skills. The interactions between instructional quality and Head Start offer, however, do not predict children's language and literacy outcomes. The emphasis on either basic or advanced literacy strategies does not affect the persistence of Head Start offer impacts by the end of the kindergarten year. We also examined whether the relationships between advanced and basic instruction and children's literacy skill development were nonlinear by constructing "high" and low "indicators" for each measure based on a median split in the distribution for the measure. No significant relationships emerged when using these forms of the measure. In sum, greater exposure to more advanced literacy instruction did not sustain the gains of the Head Start treatment group children through the kindergarten year.

We were also able to estimate a kindergarten classroom fixed effect model with the HSIS (Model 5), which tests whether Head Start participants have stronger literacy skills at the end of kindergarten relative to a control child in the same kindergarten classroom (~250 children shared a classroom with a control child). This compares children who received the same overall classroom experience, including instruction, teacher's interactions, peer ability, class size, and any other unobserved features about the classroom. In this model, the coefficient of interest is the treatment indicator itself, which denotes the difference in literacy skills between a treatment and control child within the same kindergarten classroom, also controlling for child and family characteristics. The treatment effect in the kindergarten fixed effect model was not significant, and effectively zero (.01).

**First grade.** Results presented in in Table 2b show the effect of the Head Start offer on the literacy composite at the end of first grade. As expected, we do not find that Head Start predicted children's level of literacy and language skills. However, the coefficients on advanced and basic literacy activities were very similar to those in the kindergarten models, with a positive main effect of advanced and negative main effect of basic literacy activities. Both coefficients were significant at the .10 level. Still, there was no evidence that high-quality instruction, as measured by our teacher-report indicators, sustained Head Start treatment effects through the end of the first grade year. In other words, we found no interaction between Head Start status and basic or advanced literacy activities.

**Additional models.** One concern with the previous models is that focus on specific instruction is misguided and we have overlooked some overall experience in the early school years that is especially important. Put another way, perhaps earlier explanations that focus on overall school quality are more likely to show an association with the persistence of program impacts. Appendix F presents the results for similar models that tested for moderation by several other measures of early school experiences that might affect children's learning (described in *Measures*). None of these variables, when interacted with treatment, was a statistically significant predictor of kindergarten or first grade outcomes. Nor was there a coherent or consistent pattern of associations suggesting that Head Start children fared better or worse under these varying school conditions relative to children in the control group.

[Table 2 about here]

**Preschool Intervention II: Building Blocks**

**Kindergarten.** Table 3a, presents the impacts of TRIAD on kindergarten mathematics achievement. Model 1 shows the *Building Blocks* preschool treatment effect at the end of the preschool year with an effect size of .67. At kindergarten (Model 2), the effect dropped to .33 and remained significant. When we added the instructional quality variables (COEMET overall score and number of math activities) in Model 3, the treatment effect remained unchanged, and the coefficient for the number of math activities was a significant predictor of children's math achievement, yielding an effect size of 0.12, but instructional quality, as measured by the overall average COEMET score, was not. We find no evidence that the interactions between instructional quality and treatment status entered in Model 4 were significant predictors of children's math achievement.

Models 5 and 6 take a different approach to explaining treatment impact fadeout. They examined the extent to which a kindergarten follow-through treatment with teacher PD including aspects that could reduce instructional repetition predicted math achievement. Note that teachers were resistant to this aspect of the PD because it meant that they were teaching a new curriculum for the first time. Unlike our measures of classroom instructional quality, teachers were randomly assigned to engage in additional PD. The treatment effect for students in the follow-through group was .38 and significant, but it was not significantly different ($p$= .54) from the end of kindergarten impact for students in the treatment group without follow-through (.32). We also examined whether the relationships between quality math instruction and children's math skills development was nonlinear by splitting the COEMET and number of math activities variables at the median, which created high and low categories. We found no significant treatment interactions with this set of dummy variables.

We also tested whether instructional quality could account for differences in mathematics achievement among children assigned to the follow-through condition at the end of preschool. Results suggest that the number of mathematics activities again produced a positive and significant coefficient (.14), but the overall teaching quality measure (COEMET) had no detectable effect on achievement. When we interacted follow-through treatment with instructional quality, these terms were not significant.

**First grade**. In Table 3b, we present results from models examining the relation between high-quality mathematics instruction in kindergarten and first grade and treatment effects at the end of first grade. Model 1 again presents the treatment effect at the end of preschool for the non-follow through treatment condition, and Model 2 shows that the effect falls to a statistically significant .16 at the end of first grade. When we added the instructional quality variables in Model 3, the Building Blocks preschool-only treatment effect did not change, and the number of math activities was significant and positive (.20), while the overall measure of instructional quality did not affect achievement. Adding the interaction with treatment in Model 4 produced a similar pattern, where the preschool-only treatment effect was nearly unchanged and significant (.17), and the main effect for number of math activities was .21 and significant. The interaction between instructional quality and treatment was not significant, nor was the interaction between number of math activities and treatment. These results also showed no evidence that high-quality instruction sustains pre-k skill advantages.

As with the last two kindergarten models in Table 3a, Model 5 in Table 3b focused on the follow-through condition in the TRIAD study. A comparison of treatment effects for both the preschool-only treatment and the follow-through group shows that the effect size for the follow-through group was .32 and significant, compared with .18 for the preschool-only group. The .32 coefficient was only slightly smaller than the .38 coefficient found for the follow-through group at the end of kindergarten, suggesting very little fadeout during first grade. Comparing the follow-through and preschool only group impacts at the end of first grade revealed that the two effect sizes are only statistically significant from one another at the .10 level ($p$ = .10).

We again tested whether differences in elementary school instructional quality explained variation in end of first grade mathematics achievement for children in the follow-through condition. Although Models 6 and 7 show positive main effects for number of math activities observed on later achievement, interactions between follow-through treatment and instructional quality did not predict first-grade mathematics achievement.

[Table 3 about here]

## Discussion

Our study tested whether measures of kindergarten and first-grade classroom quality could explain variation in program fadeout following two different preschool interventions. As other studies have reported (D. H. Clements et al., 2013; Puma et al., 2012), we found substantial treatment impact fadeout in both samples. Head Start treatment effects were insignificant by kindergarten, and TRIAD treatment effects were reduced by 70% between end of preschool and the end of first grade. Our measures of classroom quality were predictive of kindergarten and first grade achievement in both studies, where advanced instruction supported literacy skills, and basic instruction inhibited them in HSIS. Similarly in TRIAD, the number of mathematics activities observed predicted mathematics achievement in both kindergarten and first grade. However, none of these measures were significant when interacted with treatment status, indicating that kindergarten and first grade instructional quality did not explain differences in preschool impact fadeout.

In contrast, we found that the additional PD offered to teachers in the "follow-through" condition of TRIAD did sustain effects in first grade (see also Clements et al., 2013). Indeed, Clements and colleagues (2013) reported large end of first grade effects for students in the follow-through condition, and our analysis further illustrates that this treatment condition was able to abate much of the fadeout pattern observed for the treatment arm with no follow-through PD. These results suggest that targeted PD—designed to create continuity and avoid repetition between grades—can be an effective way to sustain the impacts from content specific high-quality early childhood interventions. Interestingly, our measure of instructional quality did little to explain the added benefit of the follow-through condition, as we found that the COEMET measure did not mediate the impact of the follow-through condition on later achievement. Thus, the benefits of the follow-through condition were not explained by our measure of instructional quality, leaving questions regarding which specific features of the additional PD were responsible for the sustained treatment impacts.

What might account for our primary findings that high quality early grades instruction did not moderate the persistence of preschool treatment effects? First, our measures of classroom quality may have been unable to capture the classroom experiences essential for sustaining early academic gains. Our classroom environment measure in the HSIS merely captured whether teachers reported spending time on certain types of activities. The TRIAD study included only a single observation in the kindergarten year and another during first grade, and these observations were solely focused on the quality of mathematics instruction, and not the specific content taught. We noted that the individualization of instruction is a key component of high quality instruction, but could not precisely measure whether teachers matched classroom instruction to children's skill levels because both studies only included classroom level aggregate instructional measures. In this way, our measures are likely weak proxies of true individualization, where teachers would recognize the more advanced literacy or math skills of preschool graduates and subsequently increase their overall quantity of math activities (à la TRIAD), or the amount of time on more advanced topics (à la HSIS).

However, our TRIAD follow-through condition models do suggest that classroom environments can play a role in sustaining early academic impacts if efforts are made to specifically target instruction to the needs of students with preschool experience. In other words, broadly improving instructional quality or content is not likely to sustain preschool impacts, but specifically designing instructional practices and curriculum in a manner that builds upon the competencies children have already gained may facilitate preschool effect persistence. Indeed, this strategy is likely more cost intensive, as teachers in the TRIAD follow-through condition were given incentives to attend additional PD. In addition, study designers implementing the kindergarten and first-grade PD sessions were very familiar with the knowledge students gained during preschool, and this knowledge was likely essential to successful PD. However, even though the TRIAD follow-through condition sustained effects through the end of first grade, approximately 50% of the original treatment effect still faded. Nevertheless, future research should examine the specific components of the follow-through condition that worked, and how this can be done successfully during kindergarten.

In addition to those mentioned above, other limitations of our study relate to measurement. Because both of our samples are very low income and attend schools with limited resources, we do not have wide variation in instructional quality measures. This narrow range of classrooms may have limited our ability to detect any persistence from high quality instruction (measures were standardized within each sample). Nevertheless, it should be noted that some measures of classroom quality (advanced literacy instruction, total number of math activities) produced the expected positive main effects. This suggests that the measures did capture some meaningful variation in classroom quality that related to student achievement in predictable ways. Furthermore, our measurement of literacy instructional quality is poor relative to the scientific literature on reading instruction. Here, researchers measure several more detailed dimensions of reading, such as code focused versus meaning-based, or teacher-managed versus child managed (Connor, Morrison, & Katch, 2004), and we did not have this level of detail captured in our data. Therefore, we cannot be confident based on our study and measures alone that high quality instruction does not prevent preschool program fadeout. Future studies would benefit from better measures of literacy instructional quality from both trained observers and teacher reports, and from more precise measures of individualization in both literacy and mathematics.

**Conclusion**

Through analyses with longitudinal evaluations of two enriched preschool interventions, we did not find any evidence to support the hypothesis that better instructional quality mitigates the fadeout of preschool treatment effects on children's academic skills during elementary school. However, we did find that advanced instruction was linked with positive gains, while basic instruction was linked with losses, in children's literacy skills, confirming past findings that all elementary school children benefit from higher level instruction, regardless of preschool history. We also found some evidence that the coupling of the TRIAD intervention with teacher professional supports in kindergarten and first grade all but eliminated the fadeout of effects on math achievement observed between kindergarten and first grade. Still, with both the focused PD and high instructional quality, TRIAD could not reduce fadeout effects between preschool and kindergarten. Future research should investigate aligned preschool-elementary school curricular approaches to sustain the benefits of preschool programs for low-income children.

**Acknowledgements**

This research was supported by awards from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) to [redacted]. Research was also supported by the Institute for Education Sciences (IES) from awards to [redacted]. We wish to thank [redacted] for comments on prior drafts, and to [redacted] for their contribution to the TRIAD study. The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of NICHD or IES.

**References**


Abry, T., Latham, S., Bassok, D., & LoCasale-Crouch, J. (2015). Preschool and kindergarten teachers' beliefs about early school competencies: Misalignment matters for kindergarten adjustment. *Early Childhood Research Quarterly, 31*, 78-88. doi:http://dx.doi.org/10.1016/j.ecresq.2015.01.001

Aos, S., Miller, M., & Drake, E. (2006). Evidence-based public policy options to reduce future prison construction, criminal justice costs, and crime rates. *Federal Sentencing Reporter, 19*(4), 275.

Bailey, D., Duncan, G. J., Odgers, C., & Yu, W. (2015). *Persistence and fadeout in the impacts of child and adolescent interventions*. Working Paper. Irvine, CA.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *Future of Children, 5*, 25-50.

Barnett, W. S. (2011). Effectiveness of Early Educational Intervention. *Science, 333*(6045), 975-978. doi:10.1126/science.1204534

Barnett, W. S., & Masse, L. N. (2007). Comparative benefit-cost analysis of the Abecedarian program and its policy implications. *The Economics of Early Childhood Education, 26*(1), 113-125.

Belfield, C. R., Nores, M., Barnett, W. S., & Schweinhart, L. J. (2006). The High/Scope Perry Preschool Program. *Journal of Human Resources, XLI*(1), 162-190. doi:10.3368/jhr.XLI.1.162

Bennett, N., Desforges, C., Cockburn, A., & Wilkinson, B. (1984). *Quality of Pupil Learning Experiences*. New York, NY: Routledge.

Bitler, M. P., Hoynes, H. W., & Domina, T. (2014). Experimental Evidence on Distributional Effects of Head Start. *National Bureau of Economic Research Working Paper Series, No. 20434*. doi:10.3386/w20434

Bloom, H. S., & Weiland, C. (2015, Nov 7). *Quantifying Variation in Head Start Effects on Young Children's Cognitive and Socio-emotional Skills using data from the National Head Start Impact Study*. Working Paper. New York, NY.

Bronfenbrenner, U. (1989). Ecological systems theory. In R. Vasta (Ed.), *Annals of child development* (Vol. 6, pp. 187-249). Greenwich, CT: JAI Press.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*(3), 579-620.

Campbell, F. A., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early Childhood Investments Substantially Boost Adult Health. *Science, 343*(6178), 1478-1485. doi:10.1126/science.1248429

Campbell, F. A., Pungello, E., Burchinal, M. R., Kainz, K., Pan, Y., Wasik, B. H., . . . Ramey, C. T. (2012). Adult outcomes as a function of an early childhood educational program: an Abecedarian Project follow-up. *Developmental Psychology, 48*(4), 1033.

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6*, 42-57.

Campbell, F. A., Wasik, B. H., Pungello, E., Burchinal, M. R., Barbarin, O. A., Kainz, K., . . . Ramey, C. T. (2008). Young adult outcomes of the Abecedarian and CARE early childhood educational interventions. *Early Childhood Research Quarterly, 23*(4), 452-466. doi:10.1016/j.ecresq.2008.03.003

CCSSO/NGA. (2010). Common core state standards for mathematics. Washington, DC: Council of Chief State School Officers and the National Governors Association

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics, 126*(4), 1593-1660. doi:10.1093/qje/qjr041

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

Clements, D. H., & Sarama, J. (2008). Experimental Evaluation of the Effects of a Research-Based Preschool Mathematics Curriculum. *American Educational Research Journal, 45*(2), 443-494. doi:10.3102/0002831207312908

Clements, D. H., & Sarama, J. (2014). *Learning and teaching early math: The learning trajectories approach*. New York, NY: Routledge.

Clements, D. H., Sarama, J., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: the Research-Based Early Maths Assessment. *Educational Psychology, 28*(4), 457-482.

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education, 42*(2), 127-166.

Clements, D. H., Sarama, J., & Wolfe, C. B. (2011). *TEAM--Tools for early assessment in mathematics*. Columbus, OH: McGraw-Hill Education.

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies persistence of effects in the third year. *American Educational Research Journal, 50*(4), 812-850.

Clements, M. A., Reynolds, A. J., & Hickey, E. (2004). Site-level predictors of children's school and social competence in the Chicago Child–Parent Centers. *Early Childhood Research Quarterly, 19*(2), 273-296. doi:http://dx.doi.org/10.1016/j.ecresq.2004.04.005

Connor, C. M., Morrison, F. J., & Katch, L. E. (2004). Beyond the Reading Wars: Exploring the Effect of Child-Instruction Interactions on Growth in Early Reading. *Scientific Studies of Reading, 8*(4), 305-336. doi:10.1207/s1532799xssr0804_1

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., . . . Morrison, F. J. (2009). Individualizing Student Instruction Precisely: Effects of Child × Instruction Interactions on First Graders' Literacy Development. *Child Development, 80*(1), 77-100. doi:10.1111/j.1467-8624.2008.01247.x

Crosnoe, R., & Cooper, C. E. (2010). Economically disadvantaged children's transitions into elementary school: Linking family processes, school contexts, and educational policy. *American Educational Research Journal, 47*(2), 258-291. doi:10.3102/0002831209351564

Crosnoe, R., Leventhal, T., Wirth, R. J., Pierce, K. M., Pianta, R. C., & NICHD Early Child Care Research Network. (2010). Family Socioeconomic Status and Consistent Environmental Stimulation in Early Childhood. *Child Development, 81*(3), 972-987. doi:10.1111/j.1467-8624.2010.01446.x

Currie, J. (2001). Early Childhood Education Programs. *Journal of Economic Perspectives, 15*(2), 213-238.
Currie, J., & Thomas, D. (1995). Does Head Start Make a Difference? *The American Economic Review, 85*(3), 341-364. doi:10.2307/2118178
Currie, J., & Thomas, D. (2000). School Quality and the Longer-term Effects of Head Start. *Journal of Human Resources, 35*(4), 755-774.
Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics, 1*(3), 111-134. doi:10.2307/25760174
Duncan, G. J., & Magnuson, K. (2013). Investing in Preschool Programs. *The Journal of Economic Perspectives, 27*(2), 109-132. doi:10.1257/jep.27.2.109
Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test--Third Edition (PPVT-III)*. Upper Saddle River, NJ: Pearson Publishing.
Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D.-M. C., Crawford, G. M., . . . Pianta, R. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 25*(2), 177-193. doi:http://dx.doi.org/10.1016/j.ecresq.2009.10.003
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. doi:10.3102/0162373712461850
Engel, M., Claessens, A., Watts, T. W., & Farkas, G. (2014). *The misalignment of kindergarten mathematics content*. Working Paper. Irvine, CA.
Fox, S. E., Levitt, P., & Nelson, C. A. (2010). How the Timing and Quality of Early Experiences Influence the Development of Brain Architecture. *Child Development, 81*(1), 28-40. doi:10.1111/j.1467-8624.2009.01380.x
Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of Head Start. *The American Economic Review, 92*, 999-1012.
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* (pp. 47-64). New York, NY: Springer.
Griffin, P., Snow, C. E., & Burns, M. S. (1999). *Starting Out Right:: A Guide to Promoting Children's Reading Success*: National Academies Press.
Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in cognitive sciences, 13*(2), 65-73.
Herman, J. L., Klein, D. C. D., & Abedi, J. (2000). Assessing Students' Opportunity to Learn: Teacher and Student Perspectives. *Educational Measurement: Issues and Practice, 19*(4), 16-24. doi:10.1111/j.1745-3992.2000.tb00042.x
Hillm, C. J., Gormley, W. T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly*.
Jenkins, J. M. (2014). Early childhood development as economic development: Considerations for state-level policy innovation and experimentation. *Economic Development Quarterly, 28*(1), 147-165.
Ladd, H. F., Muschkin, C. G., & Dodge, K. A. (2014). From Birth to School: Early Childhood Initiatives and Third-Grade Outcomes in North Carolina. *Journal of Policy Analysis and Management, 33*(1), 162-187.

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. doi:10.3102/01623737017001062

Li, W., Leak, J., Duncan, G. J., Magnuson, K., Schindler, H., & Yoshikawa, H. (2016). *Is Timing Everything? How Early Childhood Education Program Impacts Vary by Starting Age, Program Duration and Time Since the End of the Program*. Working Paper. National Forum on Early Childhood Policy and Programs, Meta-analytic Database Project. Center on the Developing Child, Harvard University.

Lipsey, M. W., Farran, D. C., & Hofer, K. G. (2015). *A Randomized Control Trial of the Effects of a Statewide Voluntary Prekindergarten Program on Children's Skills and Behaviors through Third Grade (Research Report)*. Nashville, TN.

Lipsey, M. W., Hofer, K. G., Dong, N., Farran, D. C., & Bilbrey, C. (2013). *Evaluation of the Tennessee Voluntary Prekindergarten program: Kindergarten and first grade follow-up results from the randomized control design*. Vanderbilt University, Peabody Research Institute, Nashville, TN.

Ludwig, J., & Phillips, D. A. (2008). Long-Term Effects of Head Start on Low-Income Children. *Annals of the New York Academy of Sciences, 1136*(Reducing the Impact of Poverty on Health and Human Development: Scientific Approaches), 257-268.

Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007). The persistence of preschool effects: Do subsequent classroom experiences matter? *Early Childhood Research Quarterly, 22*(1), 18-38. doi:http://dx.doi.org/10.1016/j.ecresq.2006.10.002

McKey, R. H., Condelli, L., Ganson, H., Barrett, B., McConkey, C., & Plantz, M. C. (1985). The Impact of Head Start on Children, Families and Communities. Final Report of the Head Start Evaluation, Synthesis and Utilization Project: Department of Health and Human Services.

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53*(2), 185-204.

Muschkin, C. G., Ladd, H. F., & Dodge, K. A. (2015). Impact of North Carolina's Early Childhood Initiatives on Special Education Placements in Third Grade. *Educational Evaluation and Policy Analysis, 37*(4), 478-500.

National Council of Teachers in Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. Retrieved from Washington, DC: http://beck.cherryhill.k12.nj.us/nctm.pdf

National Mathematics Advisory Panel. (2008). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel*. Washington, DC.

National Reading Panel. (2000). *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Washington, D.C.: National Institute of Child Health and Human Development, National Institutes of Health.

Nye, B., Hedges, L. V., & Konstantopoulos, S. (2000). The Effects of Small Classes on Academic Achievement: The Results of the Tennessee Class Size Experiment. *American Educational Research Journal, 37*(1), 123-151. doi:10.3102/00028312037001123

Penuel, W. R., & Shepard, L. A. (In press). Assessment and teaching. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of Research on Teaching*. Washington, DC: AERA.

Pianta, R., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to Learn in America's Elementary Classrooms. *Science, 315*(5820), 1795-1796. doi:10.1126/science.1139719

Puma, M., Bell, S., Cook, R., & Heid, C. (2010). *Head Start Impact Study. Final Report.* U.S. Department of Health and Human Services, Administration for Children and Families, Washington, DC.

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., . . . Downer, J. T. (2012). *Third grade follow-up to the Head Start Impact Study: Final Report*. (OPRE Report 2012-45). Washington, DC.

Reynolds, A. J., Ou, S., & Topitzes, D. (2004). Path of effects of early childhood intervention on educational attainment and delinquency: A confirmatory analysis of the Chicago Child-Parent Centers. *Child Development, 75*, 1299-1328.

Reynolds, A. J., Temple, J. A., & Ou, S. (2010). Preschool education, educational attainment, and crime prevention: Contributions of cognitive and non-cognitive skills. *Children and Youth Services Review, 32*(8), 1054-1063. doi:10.1016/j.childyouth.2009.10.019

Sarama, J., & Clements, D. H. (2015). Scaling up early mathematics interventions: Transitioning with trajectories and technologies. In B. Perry, A. MacDonald, & A. Gervasoni (Eds.), *Mathematics and transition to school* (pp. 153-169). Singapore: Springer.

Sarama, J., Clements, D. H., Wolfe, C. B., & Spitler, M. E. (2012). Longitudinal Evaluation of a Scale-Up Model for Teaching Mathematics With Trajectories and Technologies. *Journal of Research on Educational Effectiveness, 5*(2), 105-135. doi:10.1080/19345747.2011.627980

Schweinhart, L. J. (2005). *Lifetime Effects: The High/Scope Perry Preschool Study through Age 40*. High/Scope Educational Research Foundation, Ypsilanti, M.I.

Shonkoff, J. P., & Meisels, S. J. (Eds.). (2000). *The Handbook of Early Intervention* (Second ed.). New York: Cambridge University Press.

Stipek, D. (2004). Teaching practices in kindergarten and first grade: Different strokes for different folks. *Early Childhood Research Quarterly, 19*(4), 548-568. doi:http://dx.doi.org/10.1016/j.ecresq.2004.10.010

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). doi:10.1177/2332858415612751

Vygotsky, L. S. (1978). Interaction between learning and development. *Readings on the development of children, 23*(3), 34-41.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). Woodcock-Johnson-III Tests of Achievement. Itasca, IL: Riverside.

Zhai, F., Raver, C. C., & Jones, S. M. (2012). Academic performance of subsequent schools and impacts of early interventions: Evidence from a randomized controlled trial in Head Start settings. *Children and Youth Services Review, 34*(5), 946-954. doi:http://dx.doi.org/10.1016/j.childyouth.2012.01.026

Zigler, E. F., & Styfco, S. J. (Eds.). (2004). *The Head Start Debates*. Maryland: Paul H. Brookes Publishing.

**Table 1. Descriptive Statistics for the Head Start Impact Study and Building Blocks Samples**

| | *Mean* | *SD* | *p-value* |
|---|---|---|---|
| **Head Start Impact Study: Age 4 Cohort** | | | |
| Treatment | 0.62 | | |
| Preschool Entry Literacy Skills Composite Score (std.) | 0.03 | 1.00 | 0.04 |
| Male | 0.51 | | 0.64 |
| African American | 0.22 | | 0.39 |
| Hispanic | 0.43 | | 0.43 |
| Limited English Proficiency | 0.32 | | 0.35 |
| Child has Special Needs | 0.14 | | 0.51 |
| *Mothers Education* | | | |
| Below High School | 0.42 | | 0.84 |
| High School | 0.31 | | 0.51 |
| > High School | 0.26 | | 0.35 |
| **Building Blocks TRIAD Study** | | | |
| Pre-k Only Treatment | 0.35 | | |
| Follow-Through Treatment | 0.36 | | |
| Pre-k Entry Math Skills Composite Score (std.) | 0.00 | 1.00 | 0.72 |
| Male | 0.49 | | 0.66 |
| African American | 0.53 | | 0.77 |
| Hispanic | 0.22 | | 0.76 |
| Ethnicity- Other | 0.06 | | 0.31 |
| Limited English Proficiency | 0.16 | | 0.51 |
| Child is in Special Education | 0.17 | | 0.87 |
| *Mother's Education* | | | |
| Below High School | 0.15 | | 0.93 |
| High School | 0.32 | | 0.90 |
| > High School | 0.53 | | 0.98 |

*Note.* The *p*-value column for the Head Start Impact Study presents results from tests of proportions and means between the treatment and control group. The p-value column for the Building Blocks Study presents the results from an *F*-test that tested whether students assigned "Pre-k only group" and "Follow-through group" were significantly different from each other, and control group children, on each of the listed characteristics. There were no differences between either Building Blocks treatment group and the control group.

**Table 2. Head Start Impact Study Results**

**a. Sustained Classroom Environment - End of Kindergarten Literacy Composite**

| | (1) End of HS | (2) Spring of K | (3) Spring of K | (4) Spring of K | (5) Kindergarten Classroom Fixed Effect |
|---|---|---|---|---|---|
| Treatment | 0.16* | -0.11+ | -0.11+ | -0.11+ | 0.01 |
| | (0.06) | (0.06) | (0.06) | (0.06) | (0.29) |
| Total Advanced Literacy Activities in K (times per month; std.) | | | 0.11* | 0.09 | |
| | | | (0.05) | (0.07) | |
| Total Basic Literacy Activities in K (times per month; std.) | | | -0.11* | -0.09 | |
| | | | (0.05) | (0.07) | |
| Treat * Advanced Literacy Activities | | | | 0.03 | |
| | | | | (0.08) | |
| Treat * Basic Literacy Activities | | | | -0.03 | |
| | | | | (0.09) | |
| Observations | 1632 | 1077 | 1075 | 1075 | 1077 |

**b. Sustained Classroom Environment - End of First Grade Literacy Composite**

| | (1) End of HS | (2) Spring of 1st Grade | (3) Spring of 1st Grade | (4) Spring of 1st Grade |
|---|---|---|---|---|
| Treatment | 0.16* | -0.08 | -0.07 | -0.07 |
| | (0.07) | (0.06) | (0.06) | (0.07) |
| Total Advanced Literacy Activities in K and G1 combined (times per month; std.) | | | 0.12+ | 0.12 |
| | | | (0.07) | (0.11) |
| Total Basic Literacy Activities in K and G1 combined (times per month; std.) | | | -0.12+ | -0.11 |
| | | | (0.07) | (0.10) |
| Treat * Advanced Literacy Activities | | | | -0.00 |
| | | | | (0.14) |
| Treat * Basic Literacy Activities | | | | -0.02 |
| | | | | (0.12) |
| Observations | 1632 | 1065 | 1065 | 1065 |

Model 1 includes all children in the HSIS age-4 cohort; Models 2-4 only include children whose teacher responded to the survey. Analyses of outcomes based on teacher response status available in Appendix D. Analyses using each individual outcome variable comprising the literacy composite are available in Appendix G. Standard errors clustered at school level (in parentheses). All models are weighted using inverse probability of treatment weights to adjust for differential attrition and complex sampling. Weights include all the baseline child and family control variables. + p<.10; * p<.05; ** p<.01

## Table 3. Building Blocks TRIAD Scale-Up Study Results

### 3a. Sustained Classroom Environment – End of Kindergarten Math Composite

| | (1) End of Pre-k | (2) Spring of K | (3) Spring of K | (4) Spring of K | (5) End of Pre-k | (6) Spring of K | (7) Spring of K | (8) Spring of K |
|---|---|---|---|---|---|---|---|---|
| Treatment | 0.67** | 0.33** | 0.33** | 0.34** | 0.67** | 0.32*** | 0.31** | 0.31** |
| | (0.07) | (0.08) | (0.07) | (0.08) | (0.08) | (0.08) | (0.07) | (0.07) |
| Mathematics Teaching Quality | | | 0.03 | -0.00 | | | 0.03 | 0.04 |
| | | | (0.04) | (0.05) | | | (0.03) | (0.03) |
| Number of Math Activities | | | 0.12* | 0.12 | | | 0.14** | 0.12* |
| | | | (0.05) | (0.09) | | | (0.04) | (0.05) |
| Treat * Mathematics Teaching Quality | | | | 0.09 | | | | |
| | | | | (0.07) | | | | |
| Treat * Number of Math Activities | | | | 0.00 | | | | |
| | | | | (0.08) | | | | |
| Treatment with Follow-Through | | | | | 0.64** | 0.38** | 0.36** | 0.36** |
| | | | | | (0.09) | (0.09) | (0.07) | (0.07) |
| Follow-Through * Mathematics Teaching Quality | | | | | | | | -0.04 |
| | | | | | | | | (0.05) |
| Follow-Through * Number of Math Activities | | | | | | | | 0.04 |
| | | | | | | | | (0.11) |
| Observations | 563 | 563 | 563 | 563 | 876 | 876 | 876 | 876 |

*Note.* Standard errors clustered at school level (in parentheses). Mathematics teaching quality and number of math activities were measured using the COEMET. For each variable, scores were averaged from classroom observations in the kindergarten year. All models include controls for gender, ethnicity, age at preschool entry, mother's education level, free or reduced price lunch status, special education status at preschool entry, whether limited English proficient, and blocking assignment. The sample was restricted to students non-missing on all preschool, kindergarten, and first grade mathematics measures, as well as the classroom observational measure in kindergarten and first grade (COEMET). Analyses of outcomes based on classroom observation status available in Appendix E. + p<.10; * p<.05; ** p<.01

**3b. Sustained Classroom Environment – End of First Grade Math Composite**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| | End of Pre-k | Spring of 1st Grade | Spring of 1st Grade | Spring of 1st Grade | Spring of 1st Grade | Spring of 1st Grade | Spring of 1st Grade |
| Treatment | 0.67** | 0.16* | 0.15* | 0.17* | 0.18* | 0.17* | 0.17* |
| | (0.07) | (0.08) | (0.06) | (0.08) | (0.08) | (0.06) | (0.08) |
| Mathematics Teaching Quality | | | 0.02 | -0.02 | | -0.00 | 0.03 |
| | | | (0.04) | (0.06) | | (0.03) | (0.05) |
| Number of Math Activities | | | 0.14** | 0.20* | | 0.15** | 0.13** |
| | | | (0.04) | (0.08) | | (0.04) | (0.04) |
| Treat * Mathematics Teaching Quality | | | | 0.09 | | | |
| | | | | (0.09) | | | |
| Treat * Number of Math Activities | | | | -0.09 | | | |
| | | | | (0.09) | | | |
| Treatment with Follow-Through | | | | | 0.32** | 0.32** | 0.33** |
| | | | | | (0.09) | (0.07) | (0.07) |
| Follow-Through * Mathematics Teaching Quality | | | | | | | -0.12 |
| | | | | | | | (0.08) |
| Follow-Through * Number of Math Activities | | | | | | | 0.09 |
| | | | | | | | (0.11) |
| Observations | 563 | 563 | 563 | 563 | 876 | 876 | 876 |

*Note.* Standard errors clustered at school level (in parentheses). Mathematics teaching quality and number of math activities were measured using the COEMET. For each variable, scores were averaged from classroom observations in the kindergarten and first grade year. All models include controls for gender, ethnicity, age at preschool entry, mother's education level, free or reduced price lunch status, special education status at preschool entry, whether limited English proficient, and blocking assignment. The sample was restricted to students non-missing on all preschool, kindergarten, and first grade mathematics measures, as well as the classroom observational measure in kindergarten and first grade (COEMET). + p<.10; * p<.05; ** p<.01

Supplemental Online Appendix for

# Do High Quality Kindergarten and First Grade Classrooms Mitigate Preschool Fadeout?

**Appendix A: Classroom Activities Measures**

Appendix A presents each of the items from the HSIS kindergarten and first grade teacher surveys that were included in our measure of the instructional quality of literacy activities in the elementary classroom environments. We coded each activity as basic or advanced based on grade level. In the survey, teachers were asked how many times per week they did a given activity in their class (Never, Once a month or less, 2 or 3 times a month, 1-2 times per week, 3-4 times a week, Every day). For our analyses, we converted each activity into times per month by taking the mean value of the answer category multiplied by 4, and then standardized this measure to have a mean of 0 and standard deviation of 1.

**Appendix B & C: Testing for Selection into Kindergarten Classroom Environments**

Appendices B and C regress treatment status on various measures of subsequent classroom and school environments to test for differential selection. Note that the HSIS data contain a richer set of school and classroom covariates than the TRIAD study. However, both sets of results show that there is no evidence that children who were randomly assigned to preschool treatment were more likely to enter higher quality subsequent educational environments relative to children who were not assigned to the preschool treatment condition.

**Appendix D & E: Testing for Treatment effect differences by Kindergarten and First Grade Teacher Survey and Classroom Observation**

Appendix D tests for differences in the treatment impact of HS at the end of the preschool, kindergarten, and first grade years by kindergarten and first grade teachers survey response status (i.e., survey from which we created our measures of classroom literacy instruction). These analyses indicate that the treatment effect of HS varies slightly by the response status of study children's kindergarten teacher for impacts measured at the end of kindergarten only, and no systematic differences emerge. There is a marginally significant negative treatment effect at the end of kindergarten for children whose teachers responded to the kindergarten survey, but for these same students, the end of HS effect is positive and marginally significant.

Appendix E tests for differences in the treatment impact of TRIAD for students in the treatment and treatment with follow-through conditions based on whether their kindergarten and first grade classrooms had valid observational data. As Appendix C shows, the treatment effects (for both the preschool only and follow-through conditions) at the end of preschool were larger for students with observational data in kindergarten and first grade. Further, students without observational data also experienced more effect fadeout between preschool and kindergarten. These differences should be kept in mind when evaluating the main results of this study, as effect fadeout was even stronger for students excluded from the sample due to no observational data.

**Appendix F: Additional Treatment Effect Interactions**

Appendix F presents the results for models similar to those presented in Table 2a that test for moderating effects of attending full-day kindergarten, kindergarten class size, and the classroom-level proportion of children in poverty (as measured by Free and Reduced-price lunch). None of these variables, when interacted with treatment, were statistically significant.

The TRIAD scale-up evaluation did not collect information regarding whether the proportion of students in the kindergarten class that qualified for FRPL or whether classes were full-day or half-day. However, class size was obtained from the classroom observation. We tested an additional model (compare with models 2-4 of Table 3) that included a main effect for class size and an interaction between treatment and class size. We found a negative main effect for class size ($\beta$ = -.15, *SE*= .07, $p < .05$), and a positive interaction for class size and treatment

($\beta = .17$, $SE = .08$, $p < .05$), indicating that the treatment may have had a protective effect against the negative effect of larger classes on kindergarten outcomes.

## Appendix B. Head Start Impact Study: Selection into Kindergarten Classroom and School environments

| | (1) Total advanced literacy activities (times per month; standardized) | (2) Total basic literacy activities (times per month; standardized) | (3) Yrs. teaching exp. | (4) HS or below (Teacher) | (5) Some college (Teacher) | (6) Associates (Teacher) | (7) College (Teacher) | (8) College+ (Teacher) | (9) Pre-k teaching license | (10) Elementary teaching license | (11) Full-day K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Treatment | -0.029 | -0.100 | -0.803 | 0.002 | 0.010 | 0.001 | -0.012 | -0.000 | -0.025 | 0.001 | -0.052** |
| | (0.07) | (0.07) | (0.74) | (0.00) | (0.01) | (0.01) | (0.03) | (0.03) | (0.04) | (0.01) | (0.02) |
| Observations | 1075 | 1075 | 1062 | 1071 | 1071 | 1071 | 1071 | 1071 | 1003 | 1003 | 1008 |

| | (12) Classroom num. LEP students | (13) Classroom num. FRPL eligible | (14) Class size | (15) Teaching assistant | (16) Percent of school children black | (17) Percent of school children eligible for free/reduced lunch | (18) Percent of school children Hispanic | (19) Percent of school children white | (20) School proficiency level in math | (21) School proficiency level in reading |
|---|---|---|---|---|---|---|---|---|---|---|
| Treatment | 0.041 | -0.188 | -0.325 | -0.010 | 0.005 | 0.018 | 0.032 | -0.033 | -1.608 | -1.309 |
| | (0.50) | (0.45) | (0.34) | (0.03) | (0.02) | (0.02) | (0.02) | (0.02) | (1.55) | (1.85) |
| Observations | 1006 | 821 | 971 | 992 | 925 | 898 | 925 | 925 | 928 | 927 |

Coefficients and, in parentheses, standard errors are obtained from a regression of the column measure on treatment status. Standard errors clustered at center of random assignment. All models are weighted using inverse probability of treatment weights to adjust for differential attrition and complex sampling and include fixed effects for center of random assignment. Weights include all the baseline child and family control variables. + p<.10; * p<.05; ** p<.01

## Appendix A. Coding scheme for Instructional Quality of Literacy Activities in the Head Start Impact Study

| Kindergarten literacy activities | | First grade literacy activities | |
|---|---|---|---|
| Listen to stories with no print | basic | Activity related to book | basic |
| Show child how to read a book | basic | Write letters of alphabet | basic |
| Write own name | basic | Learn names of letters | basic |
| Teach directional words like over and up | basic | Have children tell you a story | basic |
| Write letters of the alphabet | basic | Practice sounds letters make | basic |
| Learn the names of letters | basic | Listen to stories w. print | basic |
| | | Read books chosen by child | basic |
| | | Read text w controlled vocab | basic |
| | | Read text w strong phonemic pattern | basic |
| | | Read patterned or predictable text | basic |
| | | Hear storytellers | basic |
| | | | |
| Discuss new words | advanced | Language activities in mixed achievement groups | advanced |
| Have children tell you a story | advanced | Discuss new words | advanced |
| Practice the sounds that letters make | advanced | Read aloud | advanced |
| Listen to stories with print | advanced | Read silently | advanced |
| Rhyming words and families | advanced | Work in reading workbook | advanced |
| | | Write words from dictation | advanced |
| | | Use invented spellings | advanced |
| | | Read thematic text | advanced |
| | | Compose stories or reports | advanced |
| | | Publish child's writing | advanced |
| | | Perform plays/skits | advanced |
| | | Write stories in journal | advanced |

Note: Literacy activities were coded into basic and advanced for grade level based by experts in early language and literacy development.

## Appendix C. TRIAD - Selection into Kindergarten Classroom Environments

| | (1)<br>Total math activities observed | (2)<br>Math instructional quality |
|---|---|---|
| Treatment | 0.01 | -0.06 |
| | (0.20) | (0.29) |
| Treatment with follow-through | 0.38+ | -0.04 |
| | (0.21) | (0.17) |
| Observations | 876 | 876 |

*Note.* Coefficients and, in parentheses, standard errors are obtained from a regression of the column measure on treatment status. Standard errors clustered at school level. Mathematics teaching quality and number of math activities were measured using the COEMET. For each variable, scores were averaged from classroom observations in the kindergarten and first grade year. The sample was restricted to students non-missing on all preschool, kindergarten, and first grade mathematics measures, as well as the classroom observational measure in kindergarten and first grade (COEMET). + p<.10; * p<.05; ** p<.01

## Appendix D: Head Start Impact Study- Fadeout Estimates by Kindergarten and First Grade Teacher Survey Response

| | (1) End of HS | (2) End of HS; K Teacher survey missing | (3) End of HS; K Teacher survey nonmissing | (4) Spring of K | (5) Spring of K; K Teacher survey missing | (6) Spring of K; K Teacher survey nonmissing | (7) Spring of 1st Grade | (8) Spring of 1st Grade; G1 Teacher survey missing | (9) Spring of 1st Grade; G1 Teacher survey nonmissing |
|---|---|---|---|---|---|---|---|---|---|
| Treatment | 0.16* | 0.16 | 0.17+ | -0.05 | 0.06 | -0.12+ | -0.05 | -0.10 | -0.08 |
| | (0.07) | (0.14) | (0.09) | (0.06) | (0.17) | (0.06) | (0.06) | (0.24) | (0.06) |
| Observations | 1632 | 589 | 1043 | 1525 | 448 | 1077 | 1313 | 248 | 1065 |

Coefficients and, in parentheses, standard errors are obtained from a regression identical to models 1 and 2 presented in Table 2, but with the sample inclusion criterion indicated by the column name. Standard errors clustered at center of random assignment. All models are weighted using inverse probability of treatment weights to adjust for differential attrition and complex sampling. Weights include all the baseline child and family control variables. + p<.10; * p<.05; ** p<.01

## Appendix E: TRIAD- Fadeout Estimates by Kindergarten and First Grade Classroom Observations

| | (1) End of Pre-k | (2) End of Pre-k; K Classroom obs. missing | (3) End of Pre-k: K Classroom obs. nonmissing | (4) Spring of K | (5) Spring of K; K Classroom obs missing | (6) Spring of K; K Classroom obs. nonmissing | (7) Spring of 1st Grade | (8) Spring of 1st Grade; G1 Classroom obs. missing | (9) Spring of 1st Grade; FG Classroom obs. nonmissing |
|---|---|---|---|---|---|---|---|---|---|
| Treatment | 0.61*** | 0.49*** | 0.67*** | 0.28*** | 0.10 | 0.32*** | 0.13+ | -0.06 | 0.18* |
| | (0.07) | (0.09) | (0.08) | (0.07) | (0.12) | (0.08) | (0.07) | (0.14) | (0.08) |
| Treatment with Follow-Through | 0.60*** | 0.42*** | 0.64*** | 0.37*** | 0.22+ | 0.40*** | 0.30*** | 0.17 | 0.32*** |
| | (0.07) | (0.07) | (0.09) | (0.08) | (0.12) | (0.09) | (0.07) | (0.12) | (0.09) |
| Observations | 1305 | 357 | 948 | 1218 | 270 | 948 | 1126 | 250 | 876 |

Coefficients and, in parentheses, standard errors are obtained from a regression identical to models 5 and 6 presented in Table 3a (Kindergarten) and model 5 in Table 3b ($1^{st}$ grade), but with the sample inclusion criterion indicated by the column name. Robust standard errors were clustered at the school level. All models included full baseline controls. + p<0.10 * p<0.05 ** p<0.01

## Appendix F. Additional Models Testing Moderation by Classroom and School Factors in the Head Start Impact Study: Kindergarten Literacy Composite Scores

| | (1)<br>Full-Day Kindergarten | (2)<br>Kindergarten class size | (3)<br>Classroom-level poverty (FRPL) | (4)<br>School-level poverty (FRPL) | (5)<br>School % Students Proficient in Reading | (6)<br>School % Students Proficient in Math |
|---|---|---|---|---|---|---|
| Treat | -0.21* | -0.35 | -0.09 | -0.27 | -0.24 | -0.30 |
| | (0.10) | (0.29) | (0.19) | (0.20) | (0.15) | (0.20) |
| Full-day K | 0.18 | | | | | |
| | (0.13) | | | | | |
| Treat * Full-day K | 0.22 | | | | | |
| | (0.16) | | | | | |
| Class size | | 0.01 | | | | |
| | | (0.01) | | | | |
| Treat * Class size | | 0.01 | | | | |
| | | (0.01) | | | | |
| Classroom % FRPL | | | 0.02 | | | |
| | | | (0.25) | | | |
| Treat * Class % FRPL | | | -0.30 | | | |
| | | | (0.26) | | | |
| School % FRPL | | | | 0.26 | | |
| | | | | (0.29) | | |
| Treat * School % FRPL | | | | -0.49+ | | |
| | | | | (0.29) | | |
| School % reading proficient | | | | | 0.00 | |
| | | | | | (0.00) | |
| Treat * School % reading proficient | | | | | 0.00 | |
| | | | | | (0.00) | |
| School % math proficient | | | | | | 0.00 |
| | | | | | | (0.00) |
| Treat * School % math proficient | | | | | | 0.00 |
| | | | | | | (0.00) |
| Observations | 1008 | 971 | 777 | 898 | 927 | 928 |

Coefficients, and standard errors in parentheses, presents the results for models similar to those presented in Table 2a that test for moderating effects of the column name in the HSIS. Standard errors clustered at school level. All models include fixed effects for center of random assignment and are weighted using inverse probability of treatment weights to adjust for differential attrition and complex sampling. Weights include all the baseline child and family control variables. Changes in observation counts across models reflect changes in teacher survey item non-response. FRPL-Free and Reduced-price lunch. + p<.10; * p<.05; ** p<.01