Heterogeneity in Public Preschool Efficacy: Evaluation, Student Growth Trajectories, and Approaches to Expansion

By

Sarah Kabourek

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Leadership and Policy Studies

August 31, 2019

Nashville, Tennessee

Approved:

Gary T. Henry, Ph.D.

Christopher A. Candelaria, Ph.D.

Mimi Engel, Ph.D.

Robin T. Jacob, Ph.D.

ACKNOWLEDGEMENTS

I am grateful for the Leadership, Policy, and Organizations faculty at Vanderbilt University's Peabody College. The department and faculty provided the necessary resources and support for me to complete my doctoral program. I would particularly like to thank Drs. Mimi Engel, Gary T. Henry, Christopher Candelaria, and Carolyn Heinrich. I have learned many invaluable lessons through working with you, and I appreciate the time you have taken as mentors, editors, and collaborators. Thank you as well to Dr. Robin Jacob, for pushing me to think carefully and rigorously about my work, particularly the methods and analyses used in these papers.

I would like to thank my family and friends for their support and encouragement throughout this process, for the many phone calls and care packages. I want to acknowledge my grandmother, Dr. Vivian Shapiro, who was a bus duty volunteer in the original Perry Preschool Project and always inspired me with her dedication to education. I finally want to thank my Team – I wouldn't be here without you.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

Chapter I. Persistence of Public Preschool Effects on Academic Outcomes? A Meta-Analysis

LIST OF FIGURES

Figure Page

INTRODUCTION

This dissertation probes the promise of public preschool as well as potential pitfalls in evaluation, provision, and access. Researchers and policymakers continue to point to the Abecedarian and Perry Preschool Projects, two mid-century behemoths in randomized education evaluation, as definitive proof of the cost-savings (typically first) and equity-enhancing (generally second) potential of preschool. However, the picture we get in the early twenty-first century is a much more nuanced and complex take on preschool as an educational, and social, intervention. While public preschool programs ranging from Head Start to local, school-based pre-kindergarten, show positive outcomes on school readiness measures, they have generally failed to deliver on measures that matter most in the current accountability era: namely, third grade reading and retention.

There are many reasons to provide publicly-funded preschool. Currently, the United States is the fourth-lowest OECD country for early childhood education enrollment, and one of the only OECD countries where preschool is not seen as an educational right for young children. Economically, preschool centers create jobs while allowing parents and family members to work full-time. Yet roadblocks to expansion in the United States continue, and rather than pre-kindergarten being absorbed into the K-12 system (as kindergarten classes were nearly a century ago), there is pressure to provide unequivocal evidence of short-, medium-, and long-term effects of preschool.

The essays in this dissertation ask questions related to the persistence and heterogeneity of preschool longitudinal effects, research and program design, and financing. The first study uses meta-analytic techniques to estimate the persistence of effects for public preschool programs from 1960 to present. Additionally, research design and contextual factors are explored

as predictors of end-of-treatment and longer-term effects. These include the use of experimental or quasi-experimental designs, comparison group activity, and study timing. The second study employs data from the kindergarten cohort of the Early Childhood Longitudinal Study – 2011 (*n* ~ 11,000) and longitudinal growth modeling to assess the associations between preschool participation and children's academic growth trajectories through third grade. Finally, the third study explores the potential use of Social Impact Bond financing mechanisms, where financing is tied to outcomes, to expand preschool programs. This qualitative study provides new data and perspectives from local city and state administrators regarding their goals for expansion, as well as the financial and political challenges they face.

Taken together, these studies provide new evidence on the heterogeneity of preschool impacts. The meta-analysis finds that newer, post-NCLB era programs have smaller effect sizes on average compared to their twentieth-century counterparts. More children are attending preschool, but access to quality programs is uneven, and academic and behavior-skills gaps between students attending Head Start, State preschool, and private preschool programs are observable at kindergarten entry and persist through third grade, as shown in my second study. Finally, some cities and districts with existing public-private partnerships and strong fundraising capacity are able to explore options for new financing mechanisms to support expansion, while others continue to struggle with braiding and blending together multiple funding sources to support universal access. These studies provide several new answers to existing questions, and pose many more for preschool researchers and school leaders.

CHAPTER 1

PERSISTENCE OF PUBLIC PRESCHOOL EFFECTS ON ACADEMIC OUTCOMES? A META-ANALYSIS

Introduction

While existing research shows that early childhood interventions can have lasting effects, evidence regarding the longer-term efficacy of public preschool is mixed. Current local, state, and federally funded preschool programs are highly varied in the populations they serve, services provided, funding mechanisms, curricular programming, and ultimately, short- and long- term student outcomes (Bitler, Hoynes, & Domina, 2014; Chaudry et al., 2017). While many programs show positive, immediate effects among participants at kindergarten entry (Burchinal et al., 2015; Camilli et al., 2010), substantial evidence suggests that these effects are not sustained over time (Bailey et al., 2017; Phillips et al., 2017). This decline in early childhood program effects over time is often referred to as "fadeout." A recent meta-analysis, including studies from 1960 to 2006, found that preschool academic gains fadeout or converge with comparison group outcomes at a rate of approximately 0.022 standard deviation units per year (Li et al., 2016). Achievement convergence is one main explanation as to why fadeout seems to occur. In this explanation, students who did not attend preschool simply catch up to their peers once they attend kindergarten. For example, in Clements et al.'s (2013) TRIAD evaluation of Building Blocks math intervention, a randomized controlled trial, the treatment group made significant gains in preschool, testing nearly 0.5 standard deviation units higher than control students at the end of pre-kindergarten. However, in the post-intervention period, comparison group and treatment student test scores converged by the end of first grade.

Understanding the nature and extent of effect persistence is relevant in addressing policy questions regarding publicly financed preschool. One perspective holds that preschool is an important public investment, regardless of longer-term, sustained gains (for example as described in Jenkins, 2014). In this line of thinking, preschool provides an opportunity for children to be exposed to the social setting of school while allowing their primary caregivers an opportunity to participate in the workforce. Furthermore, there is a substantial base of evidence that shows that public preschool programs reduce the academic skills gap between racial subgroups that has been evident at kindergarten entry (Phillips et al., 2017). There is also evidence that both public and private early childhood education experiences can substantially support positive high school and adult outcomes, even if the initial cognitive effects fade over time (Barnett, 1995; McCoy et al., 2017). Still, researchers and policymakers have questioned whether a significant investment in preschool is worthwhile when considering the relatively fast fadeout of effects (Samuels, 2018). It is not yet clear if the mixed results from public preschool evaluations can be attributed to specific program elements, generalizability across locations and populations, features of individual study designs, or some other set of conditions. Even terminology is not clear, with "fadeout," "convergence," or "persistence" often used interchangeably. However, the growing evidence base of preschool evaluations provides an opportunity to explore the potential influence of these factors on observed short-, medium-, and long-term student outcomes. In the current study, I aim to explore the *persistence* of preschool effects over time.

Literature Review

Meta-Analysis: Synthesizing Preschool Studies

Given the large number and wide variety of preschool evaluations in the United States, some scholars have moved to using meta-analytic techniques to organize and assess existing evidence. Meta-analysis uses statistical methods to pool quantitative results from multiple research studies. A potential benefit of using meta-analysis is that results can provide more compelling evidence than those from a single primary study. Furthermore, variation between studies can be leveraged to study questions about the association between factors such as research design, intervention implementation, timing, and aggregated effects. A 2010 meta-analysis by Camilli et al. included 123 studies of early childhood interventions, and observed outcomes including intelligence, achievement, and social/emotional skill assessments. Overall, the authors found statistically significant effects for studies with a treatment/control comparison on cognitive outcomes (average unweighted effect size of 0.231 standard deviation units), and some effect of reduced retention rate (average unweighted effect size 0.137) and social-emotional skills (0.156) (Camilli et al., 2010). More recently, a group of scholars have compiled and coded a database of early childhood education (ECE) studies fielded between 1960 and 2006 (The National Forum on Early Childhood Program Evaluation, hereafter referred to as the ECPE Forum). The ECPE Forum has published several meta-analytic studies exploring potential mediators and moderators of preschool effects. These papers have analyzed the role of child gender, child age at time of treatment, and parent education on short- and long-term cognitive, academic, and behavioral outcomes (Magnuson et al., 2016; Li, et al., 2016; Grindal et al., 2016; McCoy et al., 2017; Schindler et al., 2015; Shager et al., 2013).

Only a portion of the studies included in these meta-analyses have medium- to long-term outcomes, and none of the published meta-analyses explicitly analyze convergence. A 2016 meta-analysis studies the relationship between starting age and program duration on the impacts

of early childhood education programs on cognitive and achievement outcomes, as well as fadeout (Li et al., 2016). The study uses data from 67 studies and 1,045 effect sizes of achievement outcomes, from studies completed by 2006. Initial, end-of-program effect sizes averaged 0.23 standard deviations. This initial effect fades linearly at approximately 0.022 standard deviations per year beyond the program duration (Li et al., 2016). The authors find that starting programs earlier is associated with better outcomes at the end of pre-kindergarten—increasing program starting age by a year is associated with a -0.123 standard deviation per year decrease in end-of-program effect sizes, regardless of length of program. In other words, attending preschool at a younger age is correlated with higher end-of-treatment effect sizes. The study finds no statistically significant interactions between program length and fadeout, or program starting age and fadeout. Thus, although programs targeting infants had a larger effect, these effects declined over time in a similar pattern to programs which included "older" children. This study provides clear evidence of a "fadeout effect," and starts to test some potential predictors of fadeout. The authors do not uncover a significant interaction effect between timing variables (i.e., length of treatment) and fadeout effects; the lack of significant explanatory variables does not foreclose the possibility that other variables such as methodological variation may explain differences or that studies completed after 2006 may add new evidence and greater statistical power to detect associations.

There is reason to believe that research design plays a strong role in the mixed findings of preschool evaluations. In a 2013 meta-analysis using ECPE data, Shager et al. specifically addressed questions regarding design components of Head Start evaluation studies, and how those components were related to estimated effect sizes. Using thirty years of Head Start evaluations, the authors coded for design elements including: baseline equivalence, control group

activity, and outcome measures (including what type, reliability of measures, and how assessments were completed). The authors found that the activity level of control group children was a significant predictor of effect size. In studies where the control group actively sought alternative services (aside from Head Start), effect sizes were smaller on average (0.08) than studies in which the control group did not seek alternative treatment (0.31 on average). Furthermore, authors found that outcome features, including measures of concrete academic skills (math and literacy versus broader cognitive skills) and ratings and observations (versus performance assessments), produce larger average effect sizes. Overall, the researchers find that Head Start produces a statistically significant effect size of 0.27 for short-term (less than one year posttreatment) cognitive and achievement outcomes. Control group activity and outcome type/measurement accounted for 41% of the heterogeneity in findings between studies, and 11% of the variation within studies. A recent, broader study found evidence for the influence of methodological factors in studies of reading, mathematics, and science programs during ECE, elementary, and secondary school (Cheung & Slavin, 2016). The authors found effect sizes were nearly twice as large for published articles, small-scale trials, and experimenter-made measures (Cheung & Slavin, 2016). Finally, in addition to overall effects, Camilli et al. explored the influence of various program and research design elements, including comparison group activity (2010). The authors coded and ran analysis in two categories – studies that had a treatment/alternative treatment comparison, and studies with a treatment/control comparison. The treatment/alternative comparison included students who received alternate preschool, including private or non-center based (but academic) care. The treatment/control comparison included students who did not receive any preschool. The meta-analysis using the treatment/alternative comparison group did not have statistically significant effect sizes at the

end of treatment. However, as described above, the authors do find significant effects of preschool for studies with a treatment/control group in cognitive and social-emotional outcomes (Camilli et al., 2010). This again suggests that design-related factors such as these should be considered when trying to aggregate and compile evidence on specific interventions.

Research Design: Defining the Counterfactual in Public Preschool Evaluations

The Shager et al. (2013) and Camilli et al. (2010) studies highlight the difficulty with interpreting and generalizing across preschool studies. In their 2013 meta-analysis, Shager et al. note that control group, or counterfactual, activity level is an often overlooked ECE evaluation design feature. They define this as the level of "participation in center-based care or preschool among control group children" (Shager et al., 2013 pp. 78). Theirs was the first study to empirically address the extent to which activity level in the comparison group predicts the magnitude of program effects in a meta-analytic study of Head Start evaluations, finding a 0.23 effect size difference between comparison groups that did and did not seek alternative services. The question of how to define the counterfactual in social sciences is ongoing and relevant to evaluating preschool programs (Lemons et al., 2014; Feller et al., 2016; Bitler, Hoynes, & Domina, 2014).

There are two main ways preschool studies are conducted. One, researchers design a study that takes advantage of a natural experiment, such as a lottery for an oversubscribed program. Two, researchers design a quasi-experimental study using longitudinal data, where they utilize statistical techniques to create an artificial counterfactual, or comparison group. In either case, children who do not participate in the specific treatment program may still receive a range of alternative treatment - attending a different public preschool program, private preschool, center-based care (with or without an academic component), relative or other home-based care.

A study using data from the randomized Head Start Impact Study finds that, when separating groups of control children into those who attended other center-based care and those who stayed home, there are positive, strong effects of Head Start compared to home-based care children, but not for those in other center-based care (Feller et al., 2016). Not only does alternative treatment need to be considered, but more general contextual factors, including time and place, matter for evaluating evidence (Lemons et al., 2014).

Timing: No Child Left Behind, the Great Recession, and Public Preschool

Lemons et al. present evidence from an original and four replication studies (five total randomized control trials) to support the argument that time and place are critical for interpreting experimental and quasi-experimental research, in particular that the experiences of the counterfactual, whether controlled or business as usual, can increase or decrease effect sizes (2014). In other words, the treatment effect is always relative to the outcomes of the control or comparison sample and many changes in social norms and practices as well as the policy environment can cause the counterfactual outcomes to vary. For example, the changes due to the enhanced accountability in the era of No Child Left Behind (NCLB) may affect public preschool counterfactual outcomes.

In terms of public preschool, there is a pressing need to provide an update on the efficacy of early childhood programs in the newest accountability era. There are several reasons why we might expect differences or new information in studies published since the turn of the century. First, preschool programs in general have expanded—between 1995 and 2014 the percent of children ages three and four enrolled in center-based care increased from 49% to 55% overall; the percentage of Latino children enrolled increased from 37% to 45%, and increased from 48%

to 58% for black students (Phillips et al., 2017). Moreover, the percentage of children in the lowest family income quartile attending preschool increased from 33% to 49% during this time period (Phillips et al., 2017). These students have historically experienced a racial and income-based achievement gap as early as kindergarten (Bassok et al., 2016). Yet, having higher participation rates in preschool means that there are more students entering kindergarten with preschool experience, even if they were not included in an evaluation treatment group. This may narrow gaps between low- and high-income students, as well as potential gaps among low-income students who do and do not receive specific intervention treatment. This may make it more difficult to detect treatment effects. Second, many of the studies published since 2002 are longer-term follow-up studies to those published previously. In addition to being able to observe persistence, other changes may include a reduced contrast between comparison groups, as more students are attending preschool or academic, center-based care (Phillips et al., 2017). Further, we might expect changes in the effect estimates of more recent evaluations, due to changes in methodological approaches to social science research and the changing nature of preschool programs in the twenty-first century. Social scientists have shifted considerably towards exploring the use of quasi-experimental designs, as well as a general standardization of outcome measures, that support more plausible causal inferences (e.g., Bifulco, 2012; Cook, Shadish, & Wong, 2008; Lemons et al., 2014; Shadish, Clark, & Steiner, 2008).

Finally, Stipek (2006) argues that 2002 NCLB mandates have expanded the development of preschool standards from fewer than half the states to nearly all. Language in the 2007 Head Start reauthorization bill specifically recommended preschool alignment with K-12 content standards. Subsequent changes then could also be attributed to the NCLB "read-by-three" mandate, wherein states were incentivized to measure and reach total student reading proficiency

by third grade, and significantly expanded state preschool programs, K-2 testing, and the development of prekindergarten-elementary aligned curricula in order to meet federal accountability goals (No Child Left Behind [NCLB], 2002; Stipek, 2006; Stipek et al., 2017). However, preschool expansion was put on hold in the years during the great recession, during which time state preschool budgets declined, requiring federal stimulus funding to maintain enrollment levels (Hustedt & Barnett, 2011). State preschool budgets have since rebounded, but many states are carefully evaluating the efficacy of their early interventions. In at least one case where state preschool did not exist, a public-private financing scheme was set up to provide the highly popular program, with the state committing to pay only if the program was able to decrease expected need for special education services (Innocenti, 2015). It may be that the latest iteration of accountability initiatives that immediately preceded a severe financial crisis for states, have changed the goals, features, and policy landscape for public preschool programs. Therefore it is important to update the research base with current studies; additionally, evaluations should include detailed information to the extent possible regarding program and evaluation design, particularly comparison group activity. Additionally, understanding short-, medium-, and long-term programs can help support decision making and intervention planning at the preschool and K-12 levels.

Contribution: Addressing Gaps and Expanding Evidence

This study asks three main questions: (1) To what extent do the effects on academic achievement of participating in public preschool persist over time?, (2) To what extent are study design features associated with preschool effects?, and (3) To what extent are there differences in estimates before and after NCLB? Research design characteristics in this study include the

selection of a counterfactual group (whether an active or passive treatment group), type of assignment to treatment (randomized control trial or quasi-experimental), and level of rigor in addressing baseline equivalence and study attrition. To address these questions I conduct meta-analyses of evidence from studies produced from 1960 through 2018 that include longer-term student outcomes for children who attended publicly funded preschool. A substantial evidence base documents the decline of preschool effects as children progress through school (e.g., Bassok, Gibbs, & Latham, 2018; Bailey et al., 2017; Li et al., 2016; Lipsey, Farran, & Hofer, 2015a). The current study builds on this work by conducting a systematic analysis of total effects, expanding evidence through 2018, as well as testing potential moderators based on research design characteristics and a new wave of accountability policies. The study contributes updated evidence on the potential influence of design elements that might contribute to the variation in longer-term preschool effects from prior studies as well as assessing the role of federal accountability policies.

Methods

Research Questions

The current study addresses the following research questions:

> *RQ1: To what extent do the effects on academic achievement of participating in public preschool persist over time?*
> *RQ2: To what extent are study design features (research design type and comparison group activity) associated with preschool effects?*
> *RQ3: To what extent are there differences in estimates in pre- and post-NCLB studies?*

In order to study the persistence of effects, the current study is limited to academic and cognitive outcomes, which can be measured repeatedly throughout a child's time in and out of school. Point-in-time, binary measures, such as graduation rates, time of special education receipt, or

involvement with the justice system, are critical for understanding the full cost-benefit analysis of public preschool, but cannot tell us much about the persistence of direct preschool effects. Therefore, as a starting point, this study limits analysis to academic outcomes only. Research or study design features, as defined for this study, refer to the selection of counterfactual group, process of assignment to treatment, and methods to address baseline differences and attrition throughout the study. In the next section, I will review the eligibility criteria, search strategy, data collection process, and analytic strategy for this study.

Eligibility Criteria

This meta-analysis seeks to assess the persistence of effects and, hence, the longer-term impacts of public preschool on academic outcomes. Using a common definition from current meta-analyses on early childhood education, I will include studies on early childhood education (ECE) programs, defined as "structured, center-based early childhood education classes, day care with some educational component, or center-based care" (Li et al., 2016, p.12). The study will include programs serving three-to-five year olds, in the one or two years immediately prior to kindergarten. This may include programs referred to as "preschool" or "pre-kindergarten." Long-term, as defined for this study, includes outcome(s) measured at end of kindergarten through adulthood. Furthermore, the outcome must have been measured multiple times – at least twice after treatment – in order to measure the persistence of effects.

This analysis includes studies of ECE programs that have experimental or quasi-experimental designs (QED). In order to be included in the meta-analysis, in addition to having a well implemented RCT or QED, the intervention must include public preschool based on either universal eligibility, or selected neighborhood, family, or child criteria, and include child-level outcomes measured at least once beyond kindergarten entry. Studies where there was no

comparison group, where the intervention involved the testing of medical procedures or other health-related products, if eligibility was based on student disability status, and studies conducted outside of the U.S. were excluded from the meta-analysis (Table 1). I include both experimental and quasi-experimental studies (the majority of studies are quasi-experimental, as often is the case in education research). Studies included in the screening have at least one reported academic or cognitive measure, collected at minimum two times after treatment. Exclusion criteria include studies that are testing a pharmaceutical intervention, studies that do not include a control or comparison group, and those evaluating interventions specifically targeting children with disabilities.

Search Strategy (and Study Selection)

An electronic search was conducted to identify all peer-reviewed articles, policy reports, federally and state-funded studies, conference proceedings, working papers, and unpublished dissertations from 1960 to 2018 that examine the long-term effects of ECE participation. Table 2 provides a list of strategies for searching the literature; Figure 1 shows the PRISMA screening flowchart for the study screening process. The list encompasses databases and strategies typically used in meta-analyses, including the use of ProQuest and ERIC, conference proceedings, grey literature sources, and expert scholar consultations.

Data Collection and Coding

Data from each report (n = 26) was collected to generate effect sizes for each study (k = 13). I code each study for any academic or cognitive outcomes, and the following design factors: type of measure, student assignment, QED, testing and/or adjustment of baseline group equivalence and attrition throughout the study, and timing of the initial intervention and follow-

up (Table 3). In collecting outcome data I extract raw and standardized means, and variances for all effects presented in reports, and generate standardized mean difference effect sizes for each assessment period. In the case of multiple reports per study (typical for longitudinal studies), I use report references to collect all reports associated with the study, in order to collect data from each available time period. I then calculate a standardized mean difference effect size for each outcome within studies. Finally, I chose to separate cohorts of students within a program if they had a different treatment experience. For example, with the Abecedarian program, results are separated by students who experienced Abecedarian preschool only versus those who experienced the preschool program and the kindergarten through second-grade follow up. Students in New Jersey's Abbott preschool program experienced one or two years of preschool prior to kindergarten. As described earlier, there are reasons to believe that exposure to follow-through (Bailey et al., 2017) or longer periods of intervention (Li et al., 2016) will result in larger effect sizes.

Analytic Strategy

The studies selected for this analysis include longitudinal, empirical analyses of the effects of ECE on later student outcomes. The standardized mean difference, calculated during data collection, is considered comparable across studies (Borenstein et al., 2009). The within-study standardized mean difference is an estimate of Cohen's *d* effect size. It is appropriate to use a correction producing an estimate referred to as Hedges' g, as Cohen's d may be upwardly biased in small sample sizes; I follow this procedure for the current study. I use a random effects model assuming some true variance in effect sizes—some programs will be more effective than others and have real differences in overall effect (Borenstein et al., 2009). By design, all studies have multiple effect sizes, across domain and time period. I conduct separate random effects

meta-analyses for each type of outcome: literacy, mathematics, and general cognition. To account for correlated effect sizes, I use a robust variance estimator, described below.

There is not alignment across studies in terms of how often and how many times follow-up assessments occurred. In order to address differences in frequency, I construct two methods for standardizing follow-up periods. First, I run the meta-analyses described above at the end of treatment, kindergarten, first, second, third grade, and "older." Studies assessed in a given grade level are included as appropriate. Second, I categorize timing in terms of follow-up "waves"; this standardizes longitudinal assessment periods and maximizes use of study results by allowing for broader inclusion of effect sizes. End-of-treatment assessment is coded, and the first assessment time beyond that is "follow-up 1," the next "follow-up 2," and so on. In this structure, an assessment that occurs during "follow-up 1" may happen in kindergarten or first grade, depending on the design of the study. This secondary method of standardization across original studies addresses persistence across waves, while allowing for flexibility in when follow-up assessments occur. I run post-hoc F-tests to test for differences between estimates at end of treatment and each follow-up point. These results are shown in Table 12.

When using the random effects model, we can attempt to isolate the variation in true effects by estimating several statistics to test for heterogeneity. The *Q*-statistic tests whether observed differences in effect sizes are consistent with what we would expect due to sampling error alone. A rejection of this test suggests that there is "true" heterogeneity across studies, and a random effects model is appropriate. The I^2 statistic quantifies inconsistency in effect sizes across studies by providing a signal to noise ratio. Finally, the τ^2 statistic is the estimate of the variability of the true effect sizes around the mean of the distribution; if we assume multiple true

effect sizes in the population, τ^2 gives us an estimate of the distribution of these effect sizes. I use each of these to quantify heterogeneity in the analysis.

In addressing the second research question, I run multiple meta-regressions for outcomes at the end of treatment and for all combined follow-up periods, with research design features as predictors.

$$y_{ji} = \beta_0 + \beta_1 x_1^v + u_{ji} + e_{ji}$$

Where *y* indicates all outcomes for study *j* at time *i* (end of treatment, follow-up periods, and all time periods), and *x* indicates each predictor *v* (RCT, PSM, attrition, baseline equivalence, and comparison activity). So that, for example, in Table 11, Row 1, the coefficient estimate represents the association between the study being an RCT, and aggregate study outcomes at the end of treatment. This bivariate meta-regression is then completed for each of the following predictors: the use of propensity score matching (Row 2), testing and adjustment for attrition over time (Row 3), baseline equivalence (Row 4), and comparison group activity (active or passive, Row 5). Finally to address research question three, I run an additional bivariate meta-regression using a predictor indicating whether the study intervention took place prior to or after 2002. Due to the correlated effects structure of the longitudinal data, I use a robust variance estimation method (Hedges et al., 2010; Tipton, 2016). This estimation procedure adjusts the variance estimate to account for within-study effect size covariance by using within-study residuals. Therefore, the procedure can account for correlated effects without knowing the underlying correlation of dependent variables in the original study, which is often unpublished (Tipton, 2016). This approximation holds for small sample sizes. I utilize the *robumeta* command in Stata throughout the analysis; this command uses robust variance estimation (Hedberg, 2011; Tanner-Smith & Tipton, 2013).

Publication Bias

Publication bias is a concern when conducting a meta-analysis; it has been shown that studies with relatively higher effect sizes are more likely to be published than studies which report lower effect sizes (Borentstein et al., 2009). However, meta-analysis only provides a mathematically accurate synthesis of study effects if there is an unbiased sample of all relevant studies. Since it is easier to find studies published in peer-reviewed journals than unpublished or self-published research, it is important to consider the extent to which meta-analytic results may be biased due to the original sample of studies. Given that many preschool evaluations are conducted at the state level for specific programs, these are often not submitted for publication in a peer-reviewed journal, which is another potential source of bias. In the current study I was careful to consider grey literature and consult with leaders in the field in order to try and collect all available information. There are several standard ways to observe or test for publication bias in a meta-analysis. In this study I use two – a funnel plot and Egger test. A funnel plot is a graph used to display the relationship between study size and effect size (Borentstein et al., 2009). Individual study effect sizes are plotted with effect size on the x-axis and the sample size or variance on the y-axis (in this study, variance estimates will be plotted). The funnel plot display relies on visual inspection to assess potential publication bias; ideally there is symmetry across the x- and y-axes, indicating a representative sample of effect sizes across studies. If publication bias is present, effect sizes may be clustered on the far right (positive) side of the x-axis, indicating that only studies with large, positive effects are published. The Egger test provides an empirical test for funnel plot symmetry.

Extensions of Main Analysis

I conduct several post-hoc analysis extensions; first by re-running models from the first research question with more precisely defined assessment measures. In the main analysis I categorize assessments as "Literacy," "Math," or "General Cognition," despite the use of different measurement tools. For example, both the WJ-III Applied Problems subscale and the WJ-III Calculation subscales measure mathematics skills and are often used in combination. Yet, combining WJ-III subscales and a state-administered mathematics assessment may potentially introduce an "apples to oranges" comparison. There is enough overlap of assessments across studies to run meta-analyses with three specific assessment tools: the Peabody Picture Vocabulary Test, WJ-III Letter-Word subscale, and the WJ-III Applied Problems subscale. I expect results from these meta-analyses to follow similar patterns to the larger "Literacy" and "Math" categories, however, significant deviations may suggest that this collapse is inappropriate for a clear analysis. These results are shown in Appendix A. Next, I run post-hoc analyses with several different sample cuts to explore the sensitivity of results to sample composition (these results are shown in Appendix B). In these analyses I first restrict the sample to studies that have assessments at each follow-up point. I then re-run the main outcomes analysis and secondary predictor analysis with the sample restricted to either a) pre-NCLB studies or b) post-NCLB studies.

Results

Study Selection

Based on the search strategy described above, I identified a total of 7,788 records through ProQuest and ERIC database searches, ProQuest Dissertation and Theses, and a review of grey literature (including NAEYC, SRCD, RAND, and colleague consultation). After removing

duplicates I screened 6,218 records and excluded 6,022 as they were not longitudinal studies. I reviewed 106 full-text articles against the pre-determined eligibility criteria (Table 1). Of the 106, the final meta-analytic sample included 26 reports and 13 studies. The search and screening process is shown in Figure 1. To be included in the final sample the study needed to provide multiple post-treatment effect sizes on the same construct, as multiple follow-ups are required to address the question of persistence. Furthermore, binary outcomes such as special education receipt, grade retention, and high school completion were excluded, as these event histories do not change over time for students. This limited the number of included studies – a tradeoff of power to address the specific research questions identified.

Study Characteristics

Descriptive characteristics of final included studies are shown in Table 4, and include public preschool evaluations of programs beginning in 1962 to present. The studies are anchored by the quintessential Abecedarian and Perry Preschool programs, and include evaluations in each decade from 1960 to 2010. Study size ranges from 43 to 677 participants, have two to five follow-up periods (including end of treatment assessment), and follow children from ages three to fifteen. The preschool programs evaluated include small-scale experimental programs (Perry Preschool, Abecedarian, Howard), the national Head Start Impact Study, and statewide programs (Tennessee Voluntary Pre-K, NJ Abbott). Studies were coded for effect size as well as treatment measure type, name, form of assignment to treatment or control conditions, type of QED, whether treatment take-up, baseline equivalence, and attrition were tested, and the time elapsed since treatment. In terms of assignment to conditions, 46% of studies used a randomized control trial (RCT), while the remaining 54% used a quasi-experimental design (QED). Of the QED studies, the majority used propensity score matching (PSM). Aside from PSM, there was no

other QED category that was used by more than one study. Therefore I am only able to test for RCT or use of PSM as predictors. For research design factors, 68% of studies tested (and if necessary adjusted for) baseline equivalence at the time of assignment to treatment and 58% tested (and adjusted) for attrition from the study over time. Overall, 76% of studies had a "business as usual" comparison group (as opposed to a strict no-preschool comparison group). The average age at the final posttest across studies was 128 months (about 10.5 years), with an average of 78 months since treatment at the final posttest (6.5 years).

Results of Individual Studies

On average, studies had moderate to large, positive effects at the end of treatment (ranging from -0.5 to 1.12). Overall, scores were higher on measures of general cognition (average standardized effect size 0.212) than in mathematics (0.10) and literacy (0.06). Figure 2 shows a distribution of study effect sizes across time (from pre-treatment to tenth grade), averaged across domains. The figure shows that effect sizes narrow in range over time, and cluster toward zero the further out the follow-up. This could suggest a decline of effects, or a regression toward the mean. In only one instance, with TN-VPK, are effects consistently negative (though small) throughout follow-up periods. Earlier, smaller studies (i.e., Even Start, Howard, Perry) have the largest and most persistent effects over time. This reflects the current (post-NCLB) research base, which shows more mixed evidence on medium- to long-term efficacy of public preschool. While these studies are equally likely to have "business as usual" comparison group activity (potential preschool outside of treatment), preschool options were likely more limited in earlier years (see Figure 3, and Phillips et al., 2017). Still, without more detailed information on counterfactual group experience, we cannot say for certain that there is a

higher likelihood of preschool participation post-NCLB versus pre-NCLB. Finally, Table 5 provides a list of included programs, their starting year, and the availability of assessments at the end of treatment and at follow-up. In the last row of the table, *k* indicates the number of studies included for each model. This changes depending on whether reports provided this assessment in each period. For example, in the Abecedarian studies, the WJ-III was not given until beyond kindergarten. In the evaluation of the Child Parent Center, a general cognition assessment was given at the end of treatment but not follow up. In Michigan, the PPVT was assessed at the end of treatment but not follow-up. Finally in Arkansas, the WJ-III reading subscales were not assessed until follow-up. Due to the inconsistency in measure availability, it is difficult to systematically compare persistence in effects across programs. However, there are enough measures in early elementary school to make meaningful comparisons.

Synthesis of Results

Main Effects and Persistence Over Time

Main effects for research question one are shown in a series of Tables 6 through 11. For each outcome (Literacy, Math, and General Cognition), there are two associated tables. The first shows end-of-treatment effects followed by meta-analyses run for outcomes in Kindergarten, first, second, third grade, and "older"; the second table shows end-of-treatment effects, with follow-up "waves" as described previously.

Beginning with Literacy in Tables 6 and 7; the average effect at the end of treatment is 0.181 standard deviation units, as measured across 8 studies. This effect diminishes slightly (to 0.101 standard deviation units in Kindergarten, or 0.095 units in the first follow-up "wave"). In the last time period, including literacy scores beyond third grade (in both Tables 6 and 7), the

long-term impact of preschool, on average across studies, is approximately 0.175 standard deviation units. It is important to note that these longer-term impacts are estimated with select studies that followed students through high school (or beyond). I describe the results with a more restricted sample below, which includes only those studies with consistent measures through each time period. A χ^2 statistic testing heterogeneity indicates that random effects analysis is appropriate ($p < .001$). The I^2 statistic decreases over time (from 83.7% to 56.7%); this suggests that while there are likely observable, testable moderators immediately after treatment, over time more of the variance is random error. Finally τ^2 ranges from 0.213 to 0.0064; the smaller τ^2 estimate, the tighter the estimated distribution of variability of true effect sizes. The smaller τ^2 values appear further out in time, suggesting smaller variation in longer-term effects. Since these effects are also precisely estimated ($p < .001$), this is likely a precise estimate of longer-term effects of public preschool on literacy outcomes.

Tables 8 and 9 show results for math outcomes. The average effect at the end of treatment is 0.191 standard deviation units, measured across seven studies. This effect is statistically significant; however, in the kindergarten, second and third grade follow-ups (as well as the first, third, and fourth "waves"), there is not enough power to detect an effect. The first grade sample estimates reflect a small, positive aggregate effect (0.083 standard deviation units). Similar to literacy results, the smaller sample of studies with longer-term measures reflects strong long-term effects (0.186 standard deviation units). In terms of variance estimates, a random-effects estimator is an appropriate choice ($p < 0.05$), except for the fifth follow-up in Table 9. However, with only three studies this may be due to imprecision rather than indicating a lack of heterogeneity. Similar to literacy outcomes, the I^2 statistic is moderate to large across waves (range 29.7% to 86.3%), suggesting moderator analysis is appropriate. Post-hoc F-tests

show that estimates are significantly different from end of treatment to kindergarten (for literacy, Table 12), end of treatment and first grade (math), end of treatment and third grade (math and general cognition). Post-hoc *p*-values for F-tests are shown in Table 12. The relative consistency of results could be due to persistence of effects, or simply large confidence intervals given the relatively small sample size.

Results for general cognition are in Tables 9 and 10. The end of treatment average effect is larger than that for math or literacy, although still only a moderate effect, of 0.396 standard deviation units. Using grade-level follow-ups, there is a detectable effect in first grade less than half of the end of treatment effect (0.163 standard deviation units). Considering follow-up "waves," there is a detectable effect in the first follow-up (still rather short-term), of 0.251 standard deviation units. Overall it appears that there is insufficient power to detect changes in general cognition effects over time.

Research Design Moderators

Tables 13, 14, and 15 show the association between study design features and treatment effects, separated by time period and then aggregated in Table 14. Coefficients range from -0.169, a small negative association, to 0.095, a slight positive association. However, none of the point estimates are statistically significant. This could be due to a lack of power or variance across effects, but is somewhat surprising given the clear association between design and effect sizes found in previous studies (e.g., Shager et al., 2013; Li et al., 2016). Table 16 presents results for the third research question, testing the difference between studies conducted prior to versus after NCLB. There is a statistically significant association between average effect sizes and timing, with a coefficient of approximately -0.18. In other words, studies conducted after

2002 had, on average, a 0.18 smaller standard deviation unit effect size. This is significant considering the magnitude of effects considered throughout the study (ranging from -0.51 to 1.13 raw treatment effects, and ranging from -0.008 to 0.396 for meta-analytic results).

Risk of Bias Across Studies

Several funnel plots are shown in Figures 4-6. All plots show standardized effect sizes (measured as Hedges' g) and variances against a 95% confidence interval. In observing these plots, we are looking for visual symmetry to assess potential publication bias. Figure 4 shows a standard funnel plot with all effect sizes represented by a blue dot marker. There is a cluster of effect sizes around zero on the x-axis (effect size g) and between 0 and 0.1 on the y-axis (standard error of g). This indicates many precisely estimated zero effect sizes. There is a gap in the bottom left quadrant of the plot, indicating fewer reported negative effect sizes with large standard errors. This is confirmed with the Egger test of funnel plot asymmetry, which could not reject the null hypothesis that sample size is associated with effect size ($p=0.226$). This indicates some publication bias, where larger studies with greater positive effect sizes are more likely to be published and observed in the meta-analysis. Figure 5 separates peer-reviewed and non-peer reviewed studies with separate markers; studies in peer reviewed journals are shown with a red triangle marker, while unpublished studies are represented by a blue dot. Here the asymmetry is even more apparent, with published studies more likely to have positive effect sizes (although not necessarily measured with more precision). Figure 6 shows a funnel plot with different markers for each study. This provides a visual inspection of effect size and precision by study, and is helpful for observing the distribution of effects. For example, comparing even NJ Abbott preschool with one- versus two-years of treatment, the studies have similar precision, but the

effect sizes for the two-year treatment have larger effect sizes (one-year Abbott marked with blue dots, two-year Abbott with red triangles). This reflects prior research that earlier intervention supports longer-term results (Li et al., 2016). Observing effect sizes for HSIS (blue diamond) and TN-VPK (pink square), effects of these studies are clustered neatly around zero, with precise estimation. However, an aggregate effect size for each study masks the moderate positive, and then negative effects over time. Studies that appear in the lower left quadrant (negative effect sizes with large standard errors) – the Third National Even Start evaluation and the Howard preschool experiment – were not published in peer-reviewed journals, but published through the U.S. Department of Education.

Extension Analyses

Tables showing results from extension analyses are shown in Appendices A and B. Table A1 shows the association between preschool and PPVT outcomes, at the end of treatment and follow-up in kindergarten, first, and second grades. The end of treatment effect is 0.216 standard deviation units; this decreases in kindergarten to 0.153, 0.187 standard deviation units in first grade, and 0.251 in second grade. The results in Table A2, with follow-up waves instead of grade level, are nearly identical except for a smaller effect at the third follow-up point, only 0.199 standard deviation units. Tables A3 and A4 show results for WJ-III Letter-Word subscale outcomes. The end of treatment average effect is 0.248 standard deviation units. However, beyond that the average estimated effect size is not statistically significant, due to large standard errors and a lack of precision. Indeed, there are only three or four studies with available WJ-III Letter-Word subscale assessments at each measurement period. Finally, Tables A5 and A6 show results from the WJ-III Applied Problems subscale. The average end of treatment effect is 0.164

standard deviation units. Interestingly, Applied Problems assessments given beyond third grade have an average treatment effect of 0.222 standard deviation units as assessed across four studies (both Abecedarian cohorts, Arkansas, and CPC). This estimate is precisely estimated ($p<.001$).

Appendix B shows secondary analysis results related to sample selection. To probe potential influence of sample generation, I restrict the sample in several ways. First, I run estimates including studies that have consistent outcome measures for end of treatment, kindergarten, and first grade, with no new studies introduced if they do not have those first three measures. However, some of these studies lack measures in second grade and older; therefore I run estimates with only studies that include outcomes for each grade level. Shown in Table B2, these three studies include Perry Preschool, Howard preschool experiment, and Arkansas Better Chance program. In both cases when the sample is restricted, there is observable fadeout in effect sizes, with a smaller, but positive effect of preschool on overall academic outcomes at the final follow-up period. This difference from main analysis effects reflects influence from the change in study samples over follow-up periods in the main analysis. This is important to consider as we move forward with trying to aggregate our knowledge of preschool effects; the potential influence of study differences, such as follow-up timing, is not readily apparent in aggregate meta-analyses.

Tables B3 and B4 show the association between academic/cognitive outcomes and public preschool participation, conditional on whether the intervention occurred pre- or post-NCLB. In the post-NCLB period, results somewhat mirror main analysis results, with a moderate end of treatment effect (0.197 standard units), and declining to 0.161 standard deviation units beyond third grade. Estimates on the pre-NCLB sample are somewhat harder to interpret, with a non-significant estimate at the end of treatment and in kindergarten, followed by large treatment

effects in first grade (0.41), third grade (0.398), and older (0.20). This may be because the studies included in this subsample are larger, randomized cornerstone studies (Perry, Abecedarian, and Howard experimental preschool). Tables B5 and B6 show the association between research design factors and effect sizes, based on samples pre- and post-NCLB. With the split sample, there is not enough power to detect an association.

Discussion

Summary of Evidence

This study finds significant positive effects at the end of preschool treatment, on average. The effect for literacy remains significant and positive over time, with fadeout in the early grades. Math and general cognitive effects are less consistent, but are positive and significant at end of treatment and in first grade (and in later grades for math). There is a pattern of persistence across outcomes, although some estimates lack precision. Point estimates beyond third grade are smaller than at end of treatment, but with overlapping confidence intervals. Therefore these meta-analyses may reflect persistence, albeit with some fadeout. That effects do not fade to zero, however, is a departure from current literature, and may be due to inclusion of newer studies, or that the current study isolates effects to achievement (and separates subject area as well) rather than a wider range of effect sizes. Overall this study adds to the current evidence base, and the analysis brings to light the need to follow-through with treatment participants beyond third grade. Given the growth of standardized, yearly assessments since the early 2000s, and significant state investments in expanding data systems, we can leverage longer-term data to address questions of persistence. Furthermore, the main analysis extensions show that there are differences between medium- and long-term studies that make it difficult to make accurate

comparisons, even in early elementary years. While current meta-analyses control for length of time since preschool (e.g., Li et al., 2016), there may be additional heterogeneity between studies that is not being picked up by this predictor (i.e., year of intervention, social and political contexts, use of large administrative data, etc.).

Findings were persistent after accounting for potential research design predictors. Studies of programs completed post-NCLB have lower effect sizes on average (-0.18 standard deviation units). This is driven by end of treatment effects, which are significantly lower for post-NCLB studies (-0.29, $p < .05$). If the theory about counterfactual comparison holds, it is likely that post-NCLB comparison groups are attending more center-based or academic care. With a lack of information about comparison group activity, it is difficult to make assumptions about overall long-term efficacy of programs.

Limitations

The purpose of this study is to assess the medium- and long-term persistence of public preschool academic outcomes. In order to be included in the meta-analysis, preschool studies needed multiple achievement assessment periods beyond treatment. This limited the number of included studies and reduces power and generalizability. Additionally, a common problem with meta-analysis is having to limit the number of studies due to incomplete reporting of sample sizes, treatment effects and variances (Borenstein et al., 2009). Having encountered this problem I had to remove two studies that would have otherwise qualified for study inclusion. However, the studies ultimately included represent several decades of nationwide public preschool research.

Conclusions and Policy Implications

Overall, the main analysis results reflect a decline of academic effects over time, however with positive, significant effects remaining in later years. This supports existing evidence that preschool has lifetime positive effects that often become apparent in high school or young adulthood (e.g., Chaudry et al., 2017; Phillips et al., 2017). There are several hypotheses that may explain the early decline seen in the main analysis. First, the extension analyses reflect discrepancies across sample restrictions that are a clear factor. Still, considering the fadeout effects we do see in studies limited to preschool through third grade, it may be that the transition shock from preschool to kindergarten to elementary school destabilizes achievement gains, which then resurface after students have adjusted to the demands of regular schooling. Additionally, less rigorous teaching in kindergarten, or curricular misalignment might result in lower achievement scores. Furthermore, students may be more attuned to test-taking and develop stronger test-taking skills beginning in third grade, given school responses to accountability policies, which then provides a boost that translates to preschool study effects.

Considering these results in conversation with prior meta-analyses, specifically the Li et al. timing study (2016), I find similar, moderately-sized effects at the end of treatment. In the restricted sample analyses, I also observe fadeout as reported in the study by Li and colleagues (2016). As shown in the sensitivity analyses of the current paper, the estimates are highly sensitive to the inclusion (or non-inclusion) of any one of the thirteen studies included overall. With a much larger sample of studies and effect sizes (67 studies and over 1,000 effect sizes), Li and colleagues likely benefit from greater consistency across estimates. Still, the current study is informative– this sensitivity, as well as information provided by the variance estimates across

models – suggests that context is highly influential and should be considered closely when aggregating preschool evaluations for research or policymaking contexts.

One proposed moderator in this study was a significant predictor of overall effects. Post-NCLB effects were significantly smaller, suggesting that there is more to understand about how context changes effect persistence; Perry and Abecedarian may no longer be appropriate studies for setting expectations about public preschool effects. While more current interventions have been unable to replicate the large effects from these early programs, the present study suggests that smaller effects are persistent and not negligible. Furthermore, variance estimates in the meta-analyses (specifically I^2 estimates) suggest that there are more observable moderators that should be explored further. Therefore, it is likely not a question of whether or not to expand public preschool, but a more nuanced consideration of program elements and transition experiences that should be considered by policymakers. Future research should continue to observe potential elementary school environmental factors that may explain the persistence or lack thereof of preschool effects on academic progress. Finally, given changes in growing preschool attendance and the movement of accountability pressures from elementary school down to kindergarten and even preschool, future preschool evaluations should consider how these components (comparison group activity, intervention goals and resources, transition activities) may influence end of treatment and longer-term effects.

References

(* indicates report included in meta-analysis)

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and Fadeout in the Impacts of Child and Adolescent Interventions. *Journal of Research on Educational Effectiveness*, *10*(1), 7–39. https://doi.org/10.1080/19345747.2016.1232459

Bailey, D. H., Nguyen, T., Jenkins, J. M., Domina, T., Clements, D. H., & Sarama, J. S. (2016). Fadeout in an early mathematics intervention: Constraining content or preexisting differences? *Developmental Psychology*, *52*(9), 1457–1469. https://doi.org/10.1037/dev0000188

Barnett, W. S. (1995). Long-Term Effects of Early Childhood Programs on Cognitive and School Outcomes. *The Future of Children*, *5*(3), 25–50.

*Barnett, W.S., Jung, K., Youn, M., & Frede, E.C. (2013). Abbott Preschool Program Longidutinal Effects Study: Fifth Grade Follow-Up. New Brunswick, NJ: National Institute for Early Education Research.

Bassok, D., Finch, J. E., Lee, R., Reardon, S. F., & Waldfogel, J. (2016). Socioeconomic Gaps in Early Childhood Experiences: 1998 to 2010. *AERA Open*, *2*(3), 233285841665392. https://doi.org/10.1177/2332858416653924

Bassok, D., Gibbs, C.R., & Latham, S. (2018). Preschool participation and early schooling outcomes: Evidence from two cohorts of kindergarten entrants. *Child Development*, accepted publication.

Bifulco, R. (2012). Can nonexperimental estimates replicate estimates based on random assignment ine valuations of school choice? A within-study comparison. *Journal of Policy Analysis and Management, 31*(3), 729-751.

Bitler, M.P., Hoynes, H.W., & Domina, T. (2014). *Experimental evidence on distributional effects of Head Start* (No. w20434). National Bureau of Economic Research.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to Meta-Analysis*. West Sussex, United Kingdom: John Wiley & Sons, Ltd.

Burchinal, M., Magnuson, K., Powell, D., & Soliday Hong, S. (2015). Early child care and education. In R. Lerner (Ed.), Handbook of child psychology and developmental science (7th ed., Vol. 4, pp.1-45). Hoboken, NJ: Wiley.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Teachers College Record: Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development. Retrieved from http://www.tcrecord.org.proxy.library.vanderbilt.edu/library/content.asp?contentid=15440

Chaudry, A., Morrissey, T., Weiland, C., & Yoshikawa, H. (2017). *Cradle to Kindergarten: A New Plan to Combat Inequality*. New York, NY: Russell Sage Foundation.

Cheung, A.C.K., & Slaving, R.E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283-292.

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.

*Frede, E., Jung, K., Barnett, W.S., Lamy, C.E., & Figueras, A. (2007). The Abbott Preschool Program longitudinal effects study, interim report. New Brunswick, NJ: National Institute for Early Education Research.

*Frede, E., Jung, K., Barnett, W.S., & Figueras, A. (2009). The APPLES blossom: Abbott Preschool Program longitudinal effects study preliminary results through second grade. New Brunswick, NJ: National Institute for Early Education Research.

Friedman-Krauss, A. H., Connors, M. C., & Morris, P. A. (2017). Unpacking the Treatment Contrast in the Head Start Impact Study: To What Extent Does Assignment to Treatment Affect Quality of Care? *Journal of Research on Educational Effectiveness*, *10*(1), 68–95. https://doi.org/10.1080/19345747.2016.1147627

*Gray, S.W., & Klaus, R.A. (1970). The Early Training Project: A seventh-year report. *Child Development, 41*(4), 909-924.

*Gray, S.W., Ramsey, B.K., & Klaus, R.A. (1982). *From 3 to 20: The Early Training Project*. Baltimore, MD: University Park Press.

Grindal, T., Bowne, J. B., Yoshikawa, H., Schindler, H. S., Duncan, G. J., Magnuson, K., & Shonkoff, J. P. (2016). The added impact of parenting education in early childhood education programs: A meta-analysis. *Children and Youth Services Review*, *70*, 238–249. https://doi.org/10.1016/j.childyouth.2016.09.018

Hedberg, E.C. (2011). "ROBUMETA: Stata module to perform robust variance estimation in meta-regression with dependent effect size estimates," Statistical Software Components S457219, Boston College Department of Economics, revised 23 April 2014.

Hedges, L.V., Tipton, E., & Johnson, M.C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 2010*(1), 39-65.

*Herzog, E. (1972). Preschool and postscript: An evaluation of an inner-city program.

*Hustedt, J.T., Barnett, W.S., & Jung, K. (2008). Arkansas Better Chance program: Findings from kindergarten and first grade. New Brunswick, NJ: National Institute for Early Education Research.

Hustedt, J.T., & Barnett, W. S. (2011). Financing early childhood education programs: State, federal, and local issues. *Educational Policy, 25*(1), 167-192. DOI: 10.1177/0895904810386605

Innocenti, M. (2015). The facts: Behind Utah's social impact bond for early childhood education [Policy brief]. Retrieved from www.payforsuccess.org/sites/default/files/Utah%20Facts.pdf

Jenkins, J.M. (2014). Early childhood development as economic development: Considerations for state-level policy innovation and experimentation. *Economic Development Quarterly, 28*(2), 147-165.

*Jung, K., Barnett, S.W., Hustedt, J.T., & Francis, J. (2013). Longitudinal effects of the Arkansas Better Chance program: Findings from first grade through fourth grade. New Brunswick, NJ: National Institute for Early Education Research.

*Lamy, C., Barnett, W.S., & Jung, K. The effects of the Michigan School Readiness program on young children's abilities at kindergarten entry. New Brunswick, NJ: National Institute for Early Education Research.

Lemons, C.J., Fuchs, D., Gilbert, J.K., & Fuchs, L.S. (2014). Evidence-based practices in a changing world: Reconsidering the counterfactual in education research. *Educational Researcher, 43*(5), 242-252.

Li, W., Duncan, G.J., Magnuson, K., Schindler, H., Yoshikawa, H., Leak, J., & Shonkoff, J.P. (2016). Is timing everything? How early childhood education program cognitive and achievement impacts vary by starting age, program duration, and time since the end of the program. Manuscript under review.

Li-Grining, C. P., Votruba-Drzal, E., Maldonado-Carre?o, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology*, *46*(5), 1062–1077. https://doi.org/10.1037/a0020066

*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. Nashville, TN: Vanderbilt University, Peabody Research Institute.

*Lipsey, M.W, Farran, D.C, & Hofer, K.G. (2015). A randomized control trial of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade. Nashville, TN: Vanderbilt University, Peabody Research Institute.

*Lipsey, M.W., Farran, D.C., & Hofer, K.G. (2016). Effects of a state prekindergarten program on children's achievement and behavior through third grade. *Working Paper*. Nashville, TN: Peabody Research Institute.

Magnuson, K. A., Kelchen, R., Duncan, G. J., Schindler, H. S., Shager, H., & Yoshikawa, H. (2016). Do the effects of early childhood education programs differ by gender? A meta-analysis. *Early Childhood Research Quarterly*, *36*, 521–536. https://doi.org/10.1016/j.ecresq.2015.12.021

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. https://doi.org/10.1016/j.ecresq.2006.10.002

Magnuson, K., & Waldfogel, J. (2016). Trends in Income-Related Gaps in Enrollment in Early Childhood Education. *AERA Open*, *2*(2), 2332858416648933.

*Malofeeva, E., Daniel-Echols, M., & Xiang, Z. (2007). Findings from the Michigan School Readiness program 6 to 8 follow up study. High/Scope Educational Research Foundation. Ypsilanti, MI.

McCoy, D.C., Yoshikawa, H., Ziol-Guest, K.M., Duncan, G.J., Schindler, H.S., Magnuson, K., Yang, R., Koepp, A., & Shonkoff, J.P. (2017). Impacts of early childhood education on medium- and long-term educational outcomes. *Educational Researcher, 46*(8), 474-487. DOI: 10.3102/0013189X17737739

Nelson, G., Westhues, A., & MacLeod, J. (2003). A Meta-Analysis of Longitudinal Research on Preschool Prevention Programs for Children. *Prevention & Treatment*, *6*(1). https://doi.org/http://dx.doi.org.proxy.library.vanderbilt.edu/10.1037/1522-3736.6.1.631a

No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 *et seq*. (West 2002)

Phillips, D., A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., … Weiland, C. (2017). *The Current State of Scientific Knowledge on Preschool Effects*.

*Puma, M., Bell, S. Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A. & Downer, J. (2012). *Third grade follow-up to the Head Start Impact Study final report*. OPRE Report #2012-45, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

*Ramy, C.T., Campbell, F.A., Burchinal, M., Skinner, M.L., Gardner, D.M., & Ramey, S.L. (2000). Persistent effects of early childhood education on high-risk children and their mothers. *Applied Developmental Science, 4*(1), 2-14.

*Reynolds, A.J. (1995). One year of preschool intervention or two: Does it matter? *Early Childhood Research Quarterly, 10*(1), 1-31.

*Reynolds, A.J., & Ou, S.R. (2011). Paths of effects from preschool to adult well-being: A confirmatory analysis of the Child-Parent Center Program. *Child Development, 82*(2), 555-582.

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

*Ricciuti, A.E., St. Pierre, R.G., Lee, W., Parsad, A., & Rimdzius, T. (2004). *Third National Even Start Evaluation: Follow-up findings from the experimental design study*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Washington, DC.

Samuels, C.A. (2018, July 18). Are the Effects of State Pre-K Overrated? *Education Week.* Retrieved from https://www.edweek.org/ew/articles/2018/07/18/are-the-effects-of-state-pre-k-overrated.html

Schindler, H. S., Kholoptseva, J., Oh, S. S., Yoshikawa, H., Duncan, G. J., Magnuson, K. A., & Shonkoff, J. P. (2015). Maximizing the potential of early childhood education to prevent externalizing behavior problems: A meta-analysis. *Journal of School Psychology*, *53*(3), 243–263. https://doi.org/10.1016/j.jsp.2015.04.001

*Schweinhart, L.J., & Weikart, D.P. (1980). Effects of Perry Preschool program on youths through age 15. *Journal of the Division for Early Childhood, 4*(1), 29-39.

*Schweinhart, L.J., Barnes, H.V., & Weikart, D.P. (1993). *Significant benefits of the High-Schope Perry preschool study through age 27*. Ypsilanti, MI: High/Scope Press.

*Schweinhart, L.J. (2013). Long-term follow-up of a preschool experiment. *Journal of Experimental Criminology, 9*(4), 389-409.

Shadish, W.R., Clark, M.H., & Steiner, P.M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association 103*(484), 1334-1344.

Shager, H. M., Schindler, H. S., Magnuson, K. A., Duncan, G. J., Yoshikawa, H., & Hart, C. M. D. (2013). Can Research Design Explain Variation in Head Start Research Results? A Meta-Analysis of Cognitive and Achievement Outcomes. *Educational Evaluation and Policy Analysis*, *35*(1), 76–95. https://doi.org/10.3102/0162373712462453

Stipek, D. (2006). No Child Left Behind comes to preschool. *The Elementary School Journal, 106*(5), 455-465.

Stipek, D., Clements, D., Coburn, C., Franke, M., & Farran, D. (2017). PK-3: What does it mean for instruction? *Social Policy Report, 30*(2).

Stipek, D. (2017, March 17). The Preschool Fade-Out Effect is Not Inevitable. *Education Week Commentary.* Retrieved from https://www.edweek.org/ew/articles/2017/03/17/the-preschool-fade-out-effect-is-not-inevitable.html

Tanner-Smith, E.E., & Tipton, E. (2013). Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods, 2014*(5), 13-30.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20*(3), 375-393.

*U.S. Department of Health and Human Services, Administration for Children and Families (May 2005). Head Start Impact Study: First Year Findings. Washington, DC.

*U.S. Department of Health and Human Services, Administration for Children and Families. (2010). Head Start Impact Study. Final Report. Washington, DC.

*Weikart, D.P., Bond, J.T., & McNeil, J.T. (1978). *The Ypsilanti Perry Preschool Project: Preschool years and longitudinal results through fourth grade* (No. 3). High/Scope Foundation. Ypsilanti, MI.

*Xiang, Z., & Schweinhart, L.J. (2002). Effects five years later: The Michigan School Readiness program evaluation through age 10. High/Scope Educational Research Foundation. Ypsilanti, MI.

TABLES AND FIGURES

Table 1: Inclusion and Exclusion Criteria

Inclusion	**Exclusion**
A. Experimental study (RCT) B. Quasi-experimental using one of the following designs with existence of some type of comparison group: regression discontinuity, fixed effects, difference in difference, instrumental variables, propensity score analysis, interrupted time series C. Eligibility is based on one of the following: universal, selected neighborhood, family, or child criteria D. Includes child-level outcomes measured at least once beyond kindergarten entry	A. The intervention did not provide services for a child B. There is not a comparison group C. Children's ages during the intervention are NOT prenatal to 5 years old D. The main purpose of the intervention is to determine the efficacy or effectiveness of pharmacological agents, medical procedures, or health-related products E. The intervention is for children with diagnosed behavioral, emotional, or medical disorders or learning disabilities F. The study is conducted outside of the U.S.

Table 2. Search Strategies

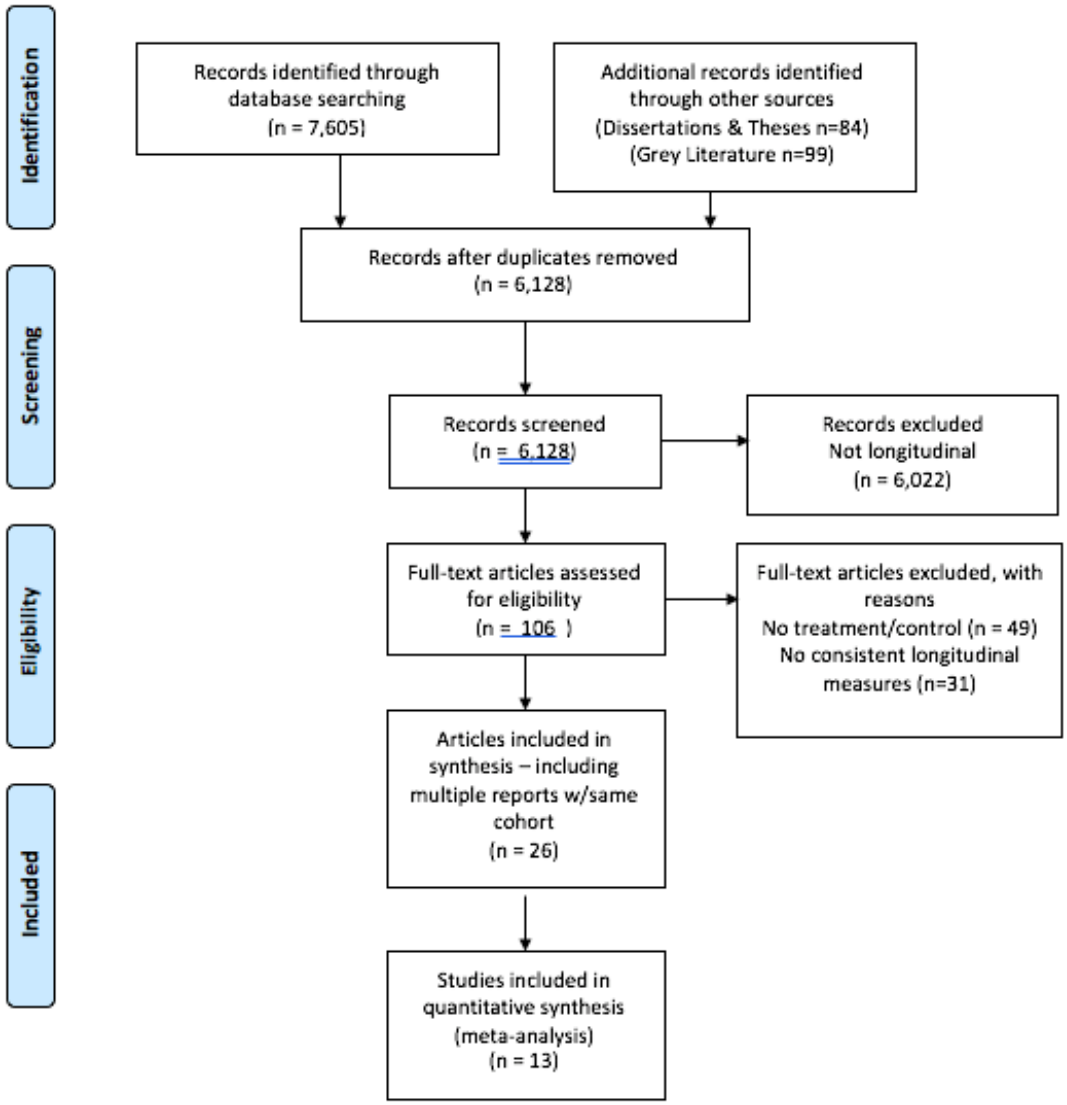
Computer and/or Search of Electronic Databases
Eric (Educational Resources Information Center database)
ProQuest
PsycINFO Psychological Abstracts
Manual search of proceedings from relevant research conferences (e.g., AERA, SCRD, Head Start, NAEYC)
Footnote Chasing
References in journals from nonreview articles
References from nonreview articles not published in journals
References in review articles
References in books/book chapters
References listed on program/program model Web sites
Topical bibliographies compiled by others
Consultation
Communications with colleagues
Attending meetings and conferences
Formal requests of scholars who are active in the field
Formal requests of foundations that fund research in the field
General requests to government agencies
Reviewing electronic networks

Table adapted from Camilli et al., Table 1: p. 584

Table 3: Codebook

Variable	**Description**
Type of Measure	1 = Self-report 2 = Rating by someone else 3 = Performance test (including direct assessment and standardized achievement) 4 = Physiological measure 5 = Observational rating 6 = Other/mixed
Name of Measure	
Form of Assignment of 2 Groups to Conditions	1 = Random assignment 2 = Quasi-experimental 3 = Groups were initially randomized, but changed during the study or post-hoc
Type of QED	1 = Family fixed effects 2 = Individual fixed effects 3 = Other longitudinal change 4 = Residualized and other kinds of change 5 = Instrumental variables 6 = Difference in difference 7 = Regression discontinuity 8 = Propensity Score Matching on demographics 9 = Propensity Score Matching on baseline measure of outcome 10 = Interrupted time series 11 = Above designs don't apply, but groups are baseline equivalent
Baseline Equivalence of Groups Tested	1 = Yes 2 = No
If "Yes" on Baseline, Any Significant differences in Groups?	Y/N, and describe
Take Up of Services (attendance by those assigned)	1 = Low 2 = Medium 3 = High 4 = Cannot be determined
Type of Comparison Group	1 = "Treatment" comparison group (alternative treatment) 2 = "Control" comparison
Activity Level of Comparison Group	1 = Passive (no treatment/wait list) 2 = Active (placebo, business as usual exposure)
Average Age at Posttest in Months	Number of months
Number of Months Elapsed Since Pretest/Initiation of Treatment	Number of months
Attrition Bias Tested?	Y/N, and what amount
Sample Size	Number of "treatment" and "control" students

Figure 1: PRISMA Flowchart



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *Preferred Reporting Items for Systematic Reviews and Meta-Analyses*: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

Table 4: Descriptive Characteristics of Studies Included in Meta-Analysis

Sample	Average	Range
Studies (total)	13	
Sample Size	956	43 – 6777
Total number of treatment effects	250	4 – 46
Effect size	.089	-.51 – 1.13
Assignment	%	
RCT	46.4	0/1
QED	53.6	0/1
PSM	22.8	0/1
Other QED	30.8	0/1
Design		
Number of follow-up assessments (beyond end-of-treatment)	3.7	1-5
Baseline equivalence tested (and adjusted) (%)	68.4	0/1
Business as usual comparison group (%)	76.4	0/1
Attrition tested (and adjusted) (%)	58.4	0/1
Average age at posttest in months	127.73	4.9 – 192
Average number of months elapsed since treatment	77.6	1 – 132

Note: Descriptive statistics include average across studies and overall range. First row, number of studies, indicates total number of studies included in the meta-analysis.

Figure 2: Distribution of Study Effect Sizes Across Time

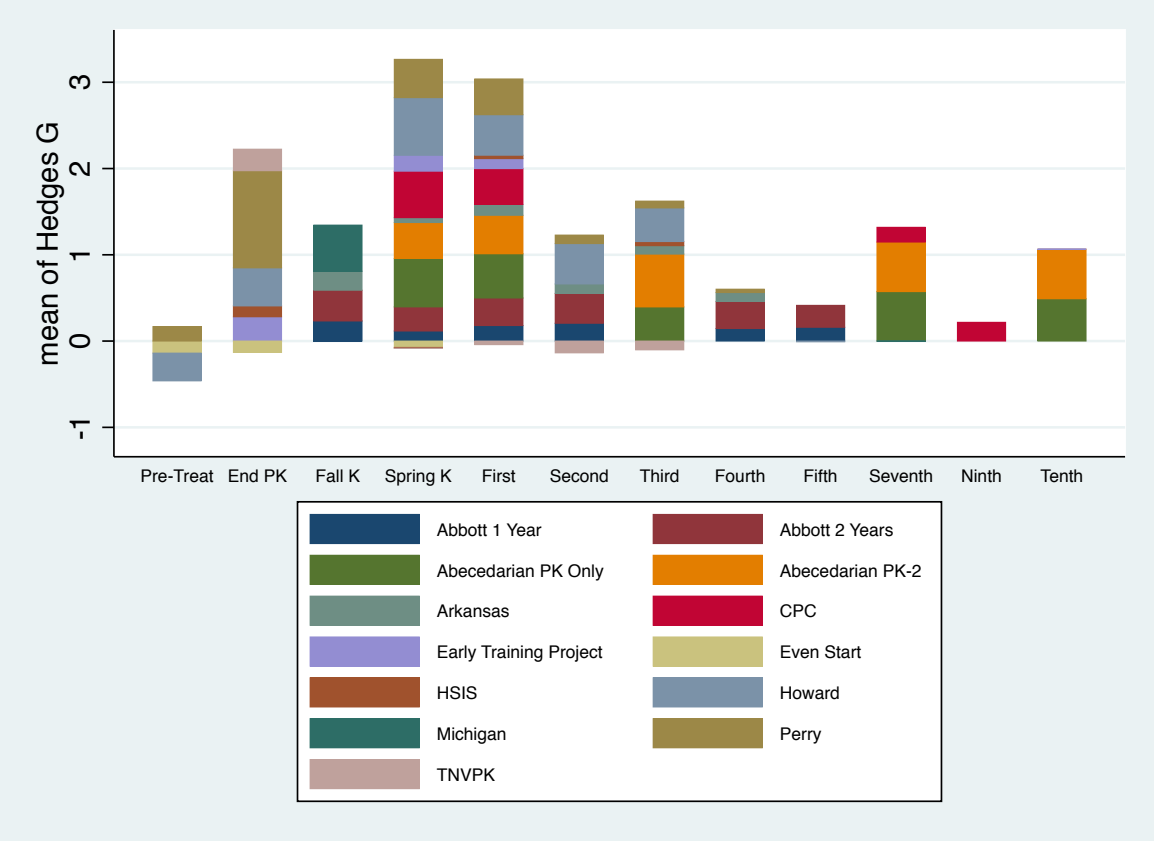


Figure 3: Percent Study Characteristics, Pre- and Post-NCLB

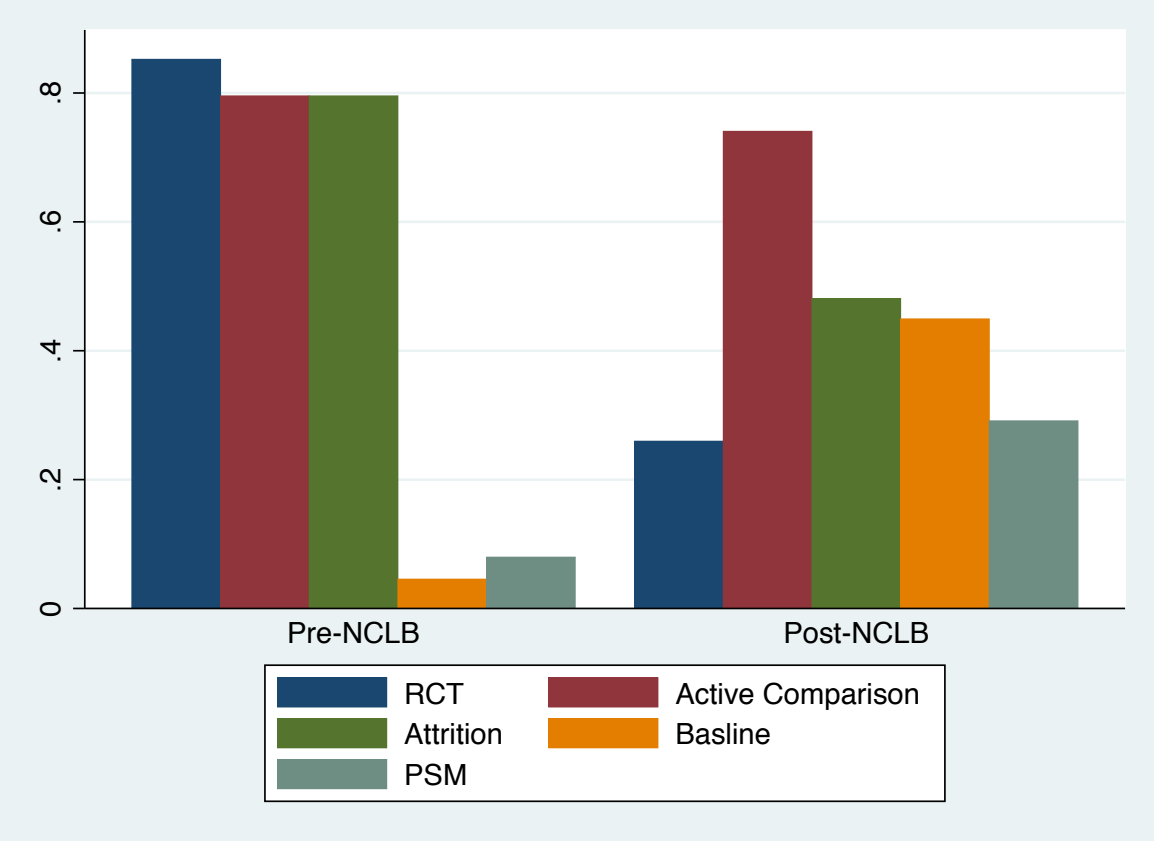


Table 5: Programs Included and Availability of Treatment Effect Assessments

		At End of Treatment					At Follow Up				
Program	**Starting Year**	**General Cognition**	**PPVT**	**WJ-III Math Subscales**	**WJ-III Reading Subscales**	**Other**	**General Cognition**	**PPVT**	**WJ-III Math Subscales**	**WJ-III Reading Subscales**	**Other**
Early Training Project	1962	x					x				
Perry Preschool	1962	x					x				
Howard	1964	x					x				
Abecedarian (PK Only)	1972	x					x		x	x	
Abecedarian (PK-2)	1972	x					x		x	x	
Child Parent Center	1986	x				x					x
Michigan	1995		x			x					x
Even Start	1999	x	x	x	x	x	x	x	x	x	x
HSIS	2002	x	x	x	x	x	x	x	x	x	x
Abbott (1 Year)	2004		x	x		x		x	x		x
Abbott (2 Years)	2004		x	x		x		x	x		x
Arkansas Better Chance	2005		x	x				x	x	x	
TNVPK	2009	x		x	x		x		x	x	
Total Studies (k)		9	6	6	3	6	8	5	8	6	6

Note: Each study listed with the starting year of program. *x* indicates assessment was given in that category.

Table 6: Association Between Public Preschool and Literacy Assessment at End of Treatment and Follow-up

		Follow-up (grade)				
	End of Treatment	Kindergarten	First	Second	Third	Older
Average Treatment Effect	0.181	0.101	0.109	0.128	0.135	0.174
Standard error						
CI lower	0.129	0.057	0.072	0.082	0.092	0.135
CI upper	0.234	0.145	0.147	0.175	0.178	0.212
k (study n)	8	9	7	7	9	10
z (test ES=0)	6.75	4.53	5.67	5.39	6.14	8.80
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Variance						
χ^2	239.00	197.08	119.31	121.47	118.51	90.58
χ^2 p-value	0.000	0.000	0.000	0.000	0.000	0.000
I^2 (%)	83.7	78.7	69.8	73.7	71.3	58.0
τ^2	0.0213	0.0147	0.0086	0.0127	0.0107	0.0076

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p < .05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table 7: Association Between Public Preschool and Literacy Assessment at End of Treatment and Follow-up

		Follow-up (wave)				
	End of Treatment	1	2	3	4	5
Average Treatment Effect	0.181	0.095	0.117	0.134	0.137	0.176
CI lower	0.129	0.053	.078	0.092	0.092	0.134
CI upper	0.234	0.137	0.155	0.177	0.183	0.218
k (study n)	8	8	10	10	9	6
z (test ES=0)	6.75	4.45	5.98	6.18	5.94	8.26
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Variance						
χ^2	239.00	199.63	134.02	136.03	112.81	62.29
χ^2 p-value	0.000	0.000	0.000	0.000	0.000	0.000
I^2 (%)	83.7	77.5	70.9	72.8	70.7	56.7
τ^2	0.0213	0.0142	0.0095	0.0116	0.0115	0.0064

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p < .05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table 8: Association Between Public Preschool and Math Assessment at End of Treatment and Follow-up

		Follow-up (grade)				
	End of Treatment	Kindergarten	First	Second	Third	Older
Average Treatment Effect	0.191	0.069	0.083	0.048	-0.008	0.186
CI lower	0.106	-0.021	0.033	-0.075	-0.075	0.117
CI upper	0.275	0.158	0.133	0.172	0.06	0.256
k (study n)	7	7	5	4	5	6
z (test ES=0)	4.43	1.50	3.28	0.77	0.22	5.26
p-value	0.000	0.133	0.001	0.440	0.823	0.000
Variance						
χ^2	29.44	60.52	29.66	58.36	28.45	24.48
χ^2 p-value	0.000	0.000	0.003	0.000	0.002	0.027
I^2 (%)	72.8	83.5	59.5	86.3	64.8	46.9
τ^2	0.0111	0.0179	0.0047	0.0304	0.0074	0.0071

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p < .05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table 9: Association Between Public Preschool and Math Assessment at End of Treatment and Follow-up

		Follow-up (wave)				
	End of Treatment	1	2	3	4	5
Average Treatment Effect	0.191	0.069	0.096	0.070	0.028	0.109
CI lower	0.106	-0.021	0.047	-0.019	-0.071	0.026
CI upper	0.275	0.158	0.146	0.159	0.126	0.192
k (study n)	7	7	8	8	5	3
z (test ES=0)	4.43	0.15	3.82	1.54	0.55	2.57
p-value	0.000	0.133	0.000	0.123	0.580	0.010
Variance						
χ^2	29.44	60.52	37.61	78.07	36.39	7.11
χ^2 p-value	0.000	0.000	0.001	0.000	0.000	0.212
I^2 (%)	72.8	83.5	60.1	83.3	75.3	29.7
τ^2	0.0111	0.0179	0.0055	0.0214	0.0167	0.0032

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p < .05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table 10: Association Between Public Preschool and General Cognition at End of Treatment and Follow-up

		Follow-up (grade)				
	End of Treatment	Kindergarten	First	Second	Third	Older
Average Treatment Effect	0.396	0.109	0.163	0.093	0.075	0.161
CI lower	0.204	-0.076	0.007	-0.26	-0.147	-0.028
CI upper	0.588	0.293	0.318	0.446	0.297	0.351
k (study n)	9	6	7	3	5	4
z (test ES=0)	4.04	1.15	2.05	0.51	0.66	1.67
p-value	0.000	0.248	0.041	0.606	0.507	0.095
Variance						
χ^2	55.65	18.21	13.67	8.11	7.17	3.23
χ^2 p-value	0.000	0.003	0.034	0.017	0.127	0.779
I^2 (%)	85.6	72.5	56.1	75.3	44.2	0.000
τ^2	0.0557	0.0297	0.0177	0.0719	0.0267	0.000

Note: Robust variance estimation to assess average treatment effect at each time point. Confidence intervals ($p < .05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table 11: Association Between Public Preschool and General Cognition Assessment at End of Treatment and Follow-up

		Follow up (wave)			
	End of Treatment	1	2	3	4
Average Treatment Effect	0.396	0.251	0.06	0.229	0.12
CI lower	0.204	0.026	-0.05	-0.075	-0.125
CI upper	0.588	0.476	0.17	0.533	0.364
k (study n)	9	8	6	6	5
z (test ES=0)	4.04	2.19	1.07	1.48	0.96
p-value	0.000	0.029	0.285	0.140	0.337
Variance					
χ^2	55.65	38.16	6.88	18.12	8.69
χ^2 p-value	0.000	0.000	0.229	0.003	0.069
I^2 (%)	85.6	81.7	27.4	72.4	54.0
τ^2	0.0557	0.0663	0.0047	0.0942	0.0387

Note: Robust variance estimation used with robumeta command in Stata to assess average treatment effect at each time point. Confidence intervals (p<.05), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table 12. Post-hoc F-test for Tables 6-11

Grade level follow-up		ET==K p-value	ET==First p-value	ET==Second p-value	ET==Third p-value	ET==older p-value
Table 6	Literacy	0.022	0.294	0.139	0.185	0.834
Table 8	Math	0.053	0.031	0.061	0.000	0.929
Table 10	General Cognition	0.348	0.065	0.140	0.032	0.809
Follow-up waves		ET==1 p-value	ET==2 p-value	ET==3 p-value	ET==4 p-value	ET==5 p-value
Table 7	Literacy	0.012	0.560	0.176	0.214	0.884
Table 9	Math	0.052	0.057	0.053	0.014	0.175
Table 11	General Cognition	0.337	0.003	0.363	0.082	0.000

Note: Post-hoc F-tests conducted for significant differences between estimates at end of treatment (ET) and follow-ups. P-value of F-tests shown above, at two-tailed $p < .05$ level.

Table 13: Association Between Study Design Features and Treatment Effects, at End of Treatment

End of Treatment Effects

		Coefficient	SE	p-value	N Level 1	k Level 2
RCT	(1)	-0.0542	0.1781	0.7711		
PSM	(2)	0.0953	0.1645	0.6281		
Attrition	(3)	0.0911	0.1452	0.5499	47	13
Baseline Equivalence	(4)	-0.1114	0.1122	0.3690		
Comparison Activity	(5)	-0.1685	0.1962	0.4406		

Note: Robust variance estimation used with robumeta command in Stata.

Table 14: Association Between Study Design Features and Treatment Effects, Over Time

Follow up Effects

		Coefficient	SE	p-value	N Level 1	k Level 2
Time elapsed (in months)	(1)	0.001	0.0008	0.2748		
RCT	(2)	-0.0186	0.1076	0.8691		
PSM	(3)	0.0293	0.0897	0.0917	203	13
Attrition	(4)	-0.0246	0.0917	0.7948		
Baseline Equivalence	(5)	0.0796	0.0721	0.3088		
Comparison Activity	(6)	0.0642	0.0753	0.4491		

Note: Robust variance estimation used with robumeta command in Stata.

Table 15: Association Between Study Design Features and Treatment Effects (Aggregated)

		Coefficient	SE	p-value	N Level 1	k Level 2
Time elapsed (in months)	(1)	0.0001	0.0009	0.9358		
RCT	(2)	-0.032	0.1078	0.7791		
PSM	(3)	0.0011	0.0845	0.9902	250	13
Attrition	(4)	0.027	0.0872	0.7639		
Baseline Equivalence	(5)	0.0228	0.0749	0.7704		
Comparison Activity	(6)	0.0303	0.0836	0.7379		

Note: Robust variance estimation used with robumeta command in Stata.

Table 16: Association Between Timing of Study and Treatment Effects (pre- and post-NCLB)

	Coefficient	SE	p-value	N Level 1	K Level 2
Post-NCLB (2002)	-.1799	.0696	.0328	246	13

Note: Robust variance estimation used with robumeta command in Stata.

Figure 4: Standard Funnel Plot

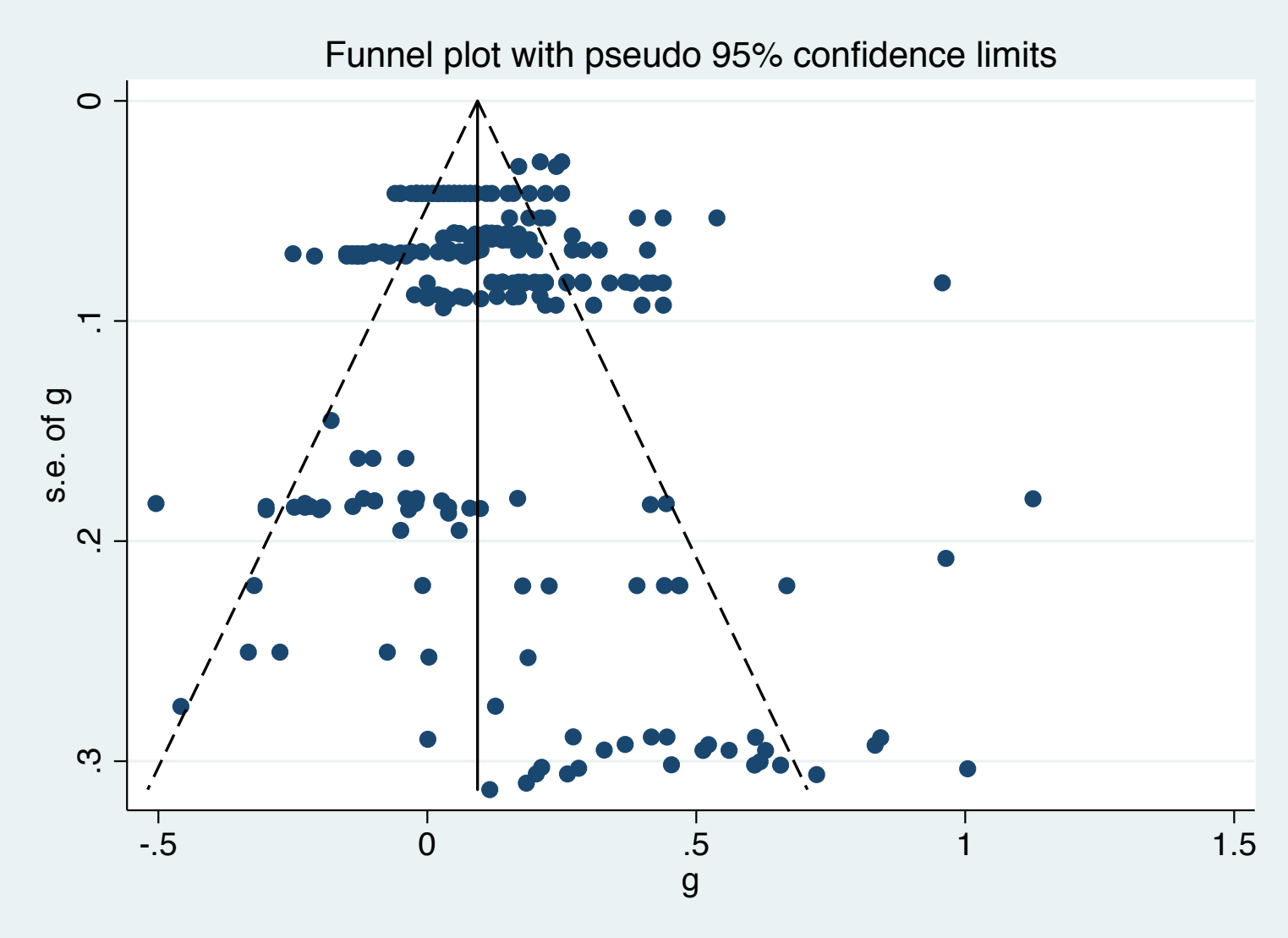


Figure 5: Funnel Plot, Effect Sizes and Standard Errors by Publication (Peer Reviewed or Not)

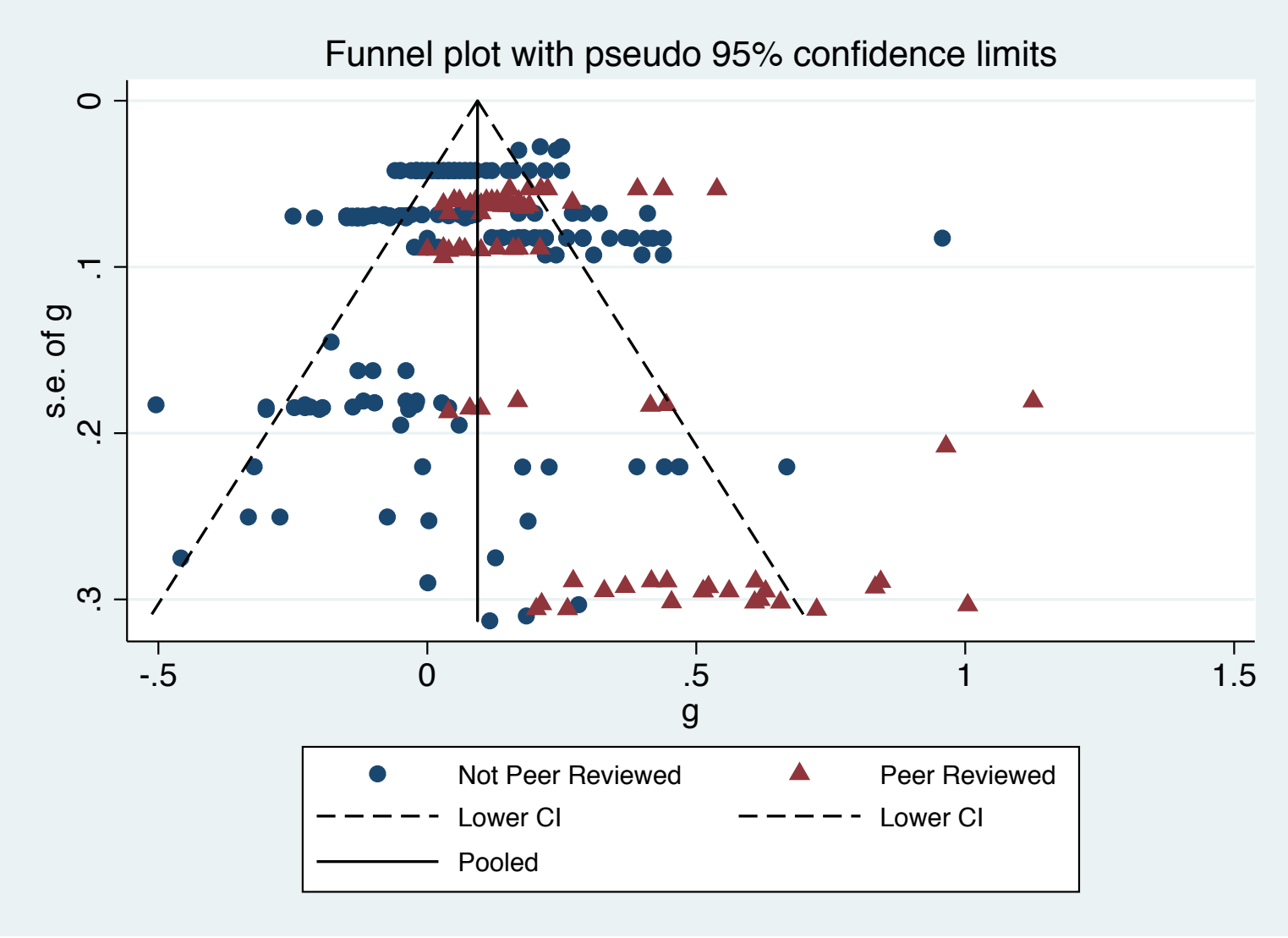
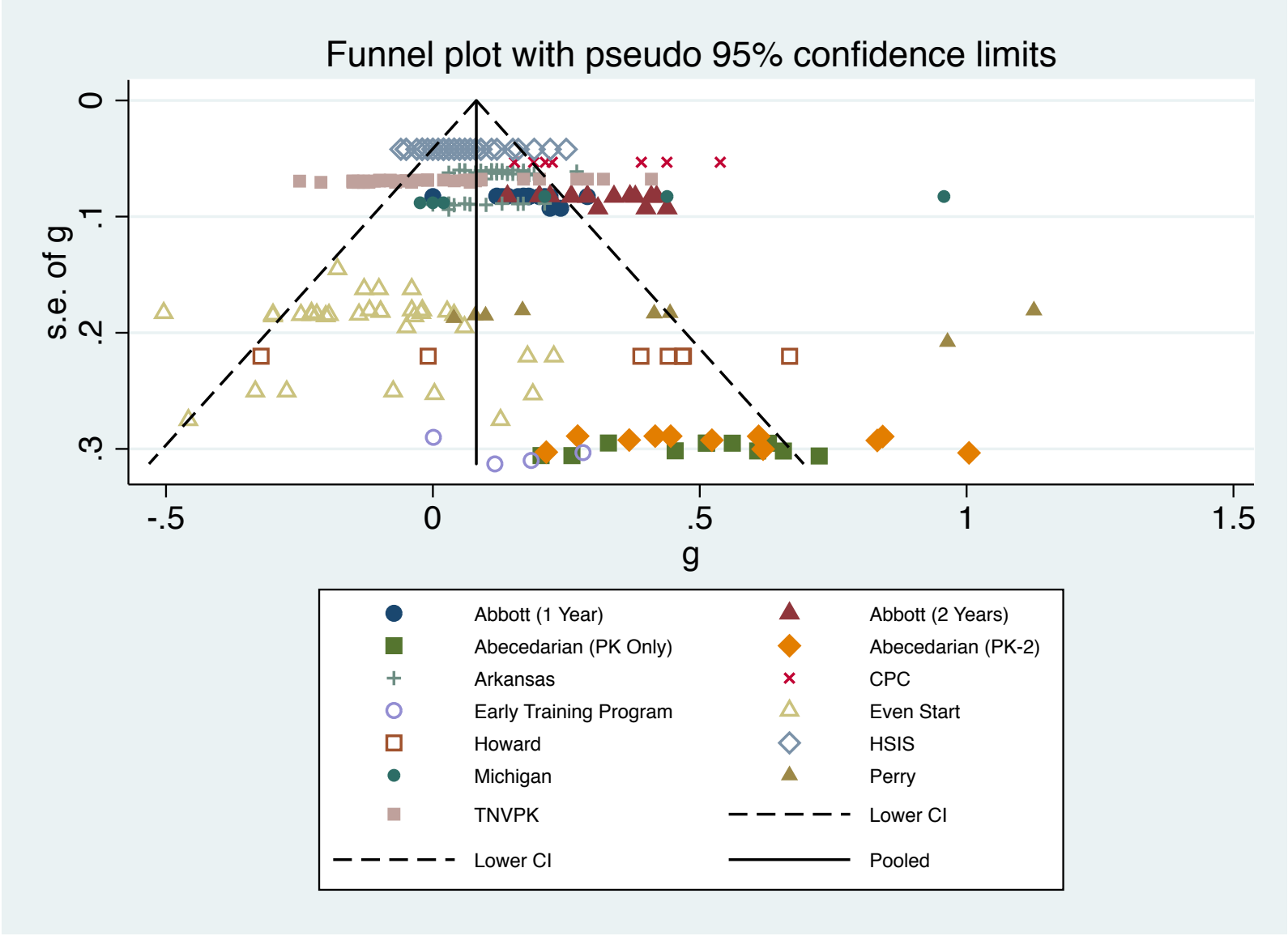


Figure 6: Funnel Plot, Effect Sizes and Standard Errors by Study



APPENDIX A

Table A1: Association Between Public Preschool and PPVT at End of Treatment and Follow-up

		Follow-up (grade)		
	End of Treatment	Kindergarten	First	Second
Average Treatment Effect	0.216	0.153	0.187	0.251
CI lower	0.108	0.006	0.071	0.117
CI upper	0.325	0.299	0.302	0.386
k (study n)	6	5	4	3
z (test ES=0)	3.91	2.04	3.16	3.66
p-value	0.000	0.041	0.002	0.000
Variance				
χ^2	15.92	18.8	9.82	4.25
χ^2 p-value	0.007	0.001	0.020	0.119
I^2 (%)	68.6	78.7	69.5	53.0
τ^2	0.0115	0.0201	0.0094	0.0075

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p < .05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table A2: Association Between Public Preschool and PPVT Assessment at End of Treatment and Follow-up

		Follow-up (wave)		
	End of Treatment	1	2	3
Average Treatment Effect	0.216	0.153	0.187	0.199
CI lower	0.108	0.006	0.071	0.073
CI upper	0.325	0.299	0.302	0.326
k (study n)	6	5	4	4
z (test ES=0)	3.91	2.04	3.16	3.09
p-value	0	0.041	0.002	0.002
Variance				
χ^2	15.92	18.8	9.82	10.63
χ^2 p-value	0.007	0.001	0.02	0.014
I^2 (%)	68.6	78.7	69.5	71.8
τ^2	0.0115	0.0201	0.0094	0.0115

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p < .05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table A3: Association Between Public Preschool and WJ-III Letter-Word Subscale at End of Treatment and Follow-up

		Follow-up (grade)		
	End of Treatment	Kindergarten	First	Third
Average Treatment Effect	0.248	-0.009	0.043	0.015
CI lower	0.065	-0.073	-0.065	-0.111
CI upper	0.432	0.056	0.151	0.141
k (study n)	3	4	3	3
z (test ES=0)	2.66	0.26	0.79	0.23
p-value	0.008	0.792	0.431	0.82
Variance				
χ^2	8.52	0.54	5.5	7.66
χ^2 p-value	0.014	0.91	0.064	0.022
I^2 (%)	76.5	0.000	63.6	73.9
τ^2	0.018	0.000	0.0058	0.0091

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p<.05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table A4: Association Between Public Preschool and WJ-III Letter-Word Subscale at End of Treatment and Follow-up

		Follow-up (wave)		
	End of Treatment	1	2	3
Average Treatment Effect	0.248	-0.009	0.043	0.029
CI lower	0.065	-0.073	-0.065	-0.105
CI upper	0.432	0.056	0.151	0.163
k (study n)	3	4	3	3
z (test ES=0)	2.66	0.26	0.79	0.42
p-value	0.008	0.792	0.431	0.673
Variance				
χ^2	8.52	0.54	5.5	8.54
χ^2 p-value	0.014	0.91	0.064	0.014
I^2 (%)	76.5	0.000	63.6	76.6
τ^2	0.018	0.000	0.0058	0.0107

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p<.05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table A5: Association Between Public Preschool and WJ-III Applied Problems Subscale at End of Treatment and Follow-up

		Follow-up (grade)				
	End of Treatment	Kindergarten	First	Second	Third	Older
Average Treatment Effect	0.164	0.126	0.116	0.164	0.000	0.222
CI lower	0.074	-0.01	0.036	-0.067	-0.148	0.105
CI upper	0.255	0.262	0.197	0.394	0.149	0.339
k (study n)	6	7	5	4	5	4
z (test ES=0)	3.56	1.81	2.84	1.39	0.01	3.71
p-value	0.000	0.070	0.004	0.164	0.995	0.000
Variance						
χ^2	12.00	41.55	7.82	27.55	14.33	10.90
χ^2 p-value	0.035	0.000	0.098	0.000	0.006	0.092
I^2 (%)	58.3	85.6	48.8	89.1	72.1	44.9
τ^2	0.0068	0.0269	0.004	0.049	0.0158	0.0085

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p<.05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table A6: Association Between Public Preschool and WJ-III Applied Problems Subscale at End of Treatment and Follow-up

		Follow-up (wave)			
	End of Treatment	1	2	3	4
Average Treatment Effect	0.164	0.126	0.14	0.175	0.137
CI lower	0.074	-0.01	0.067	0.032	-0.153
CI upper	0.255	0.262	0.214	0.317	0.427
k (study n)	6	7	8	8	4
z (test ES=0)	3.56	1.81	3.76	2.41	0.93
p-value	0.000	0.070	0.000	0.016	0.354
Variance					
χ^2	12.0	41.55	13.05	45.32	18.58
χ^2 p-value	0.035	0.000	0.071	0.000	0.000
I^2 (%)	58.3	85.6	46.4	84.6	83.9
τ^2	0.0068	0.0269	0.0045	0.0299	0.0584

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p<.05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

APPENDIX B

Table B1. Association between public preschool and academic/cognitive outcomes at end of treatment and follow-up: Restricted sample #1

		Follow-up (grade)				
	End of Treatment	Kindergarten	First	Second	Third	Older
Average Treatment Effect	0.196	0.002	0.044	-0.007	0.019	0.100
Standard Error	0.2744	0.0118	0.1235	0.0307	0.0205	0.0291
CI lower	0.142	-0.021	0.019	-0.068	-0.021	0.043
CI upper	0.249	0.025	0.068	0.053	0.060	0.157
k (study n)	5	5	5	4	5	3
z (test ES=0)	7.12	0.15	3.53	0.24	0.94	3.43
p-value	0.000	0.884	0.000	0.811	0.345	0.001
Variance						
χ^2	106.37	32.80	56.68	79.53	66.55	4.93
χ^2 p-value	0.000	0.168	0.009	0.000	0.000	0.896
I^2 (%)	81.20	20.70	40.00	74.90	63.90	0.00
τ^2	0.012	0.001	0.002	0.014	0.006	0.000

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals (p<.05), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table B2. Association between public preschool and academic/cognitive outcomes at end of treatment and follow-up periods: Restricted sample #2

		Follow-up (grade)				
	End of Treatment	Kindergarten	First	Second	Third	Older
Average Treatment Effect	0.45	0.16	0.15	0.11	0.10	0.10
Standard Error	0.1455	0.0790	0.0229	0.0206	0.0199	0.0291
CI lower	0.169	0.007	0.105	0.072	0.061	0.043
CI upper	0.739	0.317	0.195	0.153	0.139	0.157
k (study n)	3	3	3	3	3	3
z (test ES=0)	3.12	2.05	6.55	5.48	5.02	3.43
p-value	0.002	0.040	0.000	0.000	0.000	0.001
Variance						
χ^2	25.76	11.54	9.47	5.86	4.45	4.93
χ^2 p-value	0.000	0.021	0.488	0.827	0.925	0.896
I^2 (%)	88.40	65.40	0.00	0.00	0.00	0.00
τ^2	0.067	0.018	0.000	0.000	0.000	0.000

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p<.05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown. The three studies included in this analysis are: Arkansas Better Chance, Howard experimental preschool, and Perry Preschool Program.

Table B3. Association between academic/cognitive outcomes and public preschool; Pre-NCLB sample restriction

		Follow-up (grade)			
	End of Treatment	Kindergarten	First	Third	Older
Average Treatment Effect	0.16	0.073	0.41	0.398	0.20
Standard Error	0.0936	0.0737	0.1091	0.1116	0.0400
CI lower	-0.023	-0.071	0.196	0.179	0.122
CI upper	0.344	0.218	0.624	0.617	0.277
k (study n)	8	5	5	4	7
z (test ES=0)	1.71	0.99	3.76	3.57	5.02
p-value	0.086	0.320	0.000	0.000	0.000
Variance					
χ^2	177.87	75.67	1.09	9.57	39.28
χ^2 p-value	0.000	0.000	0.896	0.214	0.009
I^2 (%)	88.20	77.50	0.00	26.00	46.50
τ^2	0.152	0.064	0.000	0.026	0.011

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p<.05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown. *Second grade follow-up not shown because there were only two relevant studies; not enough to run a meta-analysis for this follow-up period.

Table B4. Association between academic/cognitive outcomes and public preschool; Post-NCLB sample restriction

	Follow-up (grade)					
	End of Treatment	Kindergarten	First	Second	Third	Older
Average Treatment Effect	0.197	0.026	0.055	0.046	0.016	0.161
Standard Error	0.0223	0.0151	0.0136	0.0330	0.0206	0.0230
CI lower	0.153	-0.004	0.029	-0.019	-0.025	0.116
CI upper	0.241	0.055	0.082	0.110	0.056	0.206
k (study n)	5	5	5	4	3	3
z (test ES=0)	8.83	1.71	4.07	1.39	0.77	7.02
p-value	0.000	0.087	0.000	0.165	0.442	0.000
Variance						
χ^2	95.40	64.68	76.83	131.64	63.77	19.48
χ^2 p-value	0.000	0.000	0.000	0.000	0.000	0.245
I^2 (%)	74.80	53.60	53.10	81.80	65.50	17.90
τ^2	0.009	0.004	0.003	0.022	0.006	0.002

Note: Robust variance estimation used to assess average treatment effect at each time point. Confidence intervals ($p<.05$), study k, and z test (Ho: ES=0) estimates shown. Variance estimates chi-squared test of heterogeneity, I-squared measure of heterogeneity, and tau-squared estimate of between-study variance shown.

Table B5. Pre-NCLB (predictors)

		Coefficient	SE	p-value	N Level 1	k Level 2
Time elapsed	(1)	-0.001	0.002	0.488	88	8
RCT	(2)	-0.008	0.111	0.948		
Attrition	(3)	0.004	0.098	0.968		
Comparison activity	(4)	0.001	0.088	0.988		

Note: Robust variance estimation used with robumeta command in Stata.

Table B6. Post-NCLB (predictors)

		Coefficient	SE	p-value	N Level 1	k Level 2
Time Elapsed	(1)	0.000	0.001	0.781	158	5
PSM	(2)	-0.159	0.059	0.084		
Attrition	(3)	0.073	0.107	0.547		
Baseline	(4)	-0.173	0.069	0.097		

Note: Robust variance estimation used with robumeta command in Stata.

Table B7. Post-hoc F-test for Tables B1-B4

Grade level follow-up		ET==K p-value	ET==First p-value	ET==Second p-value	ET==Third p-value	ET==Older p-value
Table B1	Restricted Sample 1	0.480	0.613	0.462	0.520	0.728
Table B2	Restricted Sample 2	0.078	0.039	0.020	0.016	0.017
Table B3	Pre-NCLB	0.465	0.082	0.087	0.102	0.694
Table B4	Post-NCLB	0.000	0.000	0.000	0.000	0.261

Note: Post-hoc F-tests conducted for significant differences between estimates at end of treatment (ET) and grade-level follow-ups. P-value of F-tests shown above, at two-tailed $p < .05$ level.

CHAPTER 2

HETEROGENEITY IN LEARNING TRAJECTORIES: THE ROLE OF PRESCHOOL ATTENDANCE AND SUBSEQUENT ENVIRONMENTS

Introduction

Research shows that access to high-quality early childhood education and care can reduce the achievement gap at kindergarten entry (Bailey et al., 2017; Phillips et al., 2017). Children attending a variety of preschool programs are readier for school at kindergarten entry than their non-attending peers, on short-term outcome measures including literacy, numeracy, and self-regulatory behaviors (Phillips et al., 2017). However, the evidence on long-term outcomes from early childhood interventions is mixed, with many longer-term evaluations reflecting smaller, null, or even negative effects of preschool when using a longer time horizon. Preschool "fadeout" refers to the decline of observed academic gains among students who attend public preschool, compared with their peers who did not attend. This pattern is puzzling for researchers, experts, and policymakers who view pre-kindergarten as an effective intervention for at-risk children. Recent studies find the academic skill advantages attributed to preschool begin to fade as early as first grade, and almost entirely disappear by third grade (Deming, 2009; Li et al., 2016; Phillips et al., 2017).

The current study uses growth modeling to address new research questions regarding the mechanisms of observed fadeout effects. The study relies on a nationally representative dataset to estimate the fadeout of preschool effects. While I do not observe test scores of students before preschool, I use measures of children's developmental status taken in fall of their kindergarten year as a proxy for post-preschool differences in achievement in various early childhood learning environments. These proxy measures have often been used in prior studies to address questions about early learning using the Early Childhood Longitudinal Study-Kindergarten data (ECLS-K)

(e.g., Bassok, Gibbs, & Latham, 2018; Claessens, Engel, & Curran, 2014; Loeb et al., 2007; Magnuson, Ruhm, & Waldfogel, 2007). In the current study, it is used as the first post-preschool measure of children's academic skills and the initial data point for their post-preschool academic skill trajectories. This paper adds to the current evidence base on the fadeout effects of preschool interventions by using growth modeling to test variation within and between children who attend Head Start, State preschool, Public/Private Center-based care, and no preschool, and the association between growth and subsequent environment characteristics. I aim to investigate the role of early elementary school environments through teacher instructional differentiation, as well as student social-emotional skills in influencing children's education trajectories through third grade. Overall, I address the following questions:

1. To what extent are differences in student growth in reading and math achievement over time associated with type of preschool experience?
2. To what extent is student achievement growth associated with subsequent classroom exposure to teacher instructional differentiation?
3. To what extent are growth trajectories mediated by social-emotional skills measured by teacher ratings of student approaches to learning?

Investigating Post-Preschool Learning Trajectories

Based on measures taken at the end of the program, a recent meta-analysis finds a weighted average effect size of 0.23 on measures of cognitive and achievement scores for one-year, two-year, and summer preschool programs (Li et al., 2016). After the end of treatment, however, the study finds an average fadeout effect on cognitive and achievement scores of 0.022 standard deviations per year (Li et al., 2016). This reduces the effect by nearly 10% each year after pre-kindergarten exposure. This finding is emphasized in a recent report on the scientific evidence of pre-kindergarten effects (Phillips et al., 2017). These authors find that, while evidence supports

the impact of pre-kindergarten on school readiness measures, evidence on long-term effects of public pre-kindergarten programs that relies on a wide variety of empirical methods is heavily mixed in terms of direction and significance (2017). Similarly, the randomized trial of Head Start finds a small boost for children's academic skills at the end of the program, but these gains rapidly faded when students entered formal schooling (Puma et al., 2012; Deming, 2009; Bitler, Hoynes, & Domina, 2014).

The term fadeout potentially oversimplifies what is likely substantial heterogeneity in the lasting effects of preschool. For example, some large-scale studies find positive, persistent effects only for specific subgroups, such as English Language Learners (ELLs) in the Tennessee-VPK and Head Start evaluations (Lipsey, Farran, & Hofer, 2016; Bitler et al., 2014). The current study contributes to the existing literature by exploring heterogeneity in post-preschool learning growth. The study uses multilevel growth modeling to identify associations and variation between and within preschool student subgroups. While growth modeling has been used with ECLS-K data to study early learning trajectories (e.g., Li-Grining et al., 2010; Roberts, Mohammed, & Vaughn, 2010), this will be the first study to use this method to explore trajectories based on preschool enrollment.

The remainder of this essay proceeds as follows. I first review the literature on the persistence of preschool intervention effects and provide a conceptual framework for the current study. I then describe the ECLS-K:11 dataset, followed by the study methods, including the operationalization of measures and specific growth models. Finally, I discuss the empirical findings of the study and limitations, as well as discuss potential implications for the results.

Literature Review

Pre-kindergarten is growing in both enrollment and expenditures (Barnett, 2017). The number of three- and four-year old children in preschool, defined as center-based educational care, grew by over 50% between 1989 and 2014, from 2.88 million students to 4.69 million (Chaudry & Datta, 2017). During this period, publicly funded pre-kindergarten enrollment tripled, from approximately 800,000 students to 2.7 million (Chaudry & Datta, 2017). Yet, the evidence on public preschool as an intervention is mixed. In this section, I review evidence from existing preschool evaluation studies, and develop a conceptual framework considering the mechanisms for lasting preschool impacts.

Long- and Short-Term Outcomes

According to the *skills beget skills* hypothesis, early intervention boosts capacity for subsequent interventions, supporting continual growth (Cunha & Heckman, 2007). However, evidence has shown that many preschool interventions do not follow the accelerated trajectory anticipated by this hypothesis; instead, there is an initial boost in cognitive outcomes that fades within a few years, and then evidence of some longer-term benefits to participants (Bailey et al., 2017). In this section, I review experimental and quasi-experimental evaluations of public preschool programs. In reviewing each program, I highlight the short- and long-term outcomes, and explain whether the pattern is more consistent with the skills begat skills or fade-out hypotheses.

Experimental evidence from twentieth century preschool evaluations, the Perry Preschool Project and Abecedarian, show a mix of short- to medium- term efficacy, but undeniably positive

long-term benefits both to participants and society. The Perry Preschool Project began in the 1960s and followed participants to age 40. A cost benefit analysis completed using data from participants at age 40 shows a nearly $13 gain for every $1 invested in the program (Belfield et al., 2006). A later replication adjusted for potentially compromised randomization, multiple-hypothesis testing and small sample sizes, and still found a 5.8% annual rate of return to investment from the Perry Program long-term outcomes (Heckman et al., 2010). These benefits accrue from long-term effects, with more years of education (for female participants), higher levels of employment and median income, less reported drug use, and less criminal activity at age 40 (Schweinhart, 2005). While a long-term impact in this study is universally accepted, cognitive effects of Perry as measured by child IQ almost completely faded away less than three years after end-of-treatment (Bailey et al., 2017). The Abecedarian Project, implemented a decade later, also shows significant long-term effects, with treatment group participants having more years of education, higher full-time employment, higher self-ratings of health, and a decreased likelihood of using public assistance at age 30 (Campbell et al., 2012). Participant IQ impacts persisted beyond age eight, although the effect on reading achievement dropped by more than half over the course of primary and secondary school, from an effect size of 0.28 to 0.11 (Campbell et al., 2002). Some researchers speculate that the continued IQ impacts of Abecedarian, as opposed to Perry, were due to the influence of superior (desegregated) subsequent school environments in the Chapel Hill, North Carolina area (Bailey et al., 2017).

Studies of the federal Head Start preschool program are generally positive (Currie & Thomas, 1993; Deming, 2009; Ludwig & Miller, 2007; Puma et al., 2012). A 2009 study uses differences in enrollment within families as an identification strategy for estimating the effects of Head Start (Deming, 2009). This study finds positive end-of-high school outcomes: participants

were more likely to graduate high school, not repeat grades, and less likely to have been diagnosed with a learning disorder. However, test score gains at kindergarten entry fade to less than half of their initial effect size by the end of elementary school (Deming, 2009). Here, continued, positive effects of social-emotional and learning skills may have supported academic persistence and motivation throughout early and secondary schooling, although achievement scores were subject to gradual fadeout. Experimental evidence from the National Head Start Impact Study (HSIS) reports small end-of-program impacts for three- and four-year-old participants on several cognitive construct measures (from 0.09 to 0.35 effect sizes in various language, literacy, and math skills), physical health (effect size 0.11), and parental use of educational activities at home (effect size 0.18) (Puma et al., 2006). However, all of these effects fade out by third grade (Puma et al., 2012). The HSIS has yet to observe end-of-high school or adult outcomes that may show participant benefit from the development of early learning skills.

The second-oldest (after Head Start) federally funded early childhood educational intervention is the Chicago Child-Parent Center (CPC) Program, established in 1967. The CPCs provide extensive educational and family support services, including half-day preschool, full-day kindergarten, access to health-screening resources, and a school-community representative that connects families to community resources, for children from ages three to nine (Reynolds, 1997). A longitudinal study of CPC participants during the years 1983-89 shows a strong positive association (0.17 effect size in reading and 0.19 in math achievement) between participation and school performance in eighth grade, with the strongest results (0.52 effect size in reading and 0.47 in math) for children who participated in extended intervention through third grade (Reynolds, 1997). A cost-benefit analysis using data on participants through age 21 reflects a return to society of $7.14 per dollar invested in the preschool program, and a $6.11 per dollar

return for the extended intervention (ages 4-6), with benefits accruing through increased economic participation and tax revenues and decreased costs in remedial education and criminal justice costs (Reynolds et al., 2002). The CPC intervention, by design, is continued into the early elementary school years, providing supportive subsequent environments that may extend the benefits of the preschool intervention. The reduction in return on investment (from $7.14 to $6.11) may be due to the high cost of providing continued support.

A number of states and local governments provide either targeted or universal pre-kindergarten for children. Analyses using experimental and quasi-experimental designs in Tennessee, New Jersey, Oklahoma, and North Carolina are mixed. In 2009, Tennessee began a longitudinal study to determine the effects of participation in its Voluntary Pre-K program (TN-VPK) for students at kindergarten entry, and through third grade. The state began providing TN-VPK in 2004 to four-year-old children who qualify for federal free or reduced-price lunch (FRPL), children with disabilities, and English Language Learners (ELLs). Preliminary third grade results, released in 2016, show that positive effects on achievement and school readiness at the end of pre-k dissipated in subsequent years, and by the end of third grade, pre-k participants on average scored lower than non-TNVPK participants (Lipsey et al., 2016). Fadeout of the initial positive impacts has been attributed, by the study authors, to issues of quality within the preschool programs as well as a lack of alignment between preschool and subsequent early elementary classroom learning, particularly in grades kindergarten through three (Lipsey et al., 2016).

New Jersey implemented state-funded public preschool in response to the New Jersey Supreme Court school-funding case, *Abbott v. Burke* (Farrie, 2014). The program, which began in 1999, provides full-day preschool for 3-and 4-year-old children in 31 high poverty districts.

While TN-VPK classrooms are almost entirely housed in public schools (Lipsey et al., 2016), Abbott preschool classrooms have a mixed public-private delivery system that includes private care centers, Head Start agencies, and public schools (Barnett et al., 2013). A comprehensive evaluation of Abbott preschool began in 2004. A regression discontinuity design was used to estimate effects at kindergarten entry, and a covariate adjusted regression analysis for longer-term outcomes. In the adjusted regression analysis, Abbott students were compared to students who shared their elementary school classrooms but did not attend Abbott preschool (Barnett et al., 2013). The fifth-grade follow-up study, released in 2013, finds persistent effects for students who attended Abbott pre-k, particularly for those who attended two years of pre-k (versus only one year). This longitudinal study finds significant longer-term effects, defying the overall trend of other public preschool studies. However, it is worth noting that the comparison sample of students, drawn from students attending the same kindergarten classrooms as Abbott students but without receiving Abbot pre-k, used in the evaluation is quite different than the sample of students receiving Abbott pre-k. Abbott pre-k students in the sample are more likely to be male, have parents with a high school education who are employed, speak English at home, and less likely to qualify for FRPL (Barnett et al., 2013). The differences in demographic characteristics make it difficult to discern whether positive impacts were specifically due to the groups of students who were treated (recall that ELL students experienced sustained positive impacts in the Tennessee evaluation as well), or if schools in these New Jersey districts were better equipped to provide sustained support for at-risk students, thus preventing fadeout of early gains from preschool.

An evaluation of Oklahoma's universal pre-k in Tulsa shows mixed effects. A study following two cohorts of students through third grade reflects persistent effects on math for the

later cohort (those in third grade in 2009-10), but no persistent effects for the earlier cohort (participants in third grade in 2004-05) (Hill, Gormley, & Adelstein, 2015). The authors note that the accountability movement of the early 2000s may have had some contributing effect in the later cohort, as well as the pre-kindergarten program implementation reaching maturity in the later years. These changes potentially point to growth in pre-kindergarten through third grade alignment that supported more consistent student growth, thus avoiding fadeout of initial effects.

North Carolina has provided statewide pre-kindergarten for high-risk 4-year-old children since 2001. The most recent evaluation of the North Carolina *More at Four* (MAF) program finds positive effects on reading and math achievement and reductions in special education placement and grade retention through Grade 5 (Dodge et al., 2017). These effect sizes may even be underestimated, as the identification strategy used estimated effects at the county level, regardless of specific student participation. Yet, the estimates may also disguise heterogeneity in long-term effects within districts, if students across elementary classrooms experience converging growth trajectories that we cannot disentangle at this level of aggregation. Converging growth trajectories would reflect a 'fadeout' of observable preschool effects, as achievement differences between intervention participants and non-participants become statistically indistinguishable.

Finally, a recent study using both waves of the ECLS-K finds that the association between preschool participation and literacy outcomes decreases between the fall of kindergarten and third grade (Bassok et al., 2018). The authors find a positive, statistically significant association between preschool attendance and literacy upon kindergarten entry. However, for children on average in both cohorts, this association between preschool attendance and literacy outcomes is statistically insignificant by third grade; for low-SES and Hispanic children, the

association decreases much more rapidly, and is indistinguishable from zero by the end of kindergarten (Bassok et al., 2018). The study explores a set of potential moderators, including full-day kindergarten, teacher transition practices, exposure to advanced content, and class size, and does not find statistically significant associations (Bassok et al., 2018). This investigation into the potential role of subsequent classroom support finds null results; however, further research is warranted in attempting to disentangle these potential mechanisms, and understand how these factors may influence individual student growth.

Conceptual Framework

Sustained Environments and Counterfactual Catch-up

The two most prominent explanations for decreasing differences in outcomes for students who did and did not attend preschool are counterfactual student catch-up and treatment fadeout. Counterfactual student *catch-up* occurs when students without preschool experience begin kindergarten and acquire academic skills at a faster pace than their preschool-attending peers. Non-preschool students enter kindergarten, academically behind children who attended preschool, and teachers begin their instruction at the most basic level. For students with no formal early learning experience, they begin formal learning in kindergarten. Students with preschool, according to the counterfactual catch up hypothesis, are sitting through instruction they have already covered in previous years and not continuing to learn new skills. Therefore, the counterfactual group of students, those without preschool, start to "catch up" to their preschool attending peers. When teachers pace instruction behind what students already know, higher order skills may actually atrophy and what appears is a "fadeout" effect of the preschool intervention. When this occurs achievement of the preschool participants actually falls until

there is no significant difference between preschool attendees and their non-attending peers. In practice, these two different reasons are measured by the same phenomena, reducing the difference between the skill levels of preschool participants and the children who did not attend preschool.

The *sustained environments* theory considers early elementary education from the perspective of students who attended preschool—they enter kindergarten with a higher base level of academic skill, and to maintain that advantage, students need exposure to increasingly advanced content. Many researchers point to the need for sustained (supportive) environments in order to prevent fadeout of preschool intervention effects (Bailey et al., 2017; Stipek et al., 2017). The term *sustained environments* refers to the hypothesis that alignment between pre-kindergarten and early elementary school learning is critical in maintaining, and building upon, the gains resulting from early intervention. Sustained environments are those in which high quality pre-kindergarten is followed by exposure to instructionally-aligned early elementary experiences that provide supportive learning environments for continued growth.

A commonly cited example of this effect is a study which finds, counterintuitively, that preschool benefits persist longer for students in relatively larger elementary school classrooms (Magnuson et al., 2007). Using the first wave of the ECLS-K data, Magnuson and colleagues (2007) test whether subsequent classroom factors are associated with fadeout of preschool effects. The authors specifically focus on class size and the level of academic instruction provided in kindergarten, first, and third grade. Using multivariate regression models with extensive control variables, the authors find that any initial differences in school readiness, as measured by reading and math achievement, are eliminated for children in smaller classrooms and classrooms with high levels of reading instruction (Magnuson et al., 2007). These results

suggest that non-preschool peers "catch up" faster in small classrooms with high levels of reading instruction versus non-preschool peers in larger kindergarten classrooms. The authors conclude that non-preschool attending peers are mostly benefitting from the identified classroom quality factors (Magnuson et al., 2007). This finding is in contrast with the sustained environment theory, where measures of higher quality are expected to support persistent preschool effects. This indicates that there are elements that may constitute classroom quality that are not elements of supportive environments in the *sustained environments* framework.

Mechanisms

I consider several frameworks (including, mainly, skills-beget-skills and proximal processes) for early learning trajectories, and discuss the role that kindergarten and early elementary environments play in promoting strong academic growth. The goal is to provide a framework for the concept of *sustained environments*, and describe why subsequent experiences influence the observed fadeout of preschool effects. I aim to address two questions: What do instructionally supportive subsequent environments offer? How might subsequent environments alter learning trajectories?

As previously noted, a commonly cited theory supporting early investment is the "skills-beget-skills" hypothesis (Cunha & Heckman, 2007; Heckman, Pinto, & Savelyev, 2013; Heckman, 2006). Under this concept, lost opportunities for young children impose limits on later-in-life capacity because they have failed to develop the skills required for school success. Skills developed early on are leveraged to raise productivity at later stages (Cuhna & Heckman, 2007). Cuhna and Heckman use the term *self-productivity* to refer to the idea that "skills produced at one stage augment the skills attained at later stages" (2007, p. 35). The idea that

skills are self-reinforcing and build on each other is in indirect conversation with the concept of proximal development as described by developmental psychologists Bronfenbrenner and Vygotsky. Bronfenbrenner's *bioecological model* posits that there is reciprocal interaction between individuals and their environment, and that developmentally effective activities must become increasingly more complex (Bronfenbrenner, 1994; Bronfenbrenner & Morris, 1998). Bronfenbrenner builds on Vygotsky's *zone of proximal development,* which describes the area of difference between a child's individual task capacity and their ability to perform with aided support (Vygotsky, 1978). The *zone of proximal development* is the developmental area that should be targeted by teachers to promote continued learning growth in a supportive classroom environment. This idea is further developed in Bronfenbrenner's *proximal processes* (1999). Bronfenbrenner refers to regular, enduring forms of interaction (between teachers and students, students and students, or students and parents) that occur in the immediate learning environment of an individual as *proximal processes* (Bronfenbrenner, 1999). The proximal process becomes the most important predictor of developmental outcomes, and may vary for individuals based on environmental circumstances. In terms of early education, young children interact with their home environment in early years, and if home environments fail to provide cognitively and socially stimulating exchanges, children will not develop language, interaction, or motor skills to their full capacity. Preschool and other early childhood interventions can provide stimulating school environments to "make up" for or build upon experiences in students' home environments. Young students that have attended preschool in which they developed higher skill levels than children who do not attend require, in subsequent years, exposure to more complex challenges in school in order to continue their developmental trajectory. The two figures below reflect a classroom where individual students' *zone of proximal development* is targeted by

teachers and learning continues at a steady, developmentally-appropriate rate (Figure 1), and a classroom where instruction is targeted to achievement levels, with a focus on the lowest-achieving group of students on average (Figure 2).

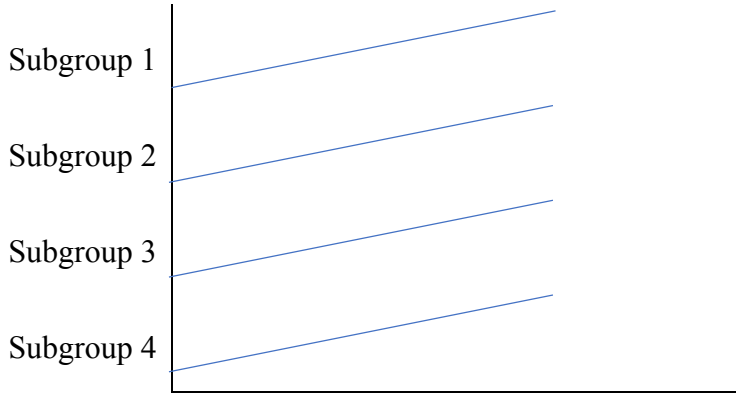


Figure 1. Sustained Environments with Steady Growth:
Consistent trajectories with significantly different initial starting points post-preschool; subsequent environments support steady growth

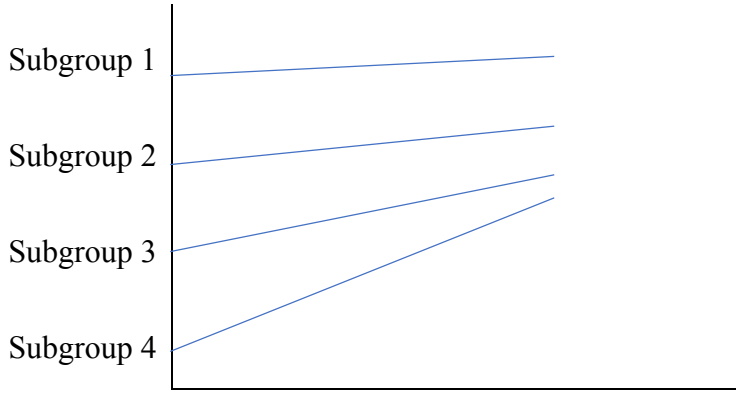


Figure 2. Converging Trajectories or 'Fadeout':
Subsequent environments support faster growth for students with inconsistent or no preschool experience, but do not support developmental growth for other students at the same pace

The question, then, is what are the qualities of subsequent environments that are most important for supporting proximal processes? One likely characteristic of a sustaining environment is differentiated content instruction. Differentiated content instruction that targets students at an appropriate developmental level may be underlying the observed associations

between exposure to advanced content and persistence of preschool benefits. Teacher use of differentiated instruction, as reported in the ECLS-K survey, has been associated with positive gains in reading achievement (McCoach, O'Connell, & Levitt, 2006). This classroom practice reflects an attempt to target students at their individual *zone of proximal development*, and building on individual initial skill levels. This practice may contribute to continued differences in growth trajectories between students based on different preschool experiences.

Another piece of the "skills-beget-skills" hypothesis addresses student motivation, persistence, and other skill-building social-emotional skills (Cuhna & Heckman, 2007; Heckman, 2006). Heckman and colleagues test the influence of personality traits on observed Perry Preschool effects, and find that changes in these traits explain a significant portion of observed adult outcomes (Heckman, Pinto, & Savelyev, 2013). This may happen through *dynamic complementarity*, where early gains in self-regulation skills or motivation are bolstered to support cognitive achievement in later stages (Cuhna & Heckman, 2007). This is again reflective of Vygotskian theories of child development. The transition from preschool to kindergarten and the more formal elementary school environment is considered a "critical period" (Vygotsky, 1978). During critical periods, children are experiencing dramatic shifts in internal processes and social interactions. The "preschool age" critical period, where children are developing self-regulation skills and new mental models of social learning, can occur between ages three to seven (Vygotsky, 1978). The alignment of preschool and kindergarten features, such as providing opportunities to practice emotional skills and self-regulation, promotes a consistent understanding of school-and-learning appropriate behavior for children in the formal learning environment. Preschool programs vary in their overall goals and methods of preparing students for elementary school. Some programs may be more successful than others in providing

intangible skills such as inquisitiveness, confidence, perseverance, and other "soft" skills that provide a foundation for lifelong learning. These skills may support continued student growth regardless of subsequent classroom environments; students who are able to self-direct their learning experiences may be more successful in sustaining learning gains even in the absence of supportive, developmentally-appropriate instruction. We may see longer-term effects on both social-emotional and academic outcomes from these programs when compared to preschools that focus more on a direct-instruction curriculum.

The approaches to learning scale used in the ECLS-K is a useful measure in this analysis, as it includes components on student persistence, emotional regulation, and attentiveness (Tourangeau et al., 2015). This scale, included in parent and teacher surveys in ECLS-K data collection, includes a set of Likert scale questions regarding student behavior. The scale was developed by the National Center for Education Statistics (NCES) and has a reliability estimate of 0.91 (Tourangeau et al., 2015). Behaviors such as paying attention in class and completing tasks independently have been associated with stronger academic skills in kindergarten and first grade (U.S. Department of Education, 2001; Snyder, de Brey, & Dillow, 2016). Using data from the first wave of ECLS-K, Li-Grining and colleagues use longitudinal growth modeling to observe the moderating influence of student approaches to learning (ATL) in their reading and math achievement trajectories (2010). This measure rates students on proficiency in skills such as persisting in completing tasks, working independently, and following classroom rules. They find a positive, beneficial link between ATL and reading and math performance for all students regardless of demographics; although there were differences based on gender and achievement levels at the beginning of kindergarten. This evidence supports the hypothesis that student motivation and early learning skills may influence how they interact with later educational

environments. The authors note that this may be particularly beneficial for at-risk students, for whom skills related to perseverance are critical in adverse learning environments (Li-Grining et al., 2010).

Methods

Data

The current study uses the NCES Early Childhood Longitudinal Study (ECLS) of Kindergarten; Kindergarten Class of 2010-11 (ECLS-K:2011). This is a nationally representative, longitudinal sample of children who attended kindergarten in 2010-11. This study systematically samples kindergarten students in the base year and will follow them through to eighth grade. The study is currently in the eighth year of data collection, and the National Center for Education Statistics (NCES) has released data on the kindergarten, first, third, and fourth grade years. Achievement assessments, teacher, parent, and school surveys are collected in each wave (twice in kindergarten and in the spring of first, second, third, fourth, fifth, and eighth grade). The ECLS-K:2011 includes approximately 18,200 students in 970 schools. The ECLS-K sample is constructed through a three-stage cluster sampling method, with stratified sampling within clusters: counties, schools, and students (Tourangeau et al., 2015). This multistage sampling design allows researchers to use the sample, with appropriate weights, to produce estimates for the full population of kindergarten students in the United States in 2010-11. Probability weights provided by NCES are used throughout the analysis. This data has been used to study academic trajectories, early predictors of later outcomes, and achievement and behavior outcomes by student subgroups (Chatterji 2005, DiPerna et al. 2007, Engel, Claessens, & Finch,

2013; Li-Grining et al., 2010; Magnuson et al., 2007; McCoach et al., 2006; Roberts, Mohammed, & Vaughn, 2010).

Measures: Reading and math achievement

The main outcome of interest in this study is academic achievement and student growth trajectories, which require repeated measures over time. The achievement outcomes are measured as IRT scores in math and reading assessments administered by NCES. Although studies on long-term preschool impacts point toward the important influence of preschool on non-achievement outcomes including grade retention, high school completion, health and behavioral measures, these measures are beyond the scope of this study. Social-emotional and behavioral outcomes in the ECLS-K:11 dataset are highly clustered and show little variation over time (McFarland et al., 2018). As the current study is interested in student growth trajectories, it is most appropriate to use achievement outcomes, particularly given the data limitations.

Measures: Preschool attendance

I rely on a parent survey indicator of student pre-kindergarten educational experiences. The parent surveys of the ECLS-K have been used extensively in quantitative research (Bassok et al., 2018; Chatterji, 2006; Claessens, Engel, & Curran, 2013), including the specific indicators of pre-kindergarten experiences used in the current study (Magnuson et al., 2004; Magnuson et al. 2007; Loeb et al. 2007). In their analysis of preschool attendance and cognitive development, Loeb and colleagues collapse the parent survey responses in order to account for some potential parent confusion regarding the difference between different types of center-based care (Loeb et al., 2007). I use these same distinctions, but include an additional category for State pre-k, which

was unavailable in the ECLS-K: 98 survey. The four categories of preschool attendance I use are: Head Start, State Pre-K, Private/Center-based care (not HS or State), or Other (including no pre-k). There are several groups of children included in the 'Other' category: some did not attend any preschool or have out-of-home care, some had relative care, and some attended nursery or daycare. The distinction between daycare and preschool is the assumption that preschool centers have some academic or developmental curriculum.

Measures: Instructional differentiation

I examine the potential influence of teacher instructional differentiation on growth trajectories. As discussed in the conceptual framework for this paper, research and theory support the notion that developmentally appropriate teaching should focus on both the child's skills and interaction with instructional content (Vygotsky, 1978; Claessens et al., 2014; Sarama & Clements, 2009). Teacher use of instructional differentiation is one measure that indicates whether teachers are attempting to teach children at their individual level. The ECLS-K teacher survey includes items specifically asking teachers if they use ability grouping in reading and math instruction. Use of within-class ability grouping implies that teachers are providing adaptive instruction to students according to their incoming skill level (McCoach et al., 2006). I use these questions, re-coded as binary dummy variables, to indicate whether a teacher does or does not use differentiated instruction in their classroom. Although this is a coarsened measure of differentiation, studies using this indicator have found a positive association between ability grouping and kindergarten achievement (Bodovski & Farkas, 2007; McCoach et al., 2006).

Measures: Social-Emotional Skills

The ECLS-K:11 data includes a composite measure of teacher-rated student *approaches to learning*. This scale asks teachers to rate students in seven behaviors: paying attention, persisting in completing tasks, showing eagerness to learn new things, working independently, adapting easily to changes in routine, keeping belongings organized, and following classroom rules. Early skills in these areas may provide a foundation for school behavior that supports faster acquisition of academic skills (Cunha & Heckman, 2007). If these behaviors are taught and practiced in preschool programs, this may affect how students respond to subsequent instruction. The *approaches to learning* measure is one of five subscales of a modified version of the Social Skills Rating System (Gresham & Elliott, 1990), adapted by NCES (Tourangeau et al., 2015). The scale, while reliable, may not fully capture the underlying behaviors of interest in the *skills beget skills* hypothesis. The scale does not distinguish between following classroom rules, a routine practice set by teachers, and showing an eagerness to learn, which may be more tied to student personality. Additionally, the rating is based on teacher observations of student behavior, which may be compromised particularly in larger classes where the teacher has many demands on her time and attention. While this measure serves as one indicator of social-emotional skill, results may be attenuated for these reasons.

Analytic Strategy

The questions addressed in this study focus on student growth trajectories in early elementary school. Ideally growth would be measured from baseline scores obtained at the beginning of preschool treatment through the end of preschool and beyond. However, due to data limitations in the ECLS-K, I use kindergarten entry scores as a baseline proxy for early childhood learning. The current study does not attempt to address preschool effects—rather, the

study focuses on student achievement growth and experiences after preschool. The results of the study add to the existing evidence on preschool fadeout or catch up by exploring the associations between both subsequent classroom factors and early skill development and academic growth trajectories beginning at the end of preschool. A multilevel model is appropriate for addressing this question because it can model individual student growth over time, assess differences in learning trajectories between student subgroups, and test the potential influence of their learning environment and early development of skills such as those measured by ATL. The growth curve model utilizes a multilevel model where *time* is a level one, time-varying variable, and time-invariant student measures are included in level 2 as predictors of growth[1]. The second level of variation allows me to estimate differences across groups of students with different types of pre-kindergarten experiences. This allows for modeling change processes that we would expect to occur in young students who have been exposed to different preschool experiences (Ram & Grimm, 2015; Singer & Willett, 2003). I estimate the following models for reading and mathematics outcomes in kindergarten through third grade.

Research Question 1: To what extent are there differences in student growth in reading and math achievement over time, based on preschool attendance?

Reading

Level 1

$$Reading_{it} = \pi_{0i} + \pi_{1i}Time_{it} + \varepsilon_{it}$$

Level 2

$$\pi_{0i} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \alpha_{0i}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{11}HeadStart_i + \gamma_{12}State_i + \gamma_{13}Other_i + \alpha_{1i}$$

Composite

[1] The number of levels for hierarchical linear modeling (HLM) are typically dictated by an examination of variation at each level (Singer & Willett, 2003). However, two-level growth models, such as those used here, are intrinsically interested in the first level growth indicators as predicted by second level, time invariant characteristics.

$$Reading_{it} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \gamma_{10}Time_{it} + \gamma_{11}HeadStart_i * Time_{it} + \gamma_{12}State_i * Time_{it} + \gamma_{13}Other_i * Time_{it} + \alpha_{0i} + \alpha_{1i}Time_{it} + \varepsilon_{it}$$

Math

Level 1

$$Math_{it} = \pi_{0i} + \pi_{1i}Time_{it} + \varepsilon_{it}$$

Level 2

$$\pi_{0i} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \alpha_{0i}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{11}HeadStart_i + \gamma_{12}State_i + \gamma_{13}Other_i + \alpha_{1i}$$

Composite

$$Math_{it} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \gamma_{10}Time_{it} + \gamma_{11}HeadStart_i * Time_{it} + \gamma_{12}State_i * Time_{it} + \gamma_{13}Other_i * Time_{it} + \alpha_{0i} + \alpha_{1i}Time_{it} + \varepsilon_{it}$$

In this model, $Reading_{it}$ and $Math_{it}$ represent the outcome measure for student *i* at time *t*. Achievement is measured twice in kindergarten and first grade, and once in third grade. $Time_{it}$ is a measure of the age of student *i,* at the time of testing, *t*. At Level 2, there are four binary categories for preschool enrollment: Head Start, State preschool, Center-based care (non-HS or State), and Other (including non-center based or relative care), indexed for student *i*, and are time invariant. The reference category is center-based (non-Head Start or State) care. The interpretation of coefficients, then, is for students in each of the other groups compared to those in center-based care. The larger policy question is about the impact of *public* preschool programs, so I am choosing to compare those programs to private early childhood education. In the composite model, the intercept is labeled as γ_0, and is centered at the time of third grade assessment. Centering at the final data point allows the interpretation of the intercept as the end-of-trajectory differences (Singer & Willet, 2003). This allows for testing of differences across groups at the end of third grade[2]. The coefficients of interest here are the interactions between

[2] I run a sensitivity check by centering *time* at kindergarten entry to test the initial differences to exposure, and find substantively similar results.

Head Start and *Time* (γ_{11}), and *State* and *Time* (γ_{12}), which indicate differences in growth trajectories by different preschool program. I initially conduct the analysis without covariates, and then add non-time-varying covariates (gender and student race/ethnicity), to control for potential imbalance among covariates.

Research Question 2: To what extent is achievement associated with subsequent classroom exposure to teacher instructional differentiation?

Reading

Level 1

$$Reading_{it} = \pi_{0i} + \pi_{1i}Time_{it} + \pi_{2i}Instruction_{it} + \varepsilon_{it}$$

Level 2

$$\pi_{0i} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \alpha_{0i}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{11}Hea\square Start_i + \gamma_{12}State_i + \gamma_{03}Other_i + \alpha_{1i}$$
$$\pi_{2i} = \gamma_{20}$$

Composite

$$Reading_{it} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \gamma_{10}Time_{it} + \gamma_{11}HeadS\square art_i * Time_{it} + \gamma_{12}State_i * Time_{it} + \gamma_{13}Other_i * Time_{it} + \gamma_{20}Instruction_{it} + \alpha_{0i} + \alpha_{1i}Time_{it} + \varepsilon_{it}$$

Math

Level 1

$$Math_{it} = \pi_{0i} + \pi_{1i}Time_{it} + \pi_{2i}Instruction_{it} + \varepsilon_{it}$$

Level 2

$$\pi_{0i} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \alpha_{0i}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{11}HeadStart_i + \gamma_{12}State_i + \gamma_{03}Other_i + \alpha_{1i}$$
$$\pi_{2i} = \gamma_{20}$$

Composite

$$Math_{it} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \gamma_{10}Time_{it} + \gamma_{11}HeadStart_i * Time_{it} + \gamma_{12}State_i * Time_{\square t} + \gamma_{13}Other_i * Time_{it} + \gamma_{20}Instruction_{it} + \alpha_{0i} + \alpha_{1i}Time_{it} + \varepsilon_{it}$$

Where, $Reading_{it}$ and $Math_{it}$ represent individual *i* achievement at time *t*, predicted by a growth model that accounts for teacher instructional differentiation (measured as ability grouping) measured in kindergarten, first, second (for a survey subsample of students and their teachers), and third grade. *Instruction* serves as a dichotomous indicator for student *i* exposure to

differentiated instruction at time *t*. Level 2 predictors test whether the type of preschool participation is associated with achievement level at the end of third grade (intercept, $\gamma_{01}, \gamma_{02}, \gamma_{03}$), and individual growth trajectory ($\gamma_{11}, \gamma_{12}, \gamma_{13}$). A statistically significant coefficient on the *Instruction* coefficient (γ_{20}) would suggest that this factor is associated with deviations from the average growth trend. This is the main coefficient of interest for research question two. As in the analysis plan for Research Question 1, I run the analysis both with and without time-invariant covariates.

Research Question 3: To what extent are growth trajectories mediated by social-emotional skills, as measured by teacher ratings of student approaches to learning?

Reading

Level 1

$$Reading_{it} = \pi_{0i} + \pi_{1i}Time_{it} + \pi_{2i}ATL_{it} + \varepsilon_{it}$$

Level 2

$$\pi_{0i} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \alpha_{0i}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{11}HeadStart_i + \gamma_{12}State_i + \gamma_{13}Other_i + \alpha_{1i}$$
$$\pi_{2i} = \gamma_{20}$$

Composite

$$Reading_{it} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \gamma_{10}Time_{it} + \gamma_{20}ATL_{it} + \gamma_{11}HeadStart_i * Time_{it} + \gamma_{12}State_i * Time_{it} + \gamma_{13}Other_i * Time_{it} + \alpha_{0i} + \alpha_{1i}Time_{it} + \varepsilon_{it}$$

Math

Level 1

$$Math_{it} = \pi_{0i} + \pi_{1i}Time_{it} + \pi_{2i}ATL_{it} + \varepsilon_{i}$$

Level 2

$$\pi_{0i} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \alpha_{0i}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{11}HeadStart_i + \gamma_{12}State_i + \gamma_{13}Other_i + \alpha_{1i}$$
$$\pi_{2i} = \gamma_{20}$$

Composite

$$Math_{it} = \gamma_{00} + \gamma_{01}HeadStart_i + \gamma_{02}State_i + \gamma_{03}Other_i + \gamma_{10}Time_{it} + \gamma_{20}ATL_{it} + \gamma_{11}HeadStart_i * Time_{it} + \gamma_{12}State_i * Time_{it} + \gamma_{13}Other_i * Time_{it} + \alpha_{0i} + \alpha_{1i}Time_{it} + \varepsilon_{it}$$

Here, a measure of student learning skills, approaches to learning (ATL) measured for student *i* over time *t*. The measure is collected from teachers at the same time as outcome testing. In this model, I use the kindergarten entry measure as a proxy for skills obtained in preschool, and as a potential mediator of the slopes that represent overall growth trajectories. A significant coefficient on the *ATL* predictor (γ_{20}), may indicate that these behavioral skills mediate the effect of preschool on short-and long-term outcomes. To determine if *ATL* is a mediator, I follow Baron and Kenny's (1986) steps to test mediation, testing for the association between the independent variable (preschool type) and mediator (approaches to learning), the independent variable and outcome (reading or math achievement), mediator and outcome, and comparing the effect of the independent variable on the outcome with and without the mediator present. The authors lay out three steps for testing mediation: (1) regress the mediator on the independent variable, (2) regress the dependent variable on the independent variable, and (3) regress the dependent variable on both the independent variable and on the mediator (Baron & Kenny, 1986, p. 1177). To confirm mediation, the independent variable must affect the mediator and dependent variable, the mediator must affect the dependent variable, and finally the effect of the independent variable on the dependent variable must be smaller when the mediator is present.

Results

Study Sample

In the final analysis sample, 14% of students attended a Head Start preschool program, 18% attended a State preschool, 51% had Center-based (private) care, and 15% did not attend preschool, or attended an informal, nonacademic program. Table 1 shows descriptive characteristics of the sample, including student characteristics, rates of preschool attendance and

conditional predictor means. Students who attended Head Start programs are more likely to be Black (34% of Head Start students) or Hispanic (37%) than students attending State preschool (8% Black, 25% Hispanic) or Center (9% Black, 11% Hispanic) preschool. Asian students are more likely to attend Center-based preschool (6%) than other categories (1 – 4%). Sample t-tests were conducted between groups for descriptive, predictor, and outcome measures, and *p*-values are shown in Appendix A, Tables A1 and A2. Interestingly, in almost every instance, students with different preschool experiences were equally likely to be exposed to instructional differentiation in early preschool. Unsurprisingly, there are significant differences in outcome scores across all groups at Kindergarten entry. Significant differences remain, with the exception of comparisons between State preschool attendees and students who did not attend preschool or had non-center-based care ('Other' category), where scores converge in later time points (Appendix Table A2).

The analytic sample was restricted to students with recorded information on preschool attendance, and had a math and reading test score for the fall and spring of Kindergarten, spring of first grade, and spring of third grade. A subsample of students was tested in the fall of first grade and fall/spring of second grade, and the growth model allows me to use this data without dropping observations. This results in approximately 10,200 overall students and approximately 49,000 student-test observations[3]. Students in the analytic sample, compared to those for whom preschool information is not identified or test scores unavailable, are more likely to be White (56% in analytic sample to 40% not in sample), and less likely to be Hispanic (20% in sample to

[3] This analytic sample size is comparable to other studies using the ECLS-K to study preschool attendance and associated elementary outcomes, i.e., Bassok, Gibbs, & Latham, 2018.

38%). Again, *p*-values of statistical differences between sample and non-sample children for descriptive and predictor characteristics are shown in Appendix Tables A1 and A2.

Model Building and Exploring Differences in Growth

Tables 3 and 4 show the results of the basic growth model to address my first research question, where test score growth in reading (Table 3) and math (Table 4) is modeled as a function of time and preschool attendance. I begin with a null model and an unconditional growth model (Models A and B, respectively). Moving from Model A to Model B, including *Time* in the unconditional growth model, the within-person variance estimate decreases by 87.6% (reading score) and 91.5% (math), indicating that approximately 88-92% of within-person variation is systematically associated with linear *Time*. As expected, the growth model is highly appropriate for modeling this change.

In Model C, I include the preschool predictors for the Level 1 slope and intercept. Interpreting the results for reading first from Table 3, a student who attended Center-based preschool ends third grade, on average, with a reading scale score of 119.54, with a growth rate from kindergarten to third grade of 1.44 scale points monthly, on average. The intercept (third grade score) for students who did not attend Center based preschool is significantly lower (p <.001), from 5 to 10 points on average. There are significant differences in outcomes for students in Head Start and State preschool, and Head Start and 'Other', but not between students attending State or 'Other' preschool. Model D includes student-level covariates, race/ethnicity and gender. When covariates are included, point estimates on preschool attendance are qualitatively similar but somewhat smaller (decreasing by about 1 unit from Model C to Model D); some of the differences across preschool attendance group are picking up differences across

racial groups (as expected, since there are significant differences in preschool attendance by race).

Identical models are used for math outcomes, shown in Table 4. Column 4 shows the full growth model with covariates; here, a student who attended Center-based preschool has a third-grade math score of 105.85, on average, and an average growth rate of 1.56 scale points per month. Students who attended Head Start have a significantly different third-grade score, on average, 10.33 points below Center-based attendees. State preschool and 'Other' preschool students also perform worse, on average, than their Center-based attending peers (by approximately 5 scale points, Table 4 Column 4). The differences between Head Start and State, and Head Start and 'Other' estimates are statistically significant (p <.001), but there is not a significant difference between State and 'Other' preschool attendees (p = .187). This indicates convergence between these two groups, while there are still significant differences across others.

Next, I test for differences in average growth trajectories, or student rate of growth. Based on my conceptual framework and study of previous literature, I hypothesize that student trajectories will converge across groups, as students are exposed to subsequent early elementary environments. The estimated coefficients for the relevant predictors are seen in each table under the heading "Rate of change." Parameter tests of these variances can indicate whether student growth is becoming "more similar" or different (Krull et al., 2015). I find that rate of growth across groups is statistically significant and distinct for Head Start, State, and Center-based preschool; however, the differences are not necessarily meaningful. While average rate of change for students attending Center-based preschool is 1.44 score points per month for reading and 1.56 for math, the rate for Head Start students is 1.395 score points per month in reading and 1.502 for math (Columns C in Tables 3 and 4). The differences between State and Center preschool are

smaller (.027 score point difference for reading and .02 points for math). These are compared to coefficients of the average association between Head Start and State preschool attendance, ranging from 5.63 to 12.33 point differences.

Overall, differences in math and reading scores between each preschool group present at Kindergarten entry remain until third grade, with the exception being State preschool and 'Other' preschool attendees. Here assessment scores do converge with each other by the end of third grade (*p*-value testing for differences between State and Other = 0.378 for reading, = 0.187 for math; this can also be seen visually in Figures 3 and 4). The rate of change (slope) is statistically different for Center-based preschool and Head Start, State, and 'Other' preschool ($p < .001$). For reading scores, growth is also different between Head Start and 'Other' ($p < .001$) and State and 'Other' ($p < .01$). For math, these differences are significant for Head Start and State preschool ($p < .001$) and Head Start and 'Other' preschool ($p < .001$). However, for students in Head Start and State preschool, their trajectories in reading growth are indistinguishable from each other. This is also true for State and 'Other' students in math. The persistence of differences, both in outcome scores and growth rate, rejects hypotheses of 'fadeout' or 'convergence' across groups.

Subsequent Environments: Instructional Differentiation

Table 5 shows the association between reading and math outcomes and the use of classroom ability grouping, controlling for preschool attendance. In this model, I hypothesize that there is some heterogeneity being masked by an incomplete growth model. Here I include differentiated instruction (measured as ability grouping) as a predictor of achievement, in addition to preschool attendance. Indeed, the use of differentiated reading groups has a significant, positive association with final reading score, with an estimated coefficient of 4.9,

which is approximately half the size of the association between Head Start participation and final score (-10.7), and nearly equal to State preschool (-5.61) and 'Other' preschool (-6.26) participation. The association is statistically significant but much smaller, an estimated difference of 1.14 scale score points, for the use of math groups in predicting third grade math scores. Similar to earlier results, there is no observable difference in growth trajectories across student groups in the sample, based on these predictive models. I do not assume that ability grouping would differentially change the rate of growth for individual students, so it is only included in the Level 1 model. Figures 5 and 6 show the predicted reading and math score trajectories for Head Start, State, and 'Other' preschool groups compared to Center-based attendees, controlling for ability grouping. While the use of reading and math ability grouping narrows achievement differences at the end of third grade, controlling for this factor does not result in obvious changes to average growth trajectories or achievement gaps.

Associations with Student Behavioral Skills

Table 6 shows the results of the growth model accounting for student approaches to learning skills, some of which may have been acquired in preschool. Students who attended Head Start, State, or 'Other' preschool programs have lower reading and math scores through third grade compared to their peers who attended private, Center-based preschool. There is a statistically significant and meaningful association between final achievement outcomes and behavior scores. For reading achievement, an increase in 1-scale score point on the *approaches to learning* measure is associated with approximately 3-score point increase in third grade scores. This is approximately one-third the size of the negative association from participating in Head Start, and over half of the gap for State preschool attendees when compared to private, Center-

based attendees. This association is somewhat smaller for math outcomes. This suggests that increasing student's attention and behavior skills in preschool could provide a relatively large bump towards closing the achievement gap in early elementary school. The growth rate, still, remains similar to that in previous models (i.e., controlling for attention skills does not reflect steeper or flatter growth curves).

Following Baron and Kenny's (1986) approach to testing for mediation, it does appear that the *approaches to learning* measure serves as a mediator[4]. First, the independent variable, preschool attendance category, predicts the mediator (*approaches to learning* score) at the beginning of kindergarten (statistically significant across Head Start, State preschool, Other, and Center with *p*-values <.001). Next, as seen in the multilevel model results in Tables 3 and 4, as well as a typical ordinary least squares (OLS) regression model, preschool attendance category predicts the outcomes, reading and math achievement, across early elementary school (coefficient estimates ranging from -9.04 to 7.87, *p*-values <.001). Third, a standard OLS regression model shows that *approaches to learning* predicts reading and math achievement (point estimates 9.6 and 9.68 for reading and math, respectively, *p*-values <.001). Finally, comparing estimates from models with and without the mediator, preschool attendance estimates are smaller when the mediator is included (ranging from -5.0 to -9.515, *p*-value <.001, versus ranging from -5.63 to -12.33, *p*-value <.001 without the mediator). Estimates from these regressions are shown in Appendix Table A3. This evidence suggests that some of the influence attributed to preschool attendance category in the initial models is reflecting student attention and behavioral skills.

[4] Table of steps and results shown in Appendix A, Table A3.

Robustness Check: Centering and School-Level Fixed Effects

I run several robustness checks, which are described here with tables shown in Appendix B. First, I test for differences at the beginning and end of the trajectories. In order to test for differences at Kindergarten entry, I re-run the models with *age* centered at kindergarten entry rather than third grade. Results are shown in Tables B1-B3. There are significant differences across groups in student reading and math scores at the beginning of kindergarten; students who attended Head Start performed lower than their Center-based peers, as well as their peers in State or 'Other' preschool. By the end of third grade, there are still statistically significant differences in achievement between all student groups (although this difference is rather small between students who attended State preschool and 'Other'). I run an additional robustness check by using 'Other' (indicating no preschool, relative-based care, or non-center based care) as the omitted group. Results are shown in Tables B4-B6. Using this model, estimated coefficients are somewhat smaller but patterns remain consistent across the results, indicating that gaps remain when comparing the three preschool attendance groups to the group of children who did not attend preschool, or had non-center-based care. These statistically-adjusted results follow patterns seen in unadjusted group means. Table 2 shows these unadjusted means, reflecting that students who attended Center-based preschool begin and end early elementary school with higher scores in reading and math.

Across results, I surprisingly do not find evidence of converging trajectories. Students across the sample appear to have very similar growth trajectories, regardless of preschool participation, attention skills, or exposure to instructional differentiation. One reason for this may be because these models are not accounting for school clustering. The indicator for instructional differentiation compares teacher use of this across the nationwide sample, but this may look very

different by context. It may be that using school fixed effects reveals significant differences by school cluster. To test this, I re-run the theoretical models using a school fixed effect instead of the multilevel model. Results are available in Appendix B, Tables B7-B11. Table B7 shows results from the school fixed effects model of the association between preschool attendance and reading outcomes at each of the seven time points, controlling for student characteristics. Preschool attendance is a significant predictor across time points (p <.001). Similar to the results from growth modeling, the association of Head Start participation compared to Center-based preschool is negative and larger than State or 'Other' preschool, (point estimates ranging from -4.2 to -6.7). Estimates for the association between State and 'Other' preschool, compared to Center-based preschool, are also significant and negative, but smaller than those for Head Start, ranging from -1.9 to -4.0 across years. Results for math outcomes are nearly identical, shown in Table B8.

When instructional groups are included as a control, the association between Head Start and third grade reading is -5.06 scale points, on average, using the fixed effects model (Table B9). This is compared to a point estimate of -9.35 using the multilevel growth modeling. While instructional grouping has a statistically significant (p-value <.001) point estimate in the growth models, there is not a significant association when using school fixed effects. This may indicate clustering of the use of instructional grouping as a strategy within schools, to the point where its use does not have a differential effect on students within the school.

Tables B10 and B11 show the association between reading and math achievement and preschool attendance, controlling for *approaches to learning*. Teacher-rated *approaches to learning* is a statistically significant predictor (p <.001) across time waves; point estimates range from 5.5 to 9.9 for reading, and 6.5 to 9.9 for math. This reinforces findings from the growth

models suggesting that student attention skills are a significant factor in academic growth. While this is not necessarily surprising, this may be an important finding to consider when studying preschool efficacy and outcomes; further understanding of this factor could help prioritize curricular or program goals and outcomes.

Discussion

Overall, the results of this study reflect patterns similar to existing information, while adding new evidence on differences across participation subgroups, growth trajectories, and the potential influence of attention skills and differentiated instruction in early elementary school. Students who attend public preschool, Head Start, or no preschool at all, enter kindergarten with lower achievement in reading, math, and less developed attention skills than students who attend private preschool programs. Growth in reading and math is relatively similar across students with different preschool experiences. Instructional differentiation and approaches to learning are both associated with narrowed gaps in achievement by the end of third grade, but do not have a clear effect on rates of growth. This suggests that teacher instruction, when geared toward student achievement levels rather than whole-group or mixed-ability grouping, is helpful for all students, but does not differentially change rate of growth (on average) for student subgroups.

These results are in contrast to reports that do find convergence, for example in the Clements et al. (2013) Building Blocks evaluation and the recent Bassok et al. (2018) study comparing changes over time by using both waves of the ECLS-K. Yet, in contrast to the Building Blocks study, the current study is not an experiment and "instructional differentiation" is a rather broadly defined intervention compared to the specific, research-based Building Blocks curriculum. Therefore, I would not necessarily expect similar results. Bassok et al. find a

diminished preschool effect by third grade when comparing students who attended any type of preschool to those with no preschool experience, suggesting converging trajectories. However, when the authors compare private and public preschool attendees, they find that the observed effect of private preschool persists through third grade (2018), which is reflected in the current study as well. Overall, the results presented here provide some evidence that achievement gaps at kindergarten entry persistent across early elementary school; moreover, the main results and school fixed-effects results suggest that these gaps are not necessarily closed by student exposure to similar subsequent experiences. In this sample of students, growth remains stubbornly consistent within and across groups. This, I feel, is in line with our aggregate understanding of the research on preschool efficacy—context, measures, and research design matter greatly and can have a strong influence on whether or not we find "positive effects" of public preschool.

Contribution and Significance

There is currently significant attention on the question of if and how preschool effects persist (Bailey et al., 2017; Duncan & Magnuson, 2013; Stipek, 2017). The results of this study provide descriptive evidence on the association between subsequent environments, grade-three outcomes, and growth trajectories for students nationwide and across preschool programs. I find that preschool group differences are largely maintained and rather than catch up or fade out, the groups do not converge through the end of third grade (with the exception of State preschool attendees and those that did not attend preschool or had non-center based care). This new evidence on student growth trajectories explores the influence of instructional differentiation and student learning skills when addressing questions about preschool fadeout, and suggests that while these factors may narrow gaps, they are not contributing to differential growth patterns.

The use of longitudinal growth modeling to address this question is novel and supports inferences about differences between and within groups of children with different preschool experiences. Future research in this area should continue to explore the influence of various subsequent environment factors, and continue to consider differences in preschool and subsequent experiences across a wide range of pre-school exposure.

References

Bailey, D.H., Duncan, G.J., Odgers, C.L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness, 10*(1), 7-39. doi: 10.1080/19345747.2016.1232459

Barnett, W.S., Jung, K., Youn, M., & Frede, E.C. (2013). Abbott preschool program longitudinal effects study: Fifth grade follow-up. New Brunswick, NJ: National Institute for Early Education Research.

Barnett, W.S. (2017). Challenges to scaling up effective pre-kindergarten programs. In *The Current State of Scientific Knowledge on Pre-Kindergarten Effects*. Retrieved from: https://www.brookings.edu/wp-content/uploads/2017/04/duke_prekstudy_final_4-4-17_hires.pdf

Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Bassok, D., Gibbs, C.R., & Latham, S. (2018). Preschool and children's outcomes in elementary school: Have patterns changed nationwide between 1998 and 2010? *Child Development*, 1-23. DOI: 10.1111/cdev.13067

Belfield, C.R., Nores, M., Barnett, S., & Schweinhart, L. (2006). The High/Scope Perry Preschool Program: Cost-benefit analysis using data from the age-40 followup. *The Journal of Human Resources, 41*(1), 162-190.

Bitler, M.P., Hoynes, H.W., & Domina, T. (2014). Experimental evidence on distributional effects of Head Start. (NBER Working Paper No. 20434). Cambridge, MA: National Bureau of Economic Research.

Bodovski, K., & Farkas, G. (2007). Mathematics growth in early elementary school: The roles of beginning knowledge, student engagement, and instruction. *The Elementary School Journal, 109*(2), 115-130.

Bronfenbrenner, U., & Ceci, S.J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review, 101*(4), 568-586.

Bronfenbrenner, U., & Morris, P.A. (1998). The ecology of developmental processes. In W. Damon & R.M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 993-1028). Hoboken, NJ, US: John Wiley & Sons Inc.

Bronfenbrenner, U. (1999). Environments in developmental perspective: Theoretical and operational models. In Friedman, S.L., & T.D. Wachs (Eds.), *Measuring environment across the life span: Emerging methods and concepts* (pp. 3-28). Washington, DC: American Psychological Association Press.

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*(1), 42-57.

Campbell, F.A., Pungello, E.P., Kainz, K., Burchinal, M., Pan, Y., Wasik, B.H., Barbarin, O., Sparling, J.J., & Ramey, C.T. (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian project follow-up. *Developmental Psychology, 48*(4), 1033-1043. doi:10.1037/a0026644.

Chatterji, M. (2005). Achievement gaps and correlates of early mathematics achievement: Evidence from the ECLS K-first grade sample. *Education Policy Analysis Archives, 13*(46). Retrieved from http://epaa.asu.edu/epaa/v13n46 .

Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology, 98*(3), 489-507.

Chaudry, A., & Datta, A.R. (2017). The current landscape for public pre-kindergarten programs. In *The Current State of Scientific Knowledge on Pre-Kindergarten Effects*. Retrieved from: https://www.brookings.edu/wp-content/uploads/2017/04/duke_prekstudy_final_4-4-17_hires.pdf

Claessens, A., Engel, M., & Curran, F.C. (2014). Academic content, student learning, and the persistence of preschool effects. *American Educational Research Journal, 51*(2), 403-434. doi: 10.3102/0002831213513634

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.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *The American Economic Review, 97*(2), 31-47.

Currie, J., & Thomas, D. (1993). *Does Head Start make a difference?* (NBER Working Paper 4406). Cambridge, MA: National Bureau of Economic Research. Retrieved from the National Bureau of Economic Research: http://www.nber.org/papers/24406

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics, 1*(3), 111-134. http://www.aeaweb.org/articles.php7doi-10.1257/app. 1.3. Ill

Dodge, K.A., Bai, Y., Ladd, H.F., & Muschkin, C.G. (2017). Impact of North Carolina's early childhood programs and policies on educational outcomes in elementary school. *Child Development, 88*(3), 996-1014.

Duncan, G.J., & Magnuson, K.A. (2013). Investing in preschool programs. *Journal of Economic Perspectives, 27*(2), 109-132.

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

Farrie, D. (2014). *The Abbott Preschool Program: A 15-year progress report*. Education Law Center. Retrieved from: http://www.edlawcenter.org/assets/files/pdfs/publications/AbbottPreschool15YearProgressReportMay2014.pdf.

Gresham, F.M., & Elliott, S.N. (1990). *Social skills rating scale: Elementary scale A ("How often").* Circle Pines, MN: American Guidance Service.

Heckman, J.J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science, 312*(June), 1900-1902.

Heckman, J.J., Moon, S.H., Pinto, R., Savelyev, P.A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool program. *Journal of Public Economics, 94*(2010), 114-128. doi: 10.1016/j.jpubeco.2009.11.001

Heckman, J.J., Pinto, R., Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review, 103*(6), 2052-86. doi: 10.1257/aer.103.6.2052

Hill, C.J., Gormley, W.T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly, 32*(2015), 60-79. http://dx.doi.org/10.1016/j.ecresq.2014.12.005

Krull, J.L., Cheong, J., Fritz, M.S., & Mackinnon, D.P. (2015). Moderation and mediation in interindividual longitudinal analysis in D. Cicchetti (Ed.), *Developmental psychopathology, theory, and method* (pp. 922-985). Hoboken, NJ: John Wiley & Sons.

Li, W., Duncan, G.J., Magnuson, K., Schindler, H., Yoshikawa, H., Leak, J., & Shonkoff, J.P. (2016). Is timing everything? How early childhood education program cognitive and achievement impacts vary by starting age, program duration, and time since the end of the program. Manuscript under review.

Li-Grining, C.P., Votruba-Drzal, E., Maldonado-Carreño, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology, 46*(5), 1062-1077.

Lipsey, M., Farran, D., & Hofer, K. (2016). Effects of a state prekindergarten program on children's achievement and behavior through third grade. Nashville, TN: Vanderbilt University, Peabody Research Institute.

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

Ludwig, J., & Miller, D.L. (2007). Does Head Start improve children's life chances? Evidence from a regression discontinuity design. *The Quarterly Journal of Economics, February*(2007), 159-208.

Magnuson, K.A., Ruhm, C.J., & Waldfogel, J. (2004). Does prekindergarten improve school preparation and performance? (NBER Working Paper 10452). http://www.nberg.org/papers/w10452

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: 10.1016/j.ecresq.2006.10.002

McCoach, D.B., O'Connell, A.A., & Levitt, H. (2006). Ability grouping across kindergarten using an early childhood longitudinal study. *The Journal of Educational Research, 99*(6), 339-346. DOI: 10.3200/JOER.99.6.339-346

McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., Barmer, A., Forrest Cataldi, E., & Bullock Mann, F. (2018). *The Condition of Education 2018* (NCES 2018-144). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018144.

Phillips, D.A., Lipsey, M.W., Dodge, K.A., Haskins, R., Bassok, D., Burchinal, M.R., Duncan, G.J., Dynarski, M., Magnuson, K.A., & Weiland, C. (2017). Consensus Statement from the Pre-Kindergarten Task Force. In *The Current State of Scientific Knowledge on Pre-Kindergarten Effects*. Retrieved from: https://www.brookings.edu/wp-content/uploads/2017/04/duke_prekstudy_final_4-4-17_hires.pdf

Puma, M., Bell, S., Cook, R., Heid, C., & Lopez, M. (2006). *Head Start Impact Study*. New York: Nova Science Publishers.

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). *Third Grade Follow-up to the Head Start Impact Study Final Report,* OPRE Report #2012-45, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Ram, N., & Grimm, K.J. (2015). Growth curve modeling and longitudinal factor analysis. *Handbook of child psychology and developmental science,* 1-31.

Reynolds, A.J. (1997). The Chicago Child-Parent Centers: A Longitudinal Study of Extended Early Childhood Intervention. *Institute for Research on Poverty, Discussion Paper no. 1126-97.* Retrieved from https://www.ssc.wisc.edu/irpweb/publications/dps/pdfs/dp112697.pdf

Reynolds, A.J., Temple, J.A., Robertson, D.L., & Mann, E.A. (2002). Age 21 cost-benefit analysis of the Title I Chicago Child-Parent Centers. *Educational Evaluation and Policy Analysis, 24*(4), 267-303.

Roberts, G., Mohammed, S.S., & Vaughn, S. (2010). Reading achievement across three language groups: Growth estimates for overall reading and reading subskills obtained with the Early Childhood Longitudinal Survey. *Journal of Educational Psychology, 102*(3), 668-686. DOI: 10.1037/a0018983

Sarama, J., & Clements, D.H. (2009). *Early childhood mathematics education research: Learning trajectories for young children.* Routledge.

Schweinhart, L.J. (2005). The High/Scope Perry preschool study through age 40: Summary, conclusions, and frequently asked questions. Retrieved from http://nieer.org/wp-content/uploads/2014/09/specialsummary_rev2011_02_2.pdf

Singer, J.D., & Willett, J.B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* New York, NY: Oxford University Press.

Snyder, T.D., de Brey, C., & Dillow, S.A. (2016). *Digest of Education Statistics 2015* (NCES 2016-014). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Stipek, D., Franke, M., Clements, D., Farran, D., & Coburn, C. (2017). PK-3: What does it mean for instruction? *Social Policy Report, 30*(2).

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Hagedorn, M.C., Leggitt, J., & Najarian, M. (2015). *Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011), User's Manual for the ECLS-K:2011 Kindergarten-First Grade Data File and Electronic Codebook, Public Version* (NCES 2015-078). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

U.S. Department of Education, National Center for Education Statistics, *Entering Kindergarten: A Portrait of American Children When They Begin School: Findings from The Condition of Education 2000*, Nicholas Zill and Jerry West, NCES 2001-035, Washington, DC: U.S. Government Printing Office, 2001.

Vygotsky, L. (1978). Interaction between learning and development. In Gauvain & Cole (Eds.) *Readings on the development of children*. (pp. 34-30). New York: Scientific American Books.

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

Tables

Table 1: Analytic Sample Descriptive Characteristics

Student Demographics	Analytic Sample (Full)	Head Start	State	Center	Other
Female	0.49	0.50	0.46	0.50	0.51
White	0.56	0.22	0.56	0.69	0.44
Black	0.14	0.34	0.08	0.09	0.17
Hispanic	0.20	0.37	0.25	0.11	0.30
Asian	0.04	0.01	0.02	0.06	0.04
Ethnicity - Other	0.06	0.06	0.10	0.04	0.05
Preschool Attendance					
Head Start	0.14				
State	0.18				
Center	0.51				
Other	0.15				
Reading Groups					
Kindergarten	0.59	0.56	0.62	0.56	0.66
First Grade	0.90	0.90	0.90	0.89	0.91
Second Grade	0.89	0.89	0.90	0.87	0.94
Third Grade	0.81	0.88	0.80	0.79	0.85
Math Groups					
Kindergarten	0.19	0.23	0.13	0.19	0.24
First Grade	0.37	0.42	0.44	0.35	0.31
Second Grade	0.50	0.53	0.47	0.51	0.48
Third Grade	0.53	0.57	0.54	0.50	0.54

Note: N ~ 10,200. All means shown at the student level for the main analytic sample. Survey weights provided by NCES are used throughout. See Appendix A for significance test *p*-values.

Table 2: Conditional Outcome Means

Reading	Full Sample	Head Start	State	Center	Other
Fall Kindergarten	52.98	47.93	51.94	55.65	50.59
	(11.33)	(09.47)	(11.48)	(11.22)	(11.08)
Spring Kindergarten	66.92	61.26	65.15	69.62	65.87
	(12.81)	(10.82)	(12.67)	(12.76)	(12.99)
Fall First Grade	74.84	67.72	72.17	78.37	73.95
	(15.68)	(12.53)	(15.57)	(15.82)	(15.14)
Spring First Grade	91.56	83.56	89.74	95.33	89.85
	(15.32)	(14.77)	(15.77)	(14.24)	(14.72)
Fall Second Grade	96.23	89.19	93.55	99.83	95.23
	(13.92)	(14.18)	(13.90)	(12.73)	(13.58)
Spring Second Grade	104.27	98.37	101.5	107.5	103.36
	(12.40)	(13.05)	(13.56)	(10.82)	(11.79)
Spring Third Grade	111.53	106.04	109.39	114.49	110.01
	(11.26)	(11.44)	(12.73)	(09.86)	(10.41)
Math					
Fall Kindergarten	34.94	28.6	34.28	37.91	32.13
	(11.28)	(09.14)	(11.35)	(10.97)	(10.99)
Spring Kindergarten	48.52	42.96	47.21	51.36	46.61
	(11.66)	(10.65)	(11.25)	(11.28)	(11.77)
Fall First Grade	57.61	51.02	54.65	61.28	56.01
	(15.45)	(13.27)	(15.37)	(15.29)	(15.00)
Spring First Grade	72.33	64.73	70.44	76.09	69.87
	(15.60)	(13.77)	(15.30)	(15.26)	(15.05)
Fall Second Grade	77.32	70.63	74.72	81.13	74.92
	(15.17)	(14.65)	(16.35)	(13.86)	(14.63)
Spring Second Grade	87.85	81.05	85.91	91.3	85.77
	(14.22)	(14.24)	(14.95)	(12.85)	(14.19)
Spring Third Grade	99.02	93.42	97.76	101.97	96.73
	(13.13)	(12.45)	(14.22)	(11.87)	(13.66)

Note: N ~ 10,200. All means shown at the student level for the main analytic sample. Standard errors in parentheses. NCES provided weights are used throughout. See Appendix A for significance test *p*-values.

Table 3: Longitudinal growth model predicting third grade reading scores and achievement trajectories by preschool attendance

		Model A (1)	Model B (2)	Model C (3)	Model D (4)
Fixed Effects	Intercept	81.196***	114.117***	119.53***	117.87***
Final status,		(0.108)	(0.123)	(0.218)	(0.259)
π_{0i}	Head Start			-10.77***	-9.38***
				(0.439)	(0.448)
	State			-5.63***	-4.83***
				(0.417)	(0.414)
	Other			-6.17***	-5.36***
				(0.430)	(0.429)
	Black				-2.13***
					(0.390)
	Hispanic				-2.29***
					(0.328)
	Asian				7.51***
					(0.500)
	Ethnicity - Other				0.818
					(0.510)
	Female				2.75***
					(0.243)
Rate of change,					
π_{1i}	Intercept		1.417***	1.44***	1.44***
			(0.002)	(0.004)	(0.004)
	Head Start			-0.045***	-0.045***
				(0.009)	(0.009)
	State			-0.027***	-0.028***
				(0.009)	(0.009)
	Other			0.013	0.012
				(0.009)	(0.009)
Variance Components					
Level 1	Within-person	544.63	67.307	69.44	69.45
		(2.995)	(0.419)	(0.561)	(0.561)
Level 2	In final status	86.673	200.089	165.99	158.68
		(2.355)	(2.807)	(3.216)	(3.113)
	In rate of change		0.025	0.02	0.02
			(0.001)	(0.001)	(0.001)
	Covariance		1.10	0.803	0.802
			(0.045)	(0.053)	(0.053)
	Cov. Correlation		0.49	0.44	0.45
	ICC	0.137	0.75	0.71	0.70

Note: All models use survey weights. Control variables include race and gender in Models C and D, as shown. Standard errors in parentheses. *p <.05. **p < .01. ***p <.001.

Table 4: Longitudinal growth model predicting third grade math scores and achievement trajectories by preschool attendance

		Model A (1)	Model B (2)	Model C (3)	Model D (4)
Fixed Effects	Intercept	64.21***	99.83***	105.76***	105.85***
Final status,		(0.116)	(0.134)	(0.236)	(0.271)
π_{0i}	Head Start			-12.33***	-10.33***
				(0.474)	(0.477)
	State			-6.05***	-5.08***
				(0.452)	(0.445)
	Other			-6.96***	-5.75***
				(0.466)	(0.461)
	Black				-4.62***
					(0.376)
	Hispanic				-3.47***
					(0.316)
	Asian				5.88***
					(0.480)
	Ethnicity - Other				-0.361
					(0.493)
	Female				0.274
					(0.234)
Rate of change,					
π_{1i}	Intercept		1.53***	1.56***	1.56***
			(0.002)	(0.004)	(0.004)
	Head Start			-0.058***	-0.058***
				(0.008)	(0.008)
	State			-0.02**	-0.021**
				(0.008)	(0.008)
	Other			-0.007	-0.008
				(0.008)	(0.008)
Variance Components					
Level 1	Within-person	608.03	51.48	52.33	52.34
		(3.350)	(0.319)	(0.423)	(0.423)
Level 2	In final status	102.31	262.69	219.66	207.41
		(2.690)	(3.350)	(3.810)	(3.660)
	In rate of change		0.026	0.02	0.021
			(0.001)	(0.001)	(0.001)
	Covariance		2.07	1.68	1.60
			(0.048)	(0.056)	(0.055)
	Cov. Correlation		0.79	0.80	0.77
	ICC	0.14	0.84	0.81	0.80

Note: All models use survey weights. Control variables include race and gender in Models C and D, as shown. Standard errors in parentheses.* p <.05. **p < .01. ***p <.001.

Table 5: Longitudinal growth model predicting third grade reading and math scores and achievement trajectories by preschool attendance and teacher instructional differentiation

		Reading		Math	
		(1)	(2)	(3)	(4)
Fixed Effects, Final Status	Intercept	114.78***	113.26***	104.78***	105.25***
		(0.261)	(0.296)	(0.251)	(0.285)
π_{0i}	Ability Grouping	4.87***	4.88***	1.14***	1.20***
		(0.153)	(0.152)	(0.123)	(0.123)
	Head Start	-10.77***	-9.35***	-12.05***	-9.69***
		(0.460)	(0.469)	(0.490)	(0.493)
	State	-5.61***	-4.82***	-6.02***	-4.93***
		(0.436)	(0.433)	(0.467)	(0.459)
	Other	-6.26***	-5.43***	-6.83***	-5.41***
		(0.447)	(0.446)	(0.478)	(0.472)
	Black		-2.44***		-5.90***
			(0.388)		(0.390)
	Hispanic		-2.44***		-4.08***
			(0.327)		(0.330)
	Asian		6.86***		5.93***
			(0.507)		(0.510)
	Ethnicity - Other		0.76		-0.426
			(0.509)		(0.512)
	Female		2.74***		-0.163
			(0.241)		(0.243)
Rate of change,					
π_{1i}	Intercept	1.41***	1.41***	1.55***	1.55***
		(0.005)	(0.005)	(0.005)	(0.005)
	Head Start	-0.043***	-0.043***	-0.04***	-0.04***
		(0.011)	(0.011)	(0.010)	(0.010)
	State	-0.019*	-0.021*	-0.022*	-0.023*
		(0.010)	(0.010)	(0.010)	(0.010)
	Other	0.019*	0.018	0.003	0.002
		(0.010)	(0.010)	(0.009)	(0.010)
Variance Components					
Level 1	Within-person	87.72	87.72	66.75	66.99
		(0.845)	(0.847)	(0.646)	(0.725)
Level 2	In final status	151.63	144.68	201.92	188.584
		(3.284)	(3.177)	(3.842)	(2.688)
	In rate of change	0.004	0.004	0.011	0.011
		(0.001)	(0.001)	(0.001)	(0.001)
	Covariance	0.786	0.785	1.50	1.41
		(0.053)	(0.052)	(0.059)	(0.035)
	Cov. Correlation	0.99	0.99	0.99	0.98
	ICC	0.63	0.62	0.75	0.74

Note: All models use survey weights. Control variables include race and gender in columns 2 and 4, as shown. Standard errors in parentheses. * p <.05. ** p < .01. *** p <.001.

Table 6: Longitudinal growth model predicting third grade reading scores and achievement trajectories by preschool attendance and observed student behavior

		Reading		Math	
		(1)	(2)	(3)	(4)
Fixed Effects,	Intercept	108.877***	108.162***	97.944***	98.438***
Final Status		(0.362)	(0.378)	(0.348)	(0.363)
	Approaches to Learning	3.309***	3.186***	2.427***	2.432***
π_{0i}		(0.092)	(0.092)	(0.081)	(0.081)
	Head Start	-9.547***	-8.394***	-11.446***	-9.515***
		(0.429)	(0.440)	(0.468)	(0.470)
	State	-5.106***	-4.446***	-5.776***	-4.858***
		(0.406)	(0.406)	(0.446)	(0.438)
	Other	-5.721***	-5.000***	-6.668***	-5.462***
		(0.417)	(0.419)	(0.457)	(0.452)
	Black		-1.744***		-4.656***
			(0.375)		(0.363)
	Hispanic		-2.253***		-3.444***
			(0.316)		(0.307)
	Asian		6.506***		5.326***
			(0.491)		(0.475)
	Other		0.876		-0.361
			(0.492)		(0.478)
	Female		1.818***		-0.430
			(0.235)		(0.228)
Rate of change,					
π_{1i}	Intercept	1.434***	1.435***	1.553***	1.553***
		(0.005)	(0.005)	(0.004)	(0.004)
	Head Start	-0.035**	-0.035***	-0.049***	-0.050***
		(0.009)	(0.009)	(0.009)	(0.009)
	State	-0.021*	-0.022*	-0.018*	-0.019*
		(0.009)	(0.009)	(0.008)	(0.008)
	Other	0.013	0.012	-0.007	-0.008
		(0.009)	(0.009)	(0.008)	(0.008)
Variance Components					
Level 1	Within-person	69.393	69.328	52.290	52.306
		(0.590)	(0.589)	(0.444)	(0.445)
Level 2	In final status	143.692	139.968	200.202	188.625
		(3.022)	(2.959)	(3.654)	(3.513)
	In rate of change	0.019	0.019	0.020	0.020

	(0.001)	(0.001)	(0.001)	(0.001)
Covariance	0.705	0.717	1.595	1.518
	(0.053)	(0.053)	(0.056)	(0.055)
Cov. Correlation Coefficient	0.427	0.44	0.799	0.783
ICC	0.67	0.67	0.792	0.783

Note: All models use survey weights. Control variables include race and gender in Columns 2 and 4, as shown. Standard errors in parentheses. *p <.05. **p < .01. ***p <.001.

Figures

Figure 3: Predicted reading scores based on conditional growth trajectories

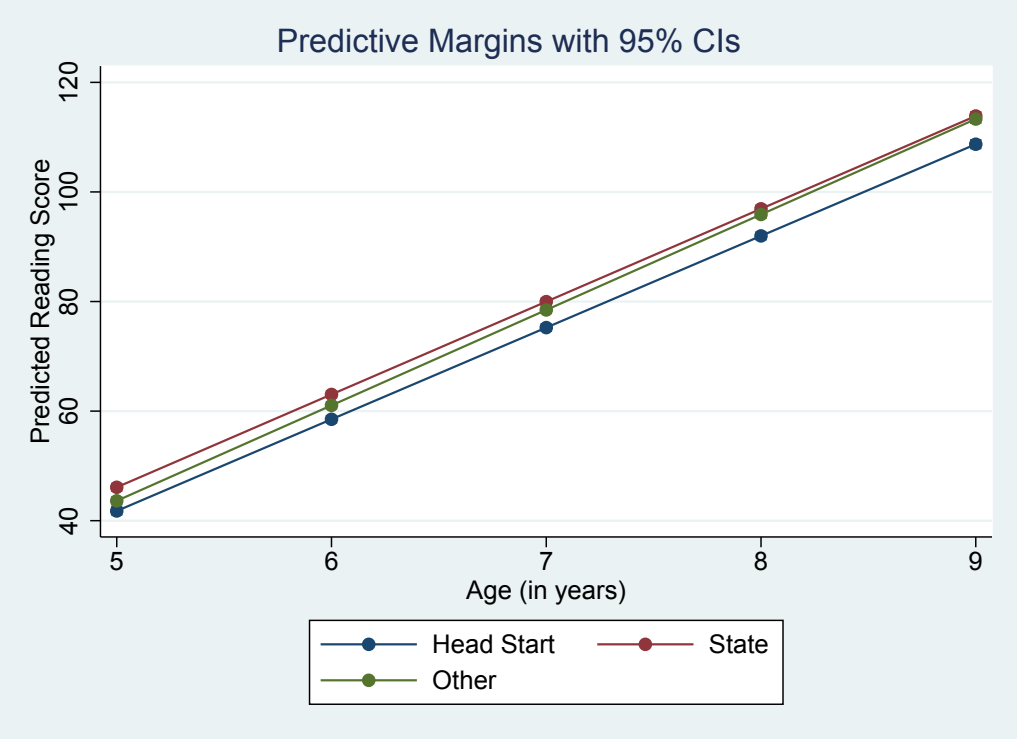


Figure 4: Predicted math scores based on conditional growth trajectories

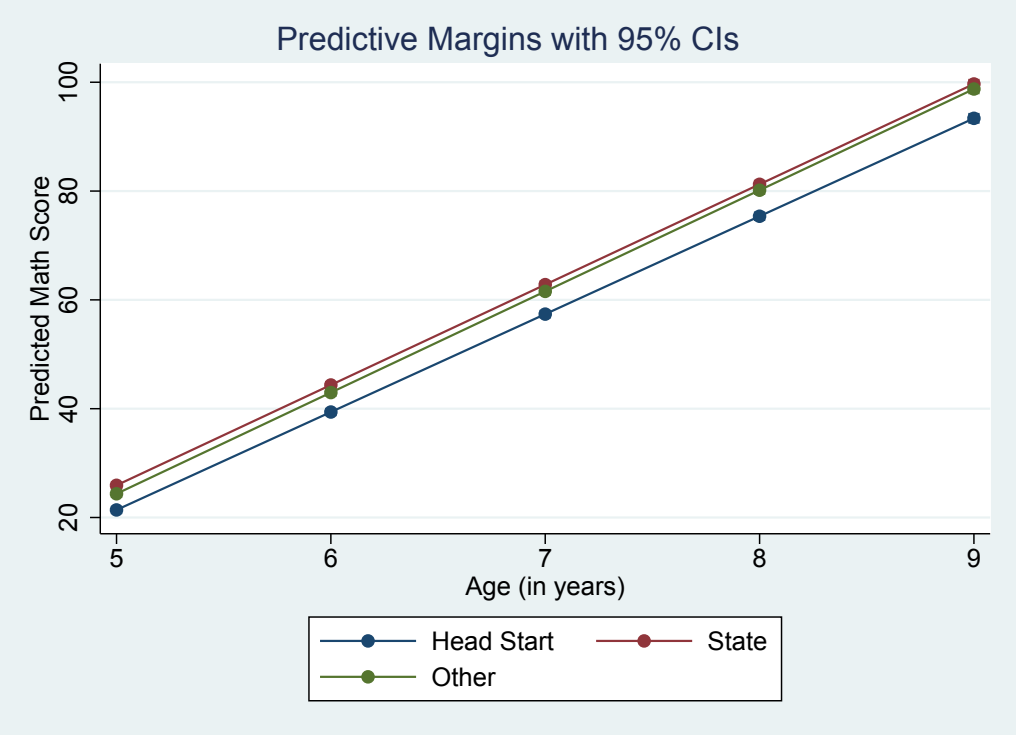


Figure 5: Predicted reading scores based on conditional growth trajectories, controlling for teacher use of reading groups

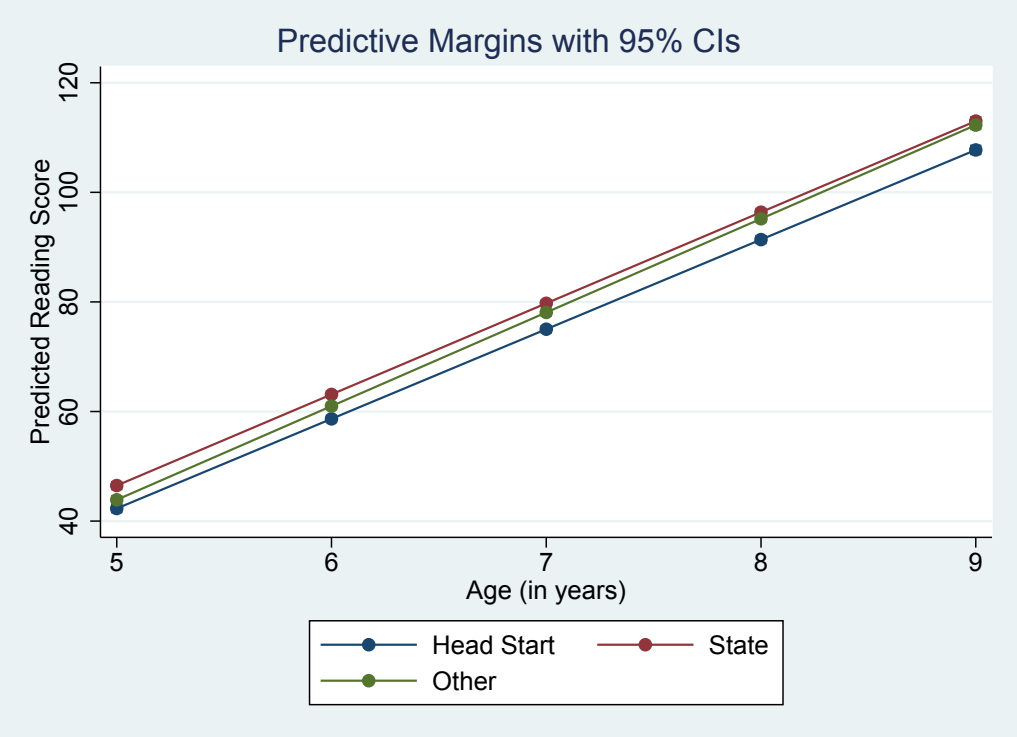


Figure 6: Predicted math scores based on conditional growth trajectories, controlling for teacher use of math groups

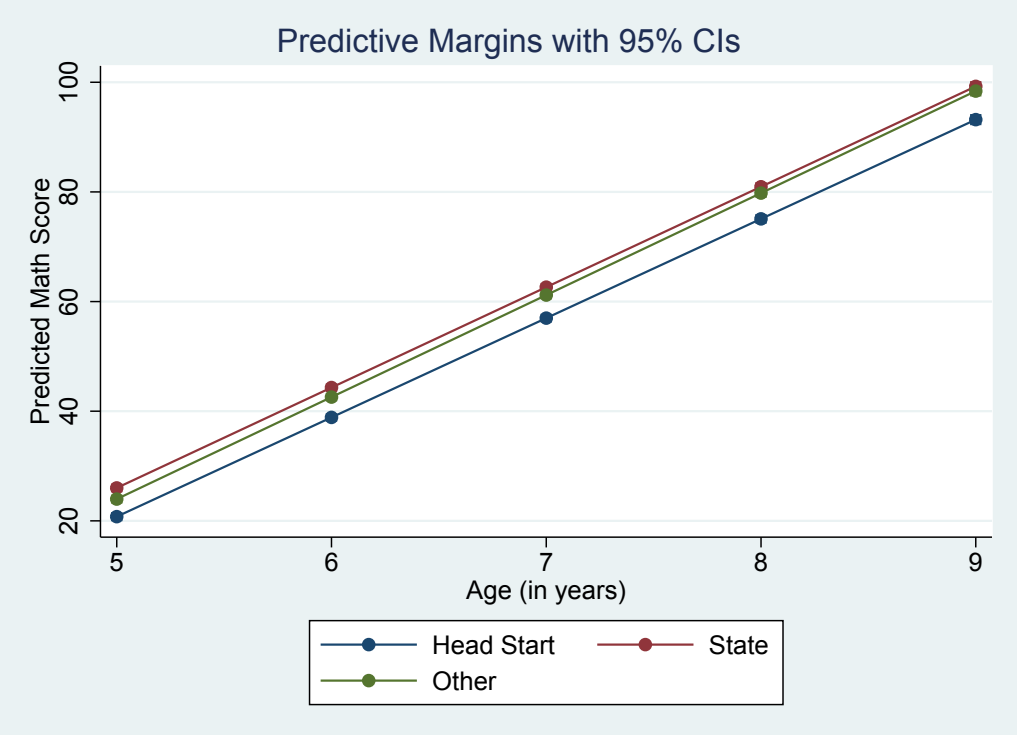


Figure 7: Predicted reading scores based on conditional growth trajectories, controlling for student approaches to learning scores

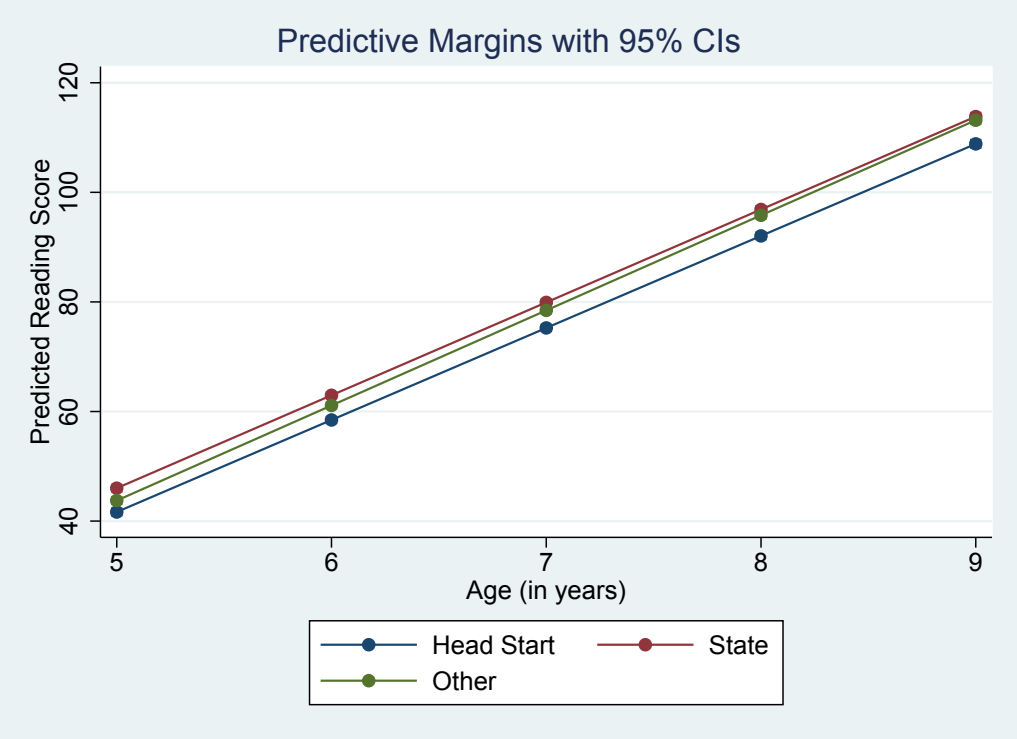
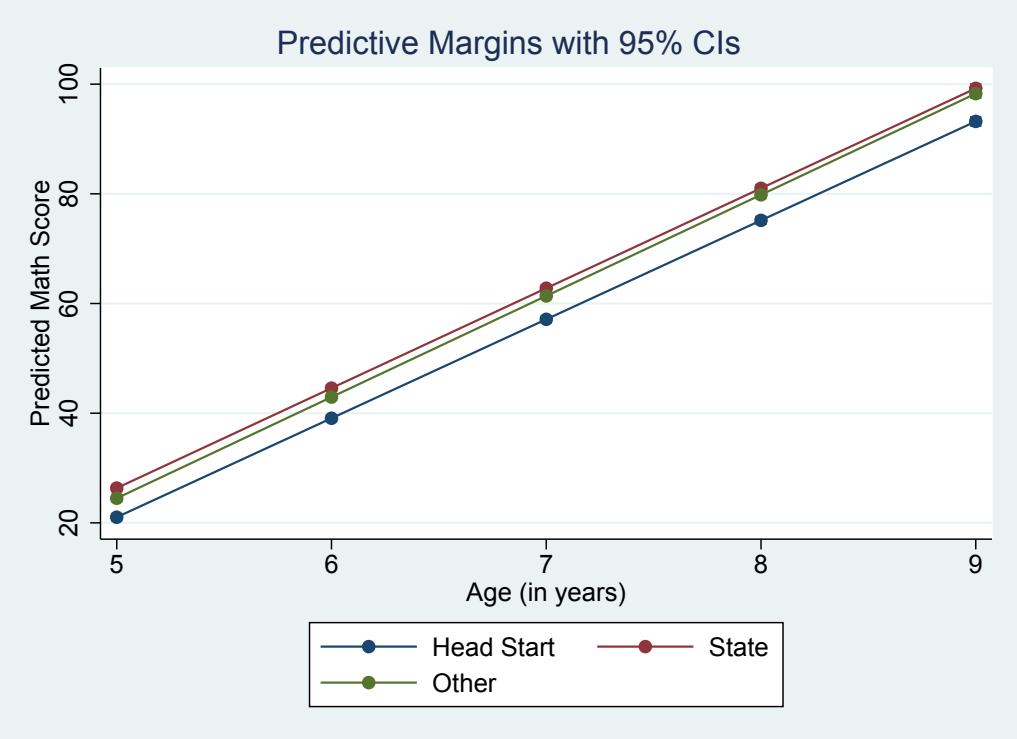


Figure 8: Predicted math scores based on conditional growth trajectories, controlling for student approaches to learning scores



Appendix A: Supplemental Analyses

Table A1: T-test *p*-values testing differences between sample groups

	HS = Center (1)	State = Center (2)	Other = Center (3)	HS = State (4)	HS = Other (5)	State = Other (6)	Excluded vs. Analytic Sample (7)
Female	0.036	0.031	0.722	0.870	0.039	0.036	0.505
White	0.000	0.000	0.000	0.000	0.000	0.000	0.001
Black	0.000	0.000	0.000	0.000	0.000	0.002	0.558
Hispanic	0.000	0.000	0.000	0.000	0.346	0.003	0.000
Asian	0.000	0.000	0.000	0.035	0.002	0.146	0.656
Ethnicity - Other	0.717	0.348	0.526	0.687	0.462	0.238	0.717
Reading Groups							
Kindergarten	0.974	0.345	0.011	0.479	0.068	0.502	0.568
First Grade	0.694	0.656	0.494	0.962	0.826	0.843	0.313
Second Grade	0.415	0.281	0.000	0.759	0.085	0.351	0.608
Third Grade	0.044	0.937	0.153	0.121	0.379	0.420	0.076
Math Groups	0.974	0.345	0.011	0.479	0.068	0.502	0.322
Kindergarten	0.298	0.220	0.058	0.059	0.817	0.003	0.786
First Grade	0.167	0.015	0.449	0.763	0.118	0.005	0.965
Second Grade	0.740	0.687	0.510	0.447	0.261	0.969	0.192
Third Grade	0.309	0.627	0.511	0.622	0.544	0.940	0.322

Note: N ~ 10,200. T-tests of means where Ho: estimates are equivalent. Survey weights used throughout.

Table A2: T-test *p*-values testing outcome differences between sample groups

	HS = Center (1)	State = Center (2)	Other = Center (3)	HS = State (4)	HS = Other (5)	State = Other (6)	Excluded vs. Analytic Sample (7)
Reading							
Fall Kindergarten	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Spring Kindergarten	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fall First Grade	0.000	0.000	0.000	0.001	0.000	0.195	0.000
Spring First Grade	0.000	0.000	0.000	0.000	0.000	0.003	0.000
Fall Second Grade	0.000	0.000	0.000	0.005	0.000	0.042	0.000
Spring Second Grade	0.000	0.000	0.000	0.000	0.000	0.133	0.000
Spring Third Grade	0.000	0.000	0.000	0.000	0.000	0.203	0.000
Math							
Fall Kindergarten	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Spring Kindergarten	0.000	0.000	0.000	0.000	0.000	0.001	0.000
Fall First Grade	0.000	0.000	0.000	0.010	0.000	0.106	0.000
Spring First Grade	0.000	0.000	0.000	0.000	0.000	0.005	0.000
Fall Second Grade	0.000	0.000	0.000	0.004	0.000	0.201	0.000
Spring Second Grade	0.000	0.000	0.000	0.000	0.000	0.302	0.000
Spring Third Grade	0.000	0.000	0.000	0.000	0.000	0.014	0.000

Note: N ~ 10,200. T-tests of means where Ho: estimates are equivalent. Survey weights used throughout.

Table A3: Mediation tests for *approaches to learning*

	IV (Preschool) on Mediator (Attention)				IV (Preschool) on DV (Reading and Math)								Mediator (Attention) on DV (Reading and Math)	
	Head Start	State	Other	Center	Head Start		State		Other		Center			
					Read	Math	Read	Math	Read	Math	Read	Math	Read	Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Coefficient	-0.225	-0.047	-0.056	0.179	-7.95	-9.04	-1.73	-1.92	-3.27	-3.33	7.13	7.87	9.60	9.68
s.e.	(0.009)	(0.008)	(0.009)	(0.006)	(0.309)	(0.327)	(0.295)	(0.312)	(0.304)	(0.322)	(0.226)	(0.239)	(0.166)	(0.176)
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: Coefficients and standard errors are point estimates from bivariate regressions.

Appendix B: Robustness Checks
Centering at Kindergarten

Table B1: Association between reading and math outcomes and preschool attendance

		Reading		Math	
		(1)	(2)	(3)	(4)
Fixed Effects	Intercept	50.09***	48.50***	30.66***	30.85***
Final status,		(0.195)	(0.243)	(0.172)	(0.222)
π_{0i}	Head Start	-8.26***	-6.85***	-9.21***	-7.19***
		(0.395)	(0.405)	(0.347)	(0.360)
	State	-3.97***	-3.16***	-4.73***	-3.76***
		(0.380)	(0.377)	(0.334)	(0.332)
	Other	-6.36***	-5.49***	-6.17***	-4.92***
		(0.415)	(0.414)	(0.365)	(0.366)
	Black		-2.27***		-4.76***
			(0.391)		(0.377)
	Hispanic		-2.44***		-3.62***
			(0.328)		(0.317)
	Asian		7.63***		5.99***
			(0.500)		(0.482)
	Other		0.702		-0.463
			(0.512)		(0.495)
	Female		2.72***		0.241
			(0.243)		(0.235)
Rate of change,					
π_{1i}	Intercept	1.44***	1.44***	1.56***	1.56***
		(0.004)	(0.004)	(0.004)	(0.004)
	Head Start	-0.046***	-0.046***	-0.057***	-0.057***
		(0.009)	(0.009)	(0.008)	(0.008)
	State	-0.028***	-0.028***	-0.019**	-0.02**
		(0.009)	(0.009)	(0.007)	(0.008)
	Other	0.012	0.011	-0.006	-0.006
		(0.009)	(0.009)	(0.008)	(0.008)
Variance Components					
Level 1	Within-person	69.44	69.44	52.33	52.34
		(0.561)	(0.561)	(0.423)	(0.423)
Level 2	In final status	137.25	129.87	107.09	101.92
		(2.812)	(2.717)	(2.174)	(2.110)
	In rate of change	0.02	0.02	0.02	0.021
		(0.001)	(0.001)	(0.001)	(0.001)
	Covariance	-0.181	-0.184	0.696	0.614
		(0.049)	(0.048)	(0.036)	(0.036)

Note: Age centered at Kindergarten entry. All models use survey weights. Control variables include race and gender in Columns 2 and 4, as shown. Standard errors in parentheses. *p <.05. **p < .01. ***p <.001.

Table B2: Association between reading and math outcomes, preschool attendance, and ability grouping

		Reading		Math	
		(1)	(2)	(3)	(4)
Fixed Effects	Intercept	46.93***	45.47***	30.02***	30.62***
Final status,		(0.218)	(0.261)	(0.187)	(0.237)
π_{0i}	Head Start	-8.37***	-6.90***	-9.73***	-7.37***
		(0.413)	(0.422)	(0.382)	(0.392)
	State	-4.31***	-3.49***	-4.56***	-3.47***
		(0.398)	(0.394)	(0.368)	(0.365)
	Other	-6.78***	-5.86***	-6.37***	-4.89***
		(0.430)	(0.429)	(0.400)	(0.400)
	Ability Group	4.87***	4.88***	1.14***	1.20***
		(0.153)	(0.152)	(0.124)	(0.123)
	Black		-2.58***		-6.10***
			(0.389)		(0.390)
	Hispanic		-2.60***		-4.25***
			(0.328)		(0.331)
	Asian		6.98***		6.04***
			(0.509)		(0.511)
	Other		0.642		-0.535
			(0.511)		(0.514)
	Female		2.71***		-0.195
			(0.242)		(0.244)
Rate of change,					
π_{1i}	Intercept	1.41***	1.41***	1.55***	1.55***
		(0.005)	(0.005)	(0.005)	(0.005)
	Head Start	-0.044***	-0.044***	-0.041***	-0.041***
		(0.011)	(0.011)	(0.010)	(0.010)
	State	-0.207*	-0.217*	-0.023**	-0.024**
		(0.009)	(0.009)	(0.009)	(0.009)
	Other	0.018	0.016	0.0003	-0.001
		(0.011)	(0.011)	(0.010)	(0.010)
Variance Components					
Level 1	Within-person	87.76	87.76	66.8	67.04
		(0.848)	(0.848)	(0.647)	(0.650)
Level 2	In final status	86.83	80.17	85.24	78.41
		(2.580)	(2.470)	(2.270)	(2.170)
	In rate of change	0.004	0.004	0.011	0.01
		(0.001)	(0.001)	(0.001)	(0.001)
	Covariance	0.589	0.579	0.966	0.904
		(0.030)	(0.029)	(0.026)	(0.025)

Note: Age centered at Kindergarten entry. All models use survey weights. Control variables include race and gender in Columns 2 and 4, as shown. Standard errors in parentheses. *p <.05. **p < .01. ***p <.001.

Table B3: Association between reading and math outcomes, preschool attendance, and attention skills

		Reading		Math	
		(1)	(2)	(3)	(4)
Fixed Effects	Intercept	38.67***	38.09***	20.82***	21.62***
Final status,		(0.495)	(0.506)	(0.420)	(0.430)
π_{0i}	Head Start	-7.46***	-6.27***	-8.54***	-6.65***
		(0.389)	(0.399)	(0.338)	(0.350)
	State	-3.73***	-3.04***	-4.49***	-3.59***
		(0.373)	(0.371)	(0.325)	(0.324)
	Other	-5.81***	-5.02***	-5.73***	-4.51***
		(0.405)	(0.406)	(0.353)	(0.355)
	Attention	3.64***	3.48***	3.15***	3.09***
		(0.147)	(0.147)	(0.123)	(0.124)
	Black		-1.86***		-4.67***
			(0.376)		(0.362)
	Hispanic		-2.39***		-3.56***
			(0.316)		(0.306)
	Asian		6.62***		5.37***
			(0.492)		(0.474)
	Other		0.768		-0.445
			(0.493)		(0.476)
	Female		1.77***		-0.544*
			(0.236)		(0.228)
Rate of change,					
π_{1i}	Intercept	1.48***	1.47***	1.65***	1.64***
		(0.015)	(0.015)	(0.013)	(0.013)
	Head Start	-0.039***	-0.039***	-0.058***	-0.058***
		(0.009)	(0.009)	(0.009)	(0.009)
	State	-0.024**	-0.025**	-0.022**	-0.022**
		(0.009)	(0.009)	(0.008)	(0.008)
	Other	0.006	0.005	-0.012	-0.013
		(0.009)	(0.009)	(0.009)	(0.009)
Variance Components					
Level 1	Within-person	69.3	69.24	52.17	52.19
		(0.590)	(0.590)	(0.443)	(0.444)
Level 2	In final status	119.96	115.21	91.7	87.59
		(2.690)	(2.620)	(2.060)	(2.000)
	In rate of change	0.019	0.019	0.021	0.021
		(0.001)	(0.001)	(0.001)	(0.001)
	Covariance	-0.196	-0.191	0.683	0.595
		(0.049)	(0.049)	(0.036)	(0.036)

Note: Age centered at Kindergarten entry. All models use survey weights. Control variables include race and gender in Columns 2 and 4, as shown. Standard errors in parentheses. *p <.05. **p < .01. ***p <.001.

Robustness Check: Omit 'Other'

Table B4: Association between reading and math outcomes and preschool attendance

		Reading		Math	
		(1)	(2)	(3)	(4)
Fixed Effects	Intercept	113.36***	112.50***	98.79***	100.11***
Final status,		(0.370)	(0.409)	(0.401)	(0.432)
π_{0i}	Head Start	-4.60***	-4.02***	-5.37***	-4.59***
		(0.531)	(0.525)	(0.574)	(0.563)
	State	0.54	0.53	0.92	0.67
		(0.513)	(0.505)	(0.556)	(0.544)
	Center	6.17***	5.36***	6.97***	5.75***
		(0.430)	(0.429)	(0.465)	(0.460)
	Black		-2.13***		-4.62***
			(0.390)		(0.376)
	Hispanic		-2.29***		-3.47***
			(0.328)		(0.316)
	Asian		7.52***		5.88***
			(0.500)		(0.480)
	Other		0.817		-0.352
			(0.511)		(0.494)
	Female		2.76***		0.274
			(0.243)		(0.234)
Rate of change,					
π_{1i}	Intercept	1.45***	1.45***	1.55***	1.55***
		(0.008)	(0.008)	(0.007)	(0.007)
	Head Start	-0.058***	-0.057***	-0.051***	-0.05***
		(0.011)	(0.011)	(0.010)	(0.010)
	State	-0.04***	-0.039***	-0.013	-0.013
		(0.011)	(0.011)	(0.009)	(0.009)
	Center	-0.013	-0.012	0.007	0.008
		(0.009)	(0.009)	(0.008)	(0.008)
Variance Components					
Level 1	Within-person	69.45	69.45	52.34	52.34
		(0.561)	(0.561)	(0.423)	(0.423)
Level 2	In final status	166.01	158.7	219.69	207.43
		(3.220)	(3.110)	(3.807)	(3.660)
	In rate of change	0.02	0.02	0.02	0.021
		(0.001)	(0.001)	(0.001)	(0.001)
	Covariance	0.803	0.802	1.68	1.6
		(0.054)	(0.053)	(0.056)	(0.055)

Note: All models use survey weights. Control variables include race and gender in Columns 2 and 4, as shown. Standard errors in parentheses. *p <.05. **p < .01. ***p <.001.

Table B5: Association between reading and math outcomes, preschool attendance, and ability grouping

		Reading		Math
		(1)	(2)	(3)
Fixed Effects	Intercept	108.52***	107.82***	97.95***
Final status,		(0.408)	(0.444)	(0.417)
π_{0i}	Head Start	-4.52***	-3.92***	-5.22***
		(0.555)	(0.550)	(0.593)
	State	0.652	0.609	0.819
		(0.535)	(0.528)	(0.573)
	Center	6.26***	5.44***	6.84***
		(0.447)	(0.446)	(0.478)
	Ability Group	4.87***	4.88***	1.14***
		(0.152)	(0.152)	(0.123)
	Black		-2.44***	
			(0.388)	
	Hispanic		-2.44***	
			(0.327)	
	Asian		6.86***	
			(0.507)	
	Other		0.768	
			(0.510)	
	Female		2.74***	
			(0.241)	
Rate of change,				
π_{1i}	Intercept	1.43***	1.42***	1.55***
		(0.009)	(0.009)	(0.008)
	Head Start	-0.062***	-0.061***	-0.043***
		(0.013)	(0.013)	(0.012)
	State	-0.039***	-0.039**	-0.025*
		(0.012)	(0.012)	(0.011)
	Center	-0.02*	-0.018	-0.003
		(0.010)	(0.010)	(0.009)
Variance Components				
Level 1	Within-person	87.74	87.74	66.76
		(0.848)	(0.847)	(0.721)
Level 2	In final status	151.64	144.68	201.95
		(3.280)	(3.180)	(2.840)
	In rate of change	0.004	0.004	0.011
		(0.000)	(0.001)	(0.001)
	Covariance	0.786	0.784	1.49
		(0.053)	(0.052)	(0.038)

Note: All models use survey weights. Control variables include race and gender in Column 2, as shown. Standard errors in parentheses. *p <.05. **p < .01. ***p <.001.

Table B6: Association between reading and math outcomes, preschool attendance, and attention skills

		Reading		Math	
		(1)	(2)	(3)	(4)
Fixed Effects	Intercept	104.09***	104.01***	93.65***	95.13***
Final status,		(0.564)	(0.586)	(0.560)	(0.580)
π_{0i}	Head Start	-3.88***	-3.44***	-4.90***	-4.19***
		(0.519)	(0.516)	(0.570)	(0.559)
	State	0.618	0.559	0.897	0.612
		(0.500)	(0.297)	(0.552)	(0.539)
	Center	5.77***	5.05***	6.79***	5.59***
		(0.419)	(0.421)	(0.461)	(0.456)
	Attention	3.00***	2.91***	1.65***	1.73***
		(0.142)	(0.143)	(0.130)	(0.130)
	Black		-1.73***		-4.53***
			(0.375)		(0.361)
	Hispanic		-2.24***		-3.40***
			(0.316)		(0.305)
	Asian		6.52***		5.26***
			(0.491)		(0.473)
	Other		0.881		-0.333
			(0.492)		(0.475)
	Female		1.80***		-0.511*
			(0.235)		(0.227)
Rate of change,					
π_{1i}	Intercept	1.48***	1.48***	1.64***	1.63***
		(0.016)	(0.016)	(0.014)	(0.014)
	Head Start	-0.049***	-0.048***	-0.047***	-0.046***
		(0.011)	(0.011)	(0.010)	(0.010)
	State	-0.033**	-0.033**	-0.011	-0.01
		(0.011)	(0.011)	(0.010)	(0.010)
	Center	-0.011	-0.009	0.012	0.012
		(0.009)	(0.009)	(0.008)	(0.008)
Variance Components					
Level 1	Within-person	69.31	69.25	52.17	52.2
		(0.590)	(0.589)	(0.443)	(0.444)
Level 2	In final status	145.19	141.2	204.51	192.38
		(3.100)	(3.026)	(3.770)	(3.621)
	In rate of change	0.019	0.019	0.021	0.021
		(0.001)	(0.001)	(0.012)	(0.001)
	Covariance	0.741	0.747	1.69	1.6
		(0.055)	(0.055)	(0.059)	(0.058)

Note: All models use survey weights. Control variables include race and gender in Columns 2 and 4, as shown. Standard errors in parentheses. *p <.05. **p < .01. ***p <.001.

Robustness Check: Fixed Effects

Table B7: Association between reading achievement and preschool attendance, using school fixed effects

	Fall K	Spring K	Fall 1	Spring 1	Fall 2	Spring 2	Spring 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Head Start	-4.246***	-5.575***	-5.200***	-6.696***	-5.267***	-5.118***	-5.264***
	(0.358)	(0.433)	(1.037)	(0.570)	(0.999)	(0.499)	(0.483)
State Preschool	-1.917***	-2.500***	-3.197***	-2.923***	-3.980***	-2.299***	-2.371***
	(0.332)	(0.398)	(0.937)	(0.511)	(0.880)	(0.444)	(0.429)
Other	-3.728***	-3.633***	-2.662**	-4.026***	-3.104***	-2.748***	-2.233***
	(0.327)	(0.395)	(0.918)	(0.507)	(0.870)	(0.441)	(0.424)
Black	-1.421**	-2.544***	-3.821**	-3.303***	-2.550	-2.252**	-3.221***
	(0.460)	(0.562)	(1.352)	(0.772)	(1.326)	(0.689)	(0.676)
Hispanic	-2.427***	-2.821***	-4.035***	-3.454***	-3.182**	-2.399***	-2.728***
	(0.371)	(0.445)	(1.005)	(0.584)	(0.967)	(0.511)	(0.497)
Asian	4.349***	4.725***	6.574***	4.243***	4.143**	1.664*	1.684*
	(0.518)	(0.620)	(1.360)	(0.804)	(1.311)	(0.697)	(0.683)
Ethnicity- Other	0.964*	1.053	0.863	1.591*	0.848	1.028	1.511*
	(0.467)	(0.567)	(1.306)	(0.747)	(1.269)	(0.666)	(0.647)
Female	0.820***	1.515***	1.370*	2.562***	2.212***	2.025***	1.870***
	(0.214)	(0.256)	(0.594)	(0.331)	(0.562)	(0.287)	(0.277)
Observations	9880	9780	2880	8630	2630	7940	7450
R-squared	0.043	0.045	0.041	0.046	0.042	0.035	0.045

Note: School fixed effects model with race/ethnicity and gender covariates. Survey weights and clustered standard errors used throughout. Observations rounded to nearest 10 in compliance with NCES restricted data use agreement. p <.05. **p < .01. ***p <.001.

Table B8: Association between math achievement and preschool participation, using school fixed effects

	Fall K	Spring K	Fall 1	Spring 1	Fall 2	Spring 2	Spring 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Head Start	-4.756***	-5.102***	-5.221***	-6.077***	-5.332***	-6.010***	-5.885***
	(0.35)	(0.39)	(0.99)	(0.59)	(1.07)	(0.56)	(0.54)
State Preschool	-1.978***	-2.372***	-4.415***	-3.023***	-5.164***	-2.587***	-2.370***
	(0.33)	(0.36)	(0.89)	(0.52)	(0.94)	(0.50)	(0.48)
Other	-3.202***	-3.052***	-3.242***	-3.562***	-3.203***	-2.966***	-2.986***
	(0.33)	(0.36)	(0.87)	(0.52)	(0.93)	(0.50)	(0.48)
Black	-3.092***	-4.687***	-8.041***	-8.825***	-7.119***	-8.028***	-7.992***
	(0.46)	(0.51)	(1.29)	(0.79)	(1.42)	(0.78)	(0.76)
Hispanic	-3.205***	-3.462***	-6.516***	-5.578***	-5.859***	-4.424***	-4.398***
	(0.37)	(0.40)	(0.96)	(0.60)	(1.03)	(0.58)	(0.56)
Asian	2.480***	1.857***	0.491	2.156**	1.881	2.650***	2.989***
	(0.52)	(0.56)	(1.30)	(0.82)	(1.40)	(0.78)	(0.77)
Ethnicity- Other	0.155	-0.024	-2.394	-0.454	-1.088	0.188	0.035
	(0.46)	(0.51)	(1.24)	(0.77)	(1.36)	(0.75)	(0.73)
Female	-0.827***	-0.659**	-1.586**	-2.554***	-2.361***	-2.638***	-3.182***
	(0.21)	(0.23)	(0.57)	(0.34)	(0.60)	(0.32)	(0.31)
Observations	9860	9770	2880	8620	2630	7940	7450
R-squared	0.046	0.046	0.053	0.057	0.056	0.061	0.076

Note: School fixed effects model with race/ethnicity and gender covariates. Survey weights and clustered standard errors used throughout. Observations rounded to nearest 10 in compliance with NCES restricted data use agreement. p <.05. **p < .01. ***p <.001.

Table B9: Association between reading and math outcomes and preschool, instructional differentiation, using school fixed effects

	Reading				Math			
	Fall K	Spring 1	Spring 2	Spring 3	Fall K	Spring 1	Spring 2	Spring 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Head Start	-4.422***	-5.688***	-5.070***	-5.059***	-4.785***	-5.045***	-5.967***	-5.724***
	(0.372)	(0.581)	(0.518)	(0.500)	(0.37)	(0.60)	(0.58)	(0.57)
State Preschool	-2.014***	-2.603***	-2.175***	-2.248***	-1.931***	-2.334***	-2.762***	-2.426***
	(0.347)	(0.517)	(0.460)	(0.444)	(0.34)	(0.53)	(0.52)	(0.50)
Other	-3.742***	-3.564***	-2.596***	-1.989***	-3.130***	-2.940***	-2.868***	-2.899***
	(0.339)	(0.513)	(0.454)	(0.436)	(0.33)	(0.53)	(0.51)	(0.49)
K Instructional Group	1.051**				-0.143			
	(0.359)				(0.39)			
Black	-1.434**	-3.687***	-2.181**	-2.961***	-3.252***	-9.366***	-8.371***	-8.416***
	(0.477)	(0.772)	(0.713)	(0.701)	(0.47)	(0.79)	(0.80)	(0.80)
Hispanic	-2.563***	-3.701***	-2.375***	-2.785***	-3.072***	-5.524***	-4.608***	-4.400***
	(0.386)	(0.594)	(0.532)	(0.515)	(0.38)	(0.61)	(0.60)	(0.58)
Asian	4.306***	3.772***	1.521*	1.637*	2.566***	2.035*	2.442**	2.719***
	(0.550)	(0.820)	(0.742)	(0.715)	(0.54)	(0.84)	(0.83)	(0.81)
Ethnicity- Other	0.922	1.935*	1.060	1.609*	0.392	-0.227	0.485	0.297
	(0.485)	(0.757)	(0.690)	(0.668)	(0.48)	(0.78)	(0.77)	(0.75)
Female	0.876***	2.058***	1.901***	1.769***	-0.728***	-2.993***	-2.646***	-3.261***
	(0.221)	(0.333)	(0.297)	(0.286)	(0.22)	(0.34)	(0.33)	(0.32)
1 Instructional Group		0.669				-0.601		
		(0.674)				(0.47)		
2 Instructional Group			-0.724				-0.258	
			(0.698)				(0.47)	
3 Instructional Group				-0.486				-0.472
				(0.557)				(0.47)
Observations	9190	7690	7300	6850	9180	7670	7270	6810
R-squared	0.045	0.041	0.033	0.042	0.045	0.060	0.063	0.077

Note: School fixed effects model with race/ethnicity and gender covariates. Survey weights and clustered standard errors used throughout. Observations rounded to nearest 10 in compliance with NCES restricted data use agreement. p <.05. **p < .01. ***p <.001.

Table B10: Association between reading achievement and preschool, approaches to learning skills, using school fixed effects

	Fall K	Spring K	Fall 1	Spring 1	Fall 2	Spring 2	Spring 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Head Start	-3.128***	-3.647***	-3.710***	-3.920***	-3.316***	-3.545***	-3.590***
	(0.351)	(0.408)	(0.973)	(0.516)	(0.909)	(0.476)	(0.459)
State Preschool	-1.639***	-2.059***	-2.273**	-1.952***	-2.417**	-1.545***	-1.481***
	(0.325)	(0.374)	(0.867)	(0.457)	(0.797)	(0.422)	(0.405)
Other	-3.075***	-2.819***	-1.681*	-2.743***	-2.243**	-1.801***	-1.331***
	(0.318)	(0.368)	(0.847)	(0.454)	(0.783)	(0.416)	(0.399)
Fall Approaches	5.506***						
	(0.167)						
Black	-0.757	-0.995	-1.059	-1.125	0.396	-0.718	-1.586*
	(0.448)	(0.526)	(1.249)	(0.686)	(1.208)	(0.654)	(0.642)
Hispanic	-2.270***	-2.443***	-3.346***	-3.634***	-2.001*	-2.352***	-2.759***
	(0.361)	(0.418)	(0.925)	(0.525)	(0.880)	(0.486)	(0.470)
Asian	3.415***	2.979***	5.238***	1.297	2.325	-0.282	-0.025
	(0.516)	(0.588)	(1.254)	(0.727)	(1.186)	(0.681)	(0.655)
Ethnicity- Other	1.224**	1.655**	1.665	1.755**	1.025	0.909	1.479*
	(0.452)	(0.528)	(1.195)	(0.670)	(1.144)	(0.630)	(0.609)
Female	-0.734***	-0.787**	-1.203*	-0.998**	-0.559	-0.552*	-0.466
	(0.213)	(0.246)	(0.562)	(0.304)	(0.523)	(0.281)	(0.271)
Spring Approaches		7.929***					
		(0.192)					
Fall 1 Approaches			9.933***				
			(0.438)				
Spring 1 Approaches				9.754***			
				(0.231)			
Fall 2 Approaches					9.458***		
					(0.409)		
Spring 2 Approaches						7.116***	
						(0.209)	
Spring 3 Approaches							6.474***
							(0.203)
Observations	9280	9200	2770	7730	2510	7340	6930

R-squared	0.153	0.208	0.213	0.246	0.243	0.191	0.196

Note: School fixed effects model with race/ethnicity and gender covariates. Survey weights and clustered standard errors used throughout. Observations rounded to nearest 10 in compliance with NCES restricted data use agreement. p <.05. **p < .01. ***p <.001.

Table B11: Association between math achievement and preschool, approaches to learning skills, using school fixed effects

	Fall K	Spring K	Fall 1	Spring 1	Fall 2	Spring 2	Spring 3
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Head Start	-3.304***	-3.211***	-3.649***	-3.335***	-3.029**	-4.098***	-3.879***
	(0.34)	(0.36)	(0.92)	(0.53)	(0.97)	(0.53)	(0.51)
State Preschool	-1.497***	-1.872***	-3.443***	-1.824***	-3.587***	-2.026***	-1.483**
	(0.31)	(0.33)	(0.82)	(0.47)	(0.85)	(0.47)	(0.45)
Other	-2.318***	-2.281***	-2.165**	-2.242***	-2.350**	-1.937***	-1.972***
	(0.30)	(0.32)	(0.80)	(0.47)	(0.84)	(0.46)	(0.45)
Fall K Approaches	6.581***						
	(0.16)						
Black	-2.454***	-3.285***	-5.679***	-6.862***	-4.220**	-6.624***	-6.352***
	(0.43)	(0.46)	(1.19)	(0.71)	(1.29)	(0.73)	(0.72)
Hispanic	-2.772***	-2.953***	-5.993***	-5.642***	-4.972***	-4.590***	-4.424***
	(0.35)	(0.37)	(0.88)	(0.54)	(0.94)	(0.54)	(0.53)
Asian	1.395**	0.088	-0.774	-0.332	0.323	0.316	0.963
	(0.50)	(0.52)	(1.19)	(0.75)	(1.27)	(0.76)	(0.74)
Ethnicity- Other	0.522	0.531	-1.721	-0.426	-0.916	-0.006	0.062
	(0.43)	(0.46)	(1.13)	(0.69)	(1.22)	(0.70)	(0.68)
Female	-2.619***	-2.888***	-4.085***	-5.958***	-5.234***	-5.551***	-5.973***
	(0.20)	(0.22)	(0.53)	(0.31)	(0.56)	(0.31)	(0.30)
Spring K Approaches		7.884***					
		(0.17)					
Fall 1 Approaches			9.388***				
			(0.42)				
Spring 1 Approaches				9.689***			
				(0.24)			
Fall 2 Approaches					9.885***		
					(0.44)		
Spring 2 Approaches						8.348***	
						(0.23)	
Spring 3 Approaches							7.559***
							(0.23)

Observations	9260	9190	2780	7720	2520	7340	6930
R-squared	0.205	0.245	0.222	0.250	0.248	0.229	0.233

Note: School fixed effects model with race/ethnicity and gender covariates. Survey weights and clustered standard errors used throughout. Observations rounded to nearest 10 in compliance with NCES restricted data use agreement. p <.05. **p < .01. ***p <.001.

CHAPTER 3

SOCIAL IMPACT BONDS FOR PUBLIC PRESCHOOL? ISSUES IN CURRENT PRESCHOOL DELIVERY, GOALS, AND FINANCING

Introduction

Despite substantial evidence that public preschool is an effective intervention for reducing the gap in school readiness, there are significant geographic, income, and racial disparities in access to public preschool (Bailey et al., 2017; Bassok, Gibbs, & Latham, 2018; Phillips et al., 2017; Weiland, 2016). While the federally funded Head Start program is implemented nationwide, state and local resources for public preschool programs vary significantly. In 2017, seven states did not offer any state preschool, and the remaining states, on average, enrolled only one-third of all 4-year-olds, and just 5% of 3-year-olds (Friedman-Krauss, et al., 2018). Program capacity is arguably the largest barrier to enrollment in public programs (Chaudry et al., 2017). State budgets for public pre-kindergarten grew steadily in the early 2000s, but plateaued and then began to decline in 2010 in response to the Great Recession (Haskins, 2017). Government budgets globally were impacted during this time, and austerity measures led to program reductions and a desire to reform public sector spending (Dodd & Moody, 2011; Fraser et al., 2018).

In this climate, an innovative alternative source of funding known as Social Impact Bonds emerged. By design, an intervention financed by a Social Impact Bond is paid for by a private investor, and government repayment only occurs if the intervention "works," or is able to generate future savings. The first Social Impact Bond (SIB) was launched in the United Kingdom in 2010; the idea materialized at least partly as an attempt to leverage public-private partnerships in response to the financial downturn and was additionally driven by the lingering

models of New Public Management reforms of the previous decades (Fraser et al., 2018; Heinrich & Kabourek, 2018; Warner, 2013). Pre-kindergarten has been used as a social intervention in the United States for decades. Rigorous cost-benefit analyses have been conducted on early, small-scale programs that indicate social investment returns ranging from $3 to $17 for every dollar invested in high-quality programs (i.e., Heckman et al., 2010; Karoly, 2016). The precedent set by these studies gives local governments and investors a predicted return on investment and provides language for thinking about preschool as a "monetizable" intervention. Therefore, governments hoping to expand public preschool programs are able to add Social Impact Bonds to their toolbox for potential funding mechanisms.

Financing Public Preschool

The landscape of preschool finance and delivery is complex due to the uncoordinated nature of funding streams and provision (Hustedt & Barnett, 2011). Federal, state, and local funding streams each have their own requirements and regulations. Providers often "braid," or blend, funding streams to create a single classroom (Chaudry & Datta, 2017; Hustedt & Barnett, 2011). In some ways, this makes preschool an appropriate venue for SIB-funding—government agencies already operate in a space where federal, state, local, and private funds and partnerships coexist to provide early care. A social impact bond (SIB) financed intervention may provide an opportunity to explore bringing these groups together in a more cohesive manner, as well as scale up service provision. Alternatively, local agencies may find a SIB project untenable given the complicated financing nature of pre-existing services. For example, if a preschool program already includes students receiving Head Start benefits (funded by Federal dollars), how would that student be included in the SIB success repayment plan, which uses local funding to repay

investors? The "wrong payor problem" described here is one of many practical and ethical concerns noted in the existing literature on SIB contracting (Jackson, 2013; McHugh et al., 2013; Tse & Warner, 2018; Warner, 2013). The current study aims to collect and analyze data about state and local efforts to make sense of, unify, and expand their preschool service delivery—at least partially in the context of understanding the feasibility of utilizing SIB funding.

Present Use of SIBs

Since the first UK project in 2010, dozens of countries have begun exploring the use of SIB-financed projects to provide both standard and preventative government services. There are 32 active SIBs in the United Kingdom and over 20 in the implementation stage in the United States, where projects are also referred to as "Pay for Success" (PFS) programs (Fraser et al., 2018; Pay for Success Learning Hub, 2018). However, with a small number of fully launched SIB projects and even fewer that have progressed to the point where outcomes are assessed, the SIBs evidence base, including their effectiveness, is largely conceptual and descriptive (Curran, 2017; Maier & Meyer, 2017). Still there is a great deal to be learned by analyzing the mechanisms of SIBs as they are being structured and operationalized within feasibility studies. Furthermore, many of the identifying features of SIBs and motivations underlying their use are not as innovative as proponents would suggest (Warner, 2013). Indeed, many of the challenges observed through the SIB feasibility study process have been subject to significant theoretical and empirical analysis within the research base on public management, performance-based contracting, and pay for results reforms.

In the United States, the Federal government has encouraged development of SIB projects through grants and legislation that support local governments in conducting feasibility

studies to determine potential uses and limitations (*Bipartisan Budget Act of 2018*; White House, 2016). In 2016, the U.S. Department of Education released a request for proposals to fund Preschool PFS feasibility pilots. The purpose of the feasibility pilot grants, awarded to eight entities (totaling approximately $3 million), was to encourage exploration by state and local governments concerning the viability, feasibility and potential effectiveness of SIB/PFS for implementing high-quality preschool programs.

In this paper, I undertake a systematic, qualitative analysis of the Preschool PFS feasibility pilot grant applications and studies, both those funded and not funded by the U.S. Department of Education (DOE) in the 2016 Fiscal Year. We received the 20 applications that were submitted to the U.S. DOE, along with the reviewer ratings and comments (three reviewers per application) and conducted interviews with 12 applicants (5 of 8 awardees, and 7 of 12 non-awardees) approximately 18 months following the funding award decisions[5]. Overall the current study aims to provide evidence about the provision and financing of public preschool programs in sites that have sought to use SIB financing to support their preschool programs. The objectives of the study are threefold: (1) to learn about the state of public preschool in districts and states attempting to expand programming, specifically in terms of current delivery capacity and quality elements; (2) to understand how grant applicants plan to use SIBs to support and/or expand preschool programs; and (3) to assess the status of the feasibility pilots, as well as the preschool program initiatives in sites that were not funded, and uncover the challenges encountered and perceived viability and sustainability of fully executed SIB preschool programs in the U.S., particularly with respect to achieving short- and long-term preschool program goals. Analysis of

[5] I collaborated with Dr. Carolyn J. Heinrich in order to generate the interview guide and coding protocols, and we conducted interviews together.

the feasibility study applications will provide information about the goals, needs, and specific program components of local preschool programs and the partnerships that support them. Additionally, the analysis of applications alongside federal review scores gives us a sense of what characteristics are considered most viable in a SIB model. Finally, interviews with applicant organizations will round out data on applicant motivations, understanding of the process, overall program goals, and updated status on their proposed SIB projects. Overall the data will help address questions regarding the capability of SIBs to expand public preschool, and how local agencies are considering their own preschool structures and investments. In the next section I provide an overview of Social Impact Bond mechanics, current use, and review of the literature.

Understanding Social Impact Bonds

In this section I review the technical structure of SIB financed projects, the potential motivation for using a SIB, and describe several former and ongoing SIB projects. Due to the relatively new and complicated nature of these contracts, it is important to clarify the identifying components of a SIB structure. While I describe a prototypical SIB project, what we see most often in the U.S. is a semi-structured version of the U.K. design, with wide variation in contracting and implementation. The implication of this variation, and the technical difficulty in implementing a SIB financed project, is described further in Heinrich and Kabourek, 2018.

Structure and Mechanics of SIBs

A social impact bond (SIBs) is a financing mechanism that allows a government agency to pay for social programs that produce positive results. Social impact bonds are not "bonds," as

such. They cannot be bought or sold on a financial market. They are more akin to their financial predecessor, performance-based contracting; indeed, target-based performance management, which became popular in the UK during the 2000s, is the financial predecessor to SIBs (Schinckus, 2017; Sinclair et al., 2014; Warner, 2013). Under these "payment-by-results schemes" contracts are linked to specific measurable outcomes. What distinguishes a SIB intervention from a similar public program, such as Head Start, is that government agencies do not finance pilot programs. Rather, private organizations provide up-front financing for implementation, with a government agreement to repay if specified outcomes are met. Governments and private organizations work together to identify research-based interventions, and develop an action plan with specific, measurable outcome goals. An intermediary serves as the arbiter of success and manages the transfer of funds. If outcomes are met, the government agency will begin repayment on the loan. If outcomes are not met, the "loan" is forgiven and becomes a charitable donation—there is no requirement for the government agency to repay the cost of the intervention. Other groups included in the process are the program developer and implementation site(s), participants in the intervention, program evaluator, and sometimes a separate philanthropic financing source that will guarantee some portion of the initial investment.

Social impact bonds allow governments to test interventions with "no risk" in terms of financing. These arrangements are attractive to government agencies because they offer low risk to taxpayers, and hold out the promise of implementing social interventions that will carry savings into the future. For example, an intervention put in place in the juvenile detention center at Rikers Island, New York, had the goal of reducing recidivism for young offenders. This would lead to increased savings down the road for the city and state by not having to pay the costs of incarceration, and decrease the social and judiciary costs associated with crime. While an

evaluation of the intervention found no measurable impact, supporters lauded the experiment a success since it allowed Rikers to pilot the intervention at, presumably, no additional cost to New York City government (Anderson & Phillips, 2016; Burton, 2015).

The PFS concept requires tangible outcomes that are reliable, easy to measure, and quantifiable. Interventions need to have a plausibly causal association with the outcome measures so that a clear pathway of inputs, outputs, and financial rewards can be established. However, identifying and measuring appropriate outcomes of social programs is notoriously difficult (McHugh et al., 2013). For example, an intervention attempting to increase high school graduation rates must establish a proximal, causal connection, set terms around what 'counts' as high school graduation (i.e., graduating with a GED, graduating in more than 4 years), and determine how to distinguish attrition from dropouts versus attrition from school switchers. The difficulty of designing a rigorous, experimental evaluation, and the associated "RCT risk" has led some intermediaries to recommend pursuing outcomes-based evaluations rather than impact-based (Williams, 2018). Impact evaluation requires a causal connection, including a theory of action or mechanism tying the outcome measure to real change; without an impact analysis, we may still observe outcomes achievement without necessarily reaching longer-term goals. In education, third grade reading proficiency may be an *outcome* of interest, because we assume it leads to long-term *impacts* such as on-time grade mobility, high school graduation, and workforce placement. Beyond establishing a measurable outcome and causal connection, a monetary value must be assigned to the outcome. A cost-benefit analysis is a rigorous method for computing monetary value of interventions but is also fraught with its own choices and challenges. sIB-structured deals are premised on government-budget savings as well as long-term investor return, and the associated cost-benefit analyses generally do not account for participant

benefit or the social value of investment. However, very few social initiatives have historically supported consistent government savings, particularly without taking into account participant perspective (Berlin, 2016).

Social impact bonds require a highly organized network of program agents. The four main categories of participants are senior investors, service providers, intermediaries, and outcome funders. In the Rikers Island experiment, the senior investor was the Goldman Sachs' Urban Investment Group, the service provider, the Osborne Association and Friends of Island Academy, the intermediary, MDRC, and the outcome funder the New York City Department of Correction. Additionally, Bloomberg Philanthropies served as an investment guarantee (up to $7.2 million of the original $9.6 million investment) and the Vera Institute of Justice conducted the program evaluation. The mechanics of social impact bonds include four general stages: feasibility study, structuring the deal, implementation, evaluation and repayment. Given the early status of projects in the United States, and the learning curve required to start new SIB-financed projects, the current study focuses on the feasibility study stage of SIB development.

Motivation: Why SIBs?

The Goldman Sachs Social Impact Fund has served as the primary investor for four SIB projects in the United States, including an expansion project for Chicago's Child-Parent Center preschool program. The fund participates in a variety of impact investing projects, in order to "provide clients with access to 'double the bottom line' investments that can provide both a financial return and measurable social impact" (Goldman Sachs, 2014). While companies may make some return on their investment, these are high-risk ventures, given that governments, researchers, and social service providers have been attempting to tackle challenges for at-risk

populations for decades with somewhat limited success (Gustafsson-Wright, Gardiner, & Putcha, 2015; Tse & Warner, 2018; Williams, 2018). Still, the social desirability of contributing to potentially highly effective programs could be a significant motivation. Furthermore, there is a possibility that once social impact bonds become more commonplace in the financial market, this new investment tool could provide new sources of revenue in the long term. Finally, private organizations may feel that they are contributing to the overall level of human capital in society, which they would be able to in turn tap into to fill their own staffing needs, and support economic expansion from a more skilled labor force. For local governments, theoretically serving as the outcome funders, the opportunity to take interventions to scale and potentially avoid future public expenditures are key motivations (Gustafsson-Wright et al., 2015).

A 2015 study surveyed a small number of stakeholders to gain an understanding of their motivation for participating in SIB programs (Gustafsson-Wright et al., 2015). Overall, investors reported to value most highly the opportunity to test the ability of SIBs to address social problems (36%), followed by an equal combination of social and financial returns (23%). A very small proportion of investors (4%) reported financial returns or savings as a primary motivation. This evidence lends itself to the hypothesis that investors are currently interested in testing the capacity for these lending vehicles as a long-term investment strategy. Among outcome funders (government agencies) surveyed, the majority report the opportunity to improve collaboration among public, private, and development funders as the most important motivation to engage in SIBs (24%). After that, responses are tied among the opportunity to test SIBs as a way to solve social problems, scaling up interventions that work, and an equal combination of social and financial returns (about 17% each). Overall, intermediaries and service providers were largely interested in the social returns of SIB programs (Gustafsson-Wright et al., 2015).

Illustrative Examples

The first SIB project, launched in the UK in 2010, was an intervention targeted at reducing recidivism rates at Peterborough prison (HMP Peterborough). Through this pilot project, SIB funding was used to finance interventions for male offenders released from HMP Peterborough who had served short (less than 12 months) prison sentences (Disley & Rubin, 2014). The caseworker model intended to reduce recidivism through an individualized approach to supporting those exiting incarceration, during and after their sentence, by providing access to caseworkers and a variety of resources for transitioning (Disley & Rubin, 2014). A quantitative evaluation used propensity score matching to determine the success of the project, as defined by whether offenders released from Peterborough experience a lower rate of recidivism than comparable offenders released from other prisons (Jolliffe & Hedderman, 2014). The evaluation finds an 8.39% reduction in reoffending rates for the first cohort of treatment individuals, which was insufficient to trigger a success payment (the contract required a minimum 10% reduction) (Jolliffe & Hedderman, 2014). A simultaneous qualitative analysis identified the flexibility of the intervention model as a key strength, while timely data collection and information sharing across partners was a major challenge (Disley & Rubin, 2014). Other reports identified in the systematic review similarly focus on evaluating the success of specific projects in meeting contracted outcomes, as measured by guidelines developed in the feasibility and contracting stages of the SIB-financed intervention (Department for Communities and Local Government, 2015; Dugger & Litan, 2012; Fraser et al., 2018; McKay, 2013; Rotheroe et al., 2013; Rudd et al., 2013).

Following the example set by the UK, the first social impact bond project in the United States began in 2012 at New York City's Rikers Island correctional facility. The motivation for this intervention was to reduce the staggering rate at which juveniles reoffend – at the time of

investment, almost 50 percent of incarcerated youth in Rikers Island return to jail within 12 months of release (City of New York, Office of the Mayor, 2012). MDRC organized the agreement between the New York City Department of Correction, the government agency, and Goldman Sachs, the investor, to provide a cognitive behavioral therapy intervention. Additional funding was provided through Bloomberg Philanthropies. The behavioral intervention used was the Adolescent Behavioral Learning Experience (ABLE), a research-based intervention that has shown promising results in other settings (Rudd et al., 2013). The intervention was given in the school setting at the detention center. The measurable goal was a reduction in recidivism. At a reduction of 10%, payments to Goldman Sachs, the investor, would begin. In this experiment, Goldman Sachs structured the investment as a loan to MDRC, and MDRC contracted with New York City to implement the program (which was delivered by the Osborne Association and Friends of the Island Academy). Bloomberg Philanthropies provided a $7.2 million grant to MDRC over the four-year period of intervention, which MDRC was to use as a payment on the loan in the event that the intervention was unable to produce the agreed upon results (a decrease in reincarcertaion by at least 10%). City payments to MDRC were scheduled to kick in at a reduction of 8.5%, although at a scale smaller than what would be required to pay back the initial loan. The Vera Institute of Justice was contracted to conduct an independent evaluation. The investment was lauded as a successful social experiment despite its inability to reduce recidivism (Anderson & Phillips, 2015; Burton, 2015). A quasi-experimental evaluation conducted by the Vera Institute for Justice found no impact of the intervention on treated teens, where treatment consisted of attending at least one ABLE session (Parsons, Weiss, & Wei, 2016).

Another early SIB-financed intervention began in 2013 in Salt Lake City, Utah, and provided preschool education for at-risk students. Financed by Goldman Sachs, the program was

able to expand preschool education to an additional 595 students in Granite School District (GSD). The existing half-day program cost $1,700 per student for the nine-month school year and was located in schools serving the most at-risk students in GSD. The preschool program is delivered through a mix of public and private providers, including the school district and private care centers. United Way of Salt Lake served as the intermediary and was responsible for overseeing implementation of the intervention, including contracting with and managing payments to service providers. The outcome metric for the intervention was special education receipt at any point from kindergarten through sixth grade by students identified as 'at-risk' for special education through an initial achievement test at preschool entry (the Peabody Picture Vocabulary Test, PPVT). For each year any of the 110 'at-risk' treatment students did not receive special education services, Salt Lake County made an annual payment to Goldman Sachs of 95 percent of the special education money saved (Stevens, 2015). During the feasibility study period, special education services were estimated at $2,600 per child per year (Innocenti, 2015a). The outcome goals are measured using validated administrative data to determine whether or not the identified students receive special education in a given year. In the first year of evaluation, just one of the 110 identified students received services. This resulted in a first-year payment of over $260,000 to Goldman Sachs.

A peer-reviewed qualitative study explores efforts in three U.S. cities to expand early childhood services through SIBs, and in doing so formulates an argument that cities "walk a razor's edge" in attempting to balance supporting public programs and increasing private financializing of the public sector (Tse & Warner, 2018). The authors code interviews with program partners in South Carolina, Utah, and Chicago intervention sites and analyze data along four metrics: systemic change, performance metrics, cost structure, and social equity. All three

sites provide "systemic change" by promoting a new, sustainable public funding stream, but only South Carolina set up a cost structure which maximizes public investment (over investor profit) (Tse & Warner, 2018). The South Carolina program does so by using philanthropic dollars and an increase in state funding for an initial expansion, and then reinvesting success payments, rather than having investor repayments. Since investor profit is not a key factor in the cost structure, the South Carolina program is not representative of the majority of SIB projects (Fraser et al., 2018; Tse & Warner, 2018). Yet, the project was successful in bolstering public support towards sustainable public funding—leading to the expansion of early childhood education services across the state (Tse & Warner, 2018). This leveraging of the SIB intervention may be a way in which effects of short-term private investment ripple out towards sustainable public investment.

Getting Started: Conducting a Feasibility Study

Initiation of a SIB project can come from the government agency, service provider, or intermediary. In practice, a government agency or intermediary will propose a project, conduct a feasibility study, and then seek funding. That said, there are no best practice or restricting guidelines for this process, and it is helpful to already have a relationship between all organizations (Rudd et al., 2013). The process can and does vary. In Rikers Island, the ABLE intervention was one of many interventions considered by the "Young Men's Initiative," an agency created by the office of Mayor Michael Bloomberg in 2011 (The City of New York, Office of the Mayor, 2012). A partnership between the city government and Bloomberg Philanthropies was actively seeking a project that could be supported using a SIB. Conversely, in Utah, the United Way of Salt Lake City was looking for a way to expand an already existing

preschool program when the organization began to look into social impact bond financing. In both instances, the introduction of a SIB program required existing infrastructure and public-private partnerships among multiple institutions.

The initial feasibility study requires undertaking a cost-benefit analysis, to provide evidence that interventions could result in government savings large enough to repay investors and determine the program capacity to be self-sustaining beyond the initial investment. There is also some expectation in the SIB model that investors would be able to earn some modest return, which is one of the motivating factors for investors (Goldman Sachs, 2014). Moving forward, financing the initial feasibility studies, and the staffing availability and financial capacity required to research and develop SIB proposals, may become a barrier to implementation. While a more traditional intervention grant may include a fee for researchers and evaluators, SIB financing requires paying for an initial analysis (feasibility study) in addition to a formal third-party evaluation of the intervention. The U.S. Department of Education (in 2016) and Congress (in 2018) set aside grant dollars toward developing such feasibility studies. The current study leverages data collected from applications and qualitative interviews with the 2016 grant applicants. Analysis of this data provides a glimpse into how local governments are trying to understand and work with this new financing mechanisms, as well as their current and future plans for preschool programming. In the next section I review the methods and analysis plan for the study, followed by results and discussion of policy and research implications.

Methods

Data Collection

The population of this study is the 20 applicants for the 2016 Preschool Pay for Success grant competition through the U.S. Department of Education (see Appendix A, and https://www.govinfo.gov/content/pkg/FR-2016-08-22/pdf/2016-20021.pdf). We obtained applicant information, including original applications as well as reviewer scores and comments (guidelines shown in Appendix B), from the Department of Education (award winner information can be found publicly here: https://www2.ed.gov/programs/pfs/awards.html). From the applications, we identified key study personnel and sent e-mail requests to conduct 1-hour interviews to discuss their experience with the grant process, Pay for Success/Social Impact Bonds, and preschool finance and expansion. We were able to complete interviews with 12 of the 20 applicant organizations, including 5 award winners and 7 additional applicants.

Application Coding

A codebook (shown in Appendix C) was generated to extract data from applications, reviewer comments, and scores. We pre-identified items of potential interest based on a review of the literature on SIB (PFS) feasibility studies and goals. Additionally, we identified items related to preschool capacity, quality, and expansion. The stated goal of the grant competition was to produce feasibility studies that would "determine if this model is an effective strategy to implement preschool programs that are high-quality and yield meaningful results" (U.S. Department of Education, 2016). Therefore, we were particularly interested in applicant and

reviewer attention towards implementing evidence-based preschool practices and rigorous evaluation methods.

Interviews

An interview procedure and guide was created (see Appendix D) and semi-structured interviews were conducted with 12 of the 20 applicants, lasting approximately one hour each. Interviewees included executive staff members at the school (1), city (3), district (1), county (5), and state level (2). The interview protocol is organized around six sections: motivation for pursuing a SIB/PFS feasibility study grant, grant application and planning, partnerships, preschool program, determining and assigning partner roles, program evidence and evaluation, and working or financing outside of the U.S. Department of Education grant. Transcribed interviews were coded using NVivo software.

Analysis

To analyze the interviews, I used an inductive (grounded theory) approach to coding. I conducted initial line-by-line coding following procedures set out in Charmaz (2014). Through initial line-by-line analysis, approximately 127 unique codes were identified (after removing duplicates or nearly identical codes, such as separate codes for "current" and "existing" program structures). Next, I developed focused, thematic codes by grouping nodes by theme, using the seven sections of the interview protocol as a general guide (Appendix D). Fourteen categories emerged, constructed around themes such as financing public preschool programs, goals for preschool expansion, data and evaluation, and building partnerships and partner capacity. The final categories and sample focus codes are shown in Appendix E. Using focus code data, I

constructed a series of iterative matrices to organize and analyze data addressing the research questions. I include a sample matrix in Appendix F.

Data from application coding supplement the interview analysis by providing additional background information and supporting empirical cross-comparison of applicant setting, capacity, and goals.

Results

Ultimately, the data collected during this study span a wide range of topic areas, which are reflected in the Interview Guide in Appendix D. Grounded theory coding was conducted on the full range of data. A companion paper, *Pay for Success Development in the U.S.: Feasible or Failing to Launch?* (Heinrich & Kabourek, 2018), discusses how grant applicants understood SIB mechanics, worked to create or support existing public-private partnerships for SIB implementation, and the specific steps taken while conducting feasibility studies. The current paper focuses on a discussion of findings related to current preschool delivery, capacity, and plans to supplement and/or expand existing programs, within the context of pursuing a Social Impact Bond. Information in the current study offers researchers and practitioners lessons learned during the feasibility study process, including performing needs assessments, identifying capacity and program goals, and determining factors contributing to the viability of SIB financing for public preschool. Future work from this study may include analysis of completed feasibility studies and any resulting SIB contracts.

The results section details findings in three areas: existing site preschool delivery and capacity, implementing and supporting quality programs, and feasibility study outcome measures

and payment challenges. The paper concludes with a discussion of policy implications, future directions for research, and study limitations.

Existing Delivery and Capacity

The current study provides an updated, detailed picture of public preschool in applicant sites. According to the National Institute for Early Education Research (NIEER) 2017 *State of Preschool* annual yearbook, 33% of 4-year-olds, and 5% of 3-year-olds are enrolled in state-funded preschool (Friedman-Krauss, et al., 2018). In comparison, NIEER estimates 8% of 4-year-olds and 8% of 3-year-olds attend federally funded Head Start programs. Aggregate enrollment estimates mask severe inequality between states—seven states have no state preschool program at all, while four serve over 70% of 4-year-olds in the state. The report additionally provides information on spending and quality elements of state preschool. In 2017 states spent, on average, $5,008 per child enrolled in public preschool. In some cases, additional spending came from local or federal contributions. Comparatively, Head Start spending on average was $9,158 per child enrolled. The following section reviews the landscape of preschool delivery in applicant sites, and their goals or motivation in pursuing SIB financing to expand their public preschool programs.

Current Preschool Delivery

Within the population of the study, public preschool is a mixed delivery system that relies on funding and delivery from both public and private partners. Nationally, 36% of preschools use a mix of private and public funding, and within those funded strictly with public dollars, sites most often “braid” or “blend” funds from Head Start, child-care subsidies, and public pre-kindergarten funds at the local or state level (Chaudry & Datta, 2017). The collected grant

applications provided information about the state of publicly provided preschool in the sample sites, including information on current capacity and delivery mechanisms. Application data and interview data confirms that service provision is largely mixed-delivery, with publicly funded preschool classrooms (or seats) in private, for-profit sites, non-profit centers, and public schools. Overall, applications indicated significant anticipated need beyond current site capacity. Needs assessments were typically based on the gap between current capacity and an estimate of the number of preschool-aged children residing in the area. There was acknowledgement that these "back of the envelope" calculations would benefit from a more thorough needs assessment, which would be conducted during the feasibility study process.

Goals and Motivation for Pursuing SIB

The majority of applicants indicated that they hoped to expand the number of preschool seats to existing programs (whether at the district, city, or state level), as opposed to creating a new program or adding program enhancements to existing program delivery. Overall, applicant sites were more likely to be planning expansion of existing programs than creating new preschool programs. Only two applicants did not currently offer preschool at the time of the 2016 grant process. This is not to say that there was no preschool provision in these areas — again, this speaks to the complex, multilayered landscape of preschool delivery. For example, although a single school or city may not offer a preschool program, there can still be state and federal (Head Start) options available to certain populations. Indeed, federal Head Start programs are available in the area occupied by the two applicants mentioned previously. Still, demand more often outpaces availability; sites aiming to provide new avenues for funding noted that this was particularly true of what they named the "working poor" population—those families whose

income levels were above the threshold for Head Start or state preschool but were still priced out of private, center-based care.

While providing additional seats was the most commonly cited goal, several sites sought to expand quality elements or provide add-on enhancements to current programs. This included adding research-based curricular programs or resources for teacher development.

> So we need to give this program 2,000 more dollars for every child that they serve. We are not paying for basic operations. The idea is that while we have a target number of kids that they'll serve, that's because we want to give them enough money to improve quality for that number of kids. But it's not the base funding.

There also appeared to be some tension between expansion and sustainability. Several applicants described the motivation to pursue SIB financing as one of multiple avenues for supporting recent expansion and demand for preschool slots.

> We got [some funding] which is really lasting us over several years, and we added – so like if we added initially 2,600 classes, sort of petered a little bit with some sites no longer being in the program, but it's maintaining the – what we call the extra slots.

The most commonly cited reason for pursuing Social Impact Bond financing was to determine whether this funding stream could provide an additional short-term means of revenue to bridge the gap between current capacity and unserved demand. There was no stated expectation among those interviewed that a SIB contract could be a permanent or even semi-permanent solution to public preschool funding.

Implementing and Supporting Quality Programs

Applicants indicated the use of certain research-based preschool features and leveraged the grant opportunity to evaluate their current programs. The structure of SIBs, with the expectation that government savings will allow for investor repayment, hinges on implementing research-based interventions with proven impact. Research on the long-term effects of preschool

reflects mixed results (e.g., Bailey et al., 2017; Phillips et al., 2017). Therefore, identifying key quality and program elements is not necessarily straightforward. There are ten quality standard benchmarks outlined by NIEER, ranging from developed early learning standards and curriculum supports, to teacher training and class size. Of particular interest in the current study is the presence of a program-wide Continuous Quality Improvement System (CQIS). This is a recent addition to the NIEER benchmark checklist and requires regular data collection on classroom quality as well as some indication that collected data is used to inform policy or practice. Evaluation of Quality Rating and Improvement Systems (QRIS) across states is ongoing and the federal government has invested significant capital to help collect and analyze QRIS validation data (Boller & Maxwell, 2015). Of this recent spate of QRIS research, there is little evidence that higher QRIS scores are associated with higher student outcomes than programs with lower QRIS scores (Karoly, 2016; Sabol et al., 2013).

Aside from the use of QRIS, throughout the grant applications and interviews, there was a focus on the use of research-based curricula or specific program interventions. While most states require sites to use "research-based" curricula, this criterion is loosely defined and leaves open many options for providers. In fact, few publishers describe the evaluation research used to substantiate claims of efficacy (Clements, 2007). For example, nearly half of Head Start centers use the Creative Curriculum, described by the publisher as evidence-based, despite it being rated by What Works Clearinghouse as having "no discernable evidence" in promoting literacy and mathematics skills (Jenkins et al., 2016). Recent research shows that focused curricula on math and literacy boost academic achievement in their respective areas, but the use of popular whole-child curricula does not provide additional advantages (Jenkins & Duncan, 2017). Moreover, preschool teacher fidelity to curricula varies, particularly based on what types of ongoing

professional development and support are available to teachers (Lieber et al., 2009; Davidson, Fields, & Yang, 2009).

This study observed applicant description of quality program features, and we were able to speak with applicants about how they understand the use of quality components in their preschool programs. The following section describes how local and state governments are working with QRIS and other quality measures, and consider how to evaluate program quality.

Quality Standards

Applicants promoted their use of research-based curriculum and QRIS systems to support early learning and monitor program quality. We coded grant applications for several research-based program features, including curriculum and observation tools or specific quality indicators (Table 1). All funded applications (and the majority of unfunded applications) indicated the current use of research-based curricula and some version of observable quality indicators (e.g., QRIS, ECERS). About half of all applicants had an existing evaluation of their preschool program, with favorable results at the end of preschool. Seven applicants (4 funded and 3 unfunded) indicated that sites could choose their own curricula, as long as it was researched-based. There is little information, however, on what resources are available to help sites choose, purchase, and implement curricula. Still, state preschool programs among these applicants were all rated as meeting curriculum support standards by NIEER in their 2017 preschool yearbook. Overall, discussion of specific quality elements in applications was brief. In the review process, applicants could receive score points for any inclusion of quantitative, qualitative, or theoretical evidence in support of their preschool design. This broad requirement explains the wide variation in level of discussion of quality elements. Indeed, reviewer score of applications focused less than 25% on the quality of the preschool design. Of 25 possible reviewer points, applicants

received only 16.5 points on average in this area (20.6 for funded applicants and 13.8 for unfunded applicants). Still, this is perhaps expected given the ongoing research debate regarding "what works" in preschool education (Phillips et al., 2017).

Table 1: Expanding Preschool Quality

	Funded Application		**Unfunded Application**	
Research-Based Feature	Yes	No	Yes	No
Curriculum	8	0	8	3
Quality Indicator	7	1	8	3
- State QRIS	3	5	4	7
Existing Program Evaluation	4	4	5	6

Conducting Evaluations

In this vein, the 2016 federal call for proposals emphasized the potential to use Feasibility Pilot funds to conduct evaluation studies and thus provide new evidence regarding high-quality programs. Award winners utilized grant resources to further determine "what works" in their current preschool program; many used this as an opportunity to conduct pilot studies or update existing evaluations and data systems. Interviewees repeatedly noted that most important feature of the grant was the ability to conduct a thorough needs assessment and support policy conversations regarding appropriate goals and measures for their public preschool programs.

> What we're trying to do right now is do a pilot of our preschool design, and seeking the philanthropic funding to be able to do a small pilot over a three-year period and be able to demonstrate through collection of data and the analysis of that data. The real positive impact that our preschool design has on those children, and then be able to approach the state or philanthropic organizations to say, okay, here it is. We've got proof. It works here.

Outcomes and Payment

In order for a SIB-financed project to be feasible, payable outcome measures must be identified, with an accompanying valuation, potential investors, and government end-payor. In

approaching feasibility studies, applicants identified a variety of potential outcomes and payment options for use in a SIB project. Applicants with existing program evaluations cited positive outcomes from those evaluations, typically higher achievement on kindergarten readiness assessments. Those without existing evaluations relied on the research base of long-term preschool effects, particularly evaluations of Abecedarian and Perry Preschool. Applicants cited this research base as a starting point for considering which outcome measures would be most relevant to their project. From there, conversations with both potential private and government funders was important to shaping an understanding of feasible, payable outcomes. In this section I describe proposed outcomes considered by applicants, as well as challenges that arose through trying to structure a SIB project, including the difficulty of costing out outcome measures, "wrong payor problems," and questions surrounding the ethics of private benefit from public service.

Proposed SIB Outcomes

Applicants largely focused on third grade achievement, early grade retention, and kindergarten readiness as potential payable outcomes. All applicants indicated that some type of academic achievement measure would be prioritized, but there was little thought given to how achievement would incur cost savings. Instead, applicants discussed savings through a decreased use of special education or English Language Learner services, or reduction in grade retention. Reduction of special education services as an outcome measure was carefully discussed by applicants, although it is unclear if they planned to drop the measure entirely or consider something more nuanced than receipt or non-receipt of services. As one interviewee stated, "The conclusions that we came to were basically that the special education outcome measures that we've seen used in the past are a blunt instrument and potentially ethically concerning."

Although most award winners have not yet finalized outcomes in their feasibility studies, through the grant applications and interviews we observe participants engaging in meaningful discussion about realistic goals for preschool implementation.

Supporting Social Emotional Learning

Attention skills, broadly defined and measured, are associated with later academic outcomes (Caprara, et al., 2000; Duncan et al., 2007). In perhaps the most-cited evidence supporting public preschool, Cuhna and Heckman hypothesize that student motivation, persistence, and other skill-building social emotional skills are critical to the persistence of early intervention and student learning (2007). One factor that emerged from the feasibility study process at multiple sites was the desire to measure social-emotional learning (SEL). An applicant describes this learning process.

> Well, one of the things that has become really clear in our community and I think are probably true across the country, is the social-emotional development is a piece that when [our state] started doing their kindergarten assessments four years ago, many superintendents and principals were saying, boy, this is going to be great, because we're going to know how children are prepared, you know, in their letter recognition, and sounds, and numbers, and that, and – what they realized pretty quickly is a more critical factor was how was their self-regulation? How was their interaction with peers and with adults that they didn't have experience with before?

Nearly all sites indicated on their applications that social emotional learning measures would be considered in their feasibility studies; however, those that are completing feasibility studies encountered challenges with how to measure and cost out SEL outcomes. As used by applicants, 'social emotional learning' was an umbrella term that encompassed regulatory behavior, social skills with peers, following directions and listening behavior, and general discipline concerns. One applicant describes how SEL became part of their logic model:

> Right now there's such a focus on strategies to get kids off on a great start in terms of their social skills and their own mental health, so you know, it may be very salable in that regard… and that's where you know, part of our logic model is thinking about okay, so

> we can support the healthy social-emotional development of three and four year olds, they'll enter kindergarten not only more socially ready but more able to effectively engage with all the academic opportunities in K-1, 2.

Costing Out Measurable Outcomes

Existing cost benefit analyses of public preschool have the benefit of experimental or strong quasi-experimental preschool studies, and a multiple-decade time horizon for calculating long-term benefits. Social Impact Bonds, while designed to have a longer time horizon than typical interventions, are still intended to begin outcome payments between 3-8 years after implementation. Furthermore, these potential benefit estimates are specific for high-quality, small-scale programs (Karoly, 2016). Therefore, cost savings must be realized in a shorter time frame. Rigorous cost benefit analyses reflect potential social benefit of $3-17 per every $1 invested in high quality preschool programs (Heckman et al., 2010; Karoly, 2016). These savings were measured as coming from higher graduation rates, reduced involvement with the criminal justice system, and reduced use of social welfare programs. This means that feasibility studies need to base potential cost savings on measures that have not yet been validated through rigorous experiments or previous cost-benefit analyses. Therefore, some measures are easier to "cost out" than others. For example, several states have implemented third grade reading benchmarks (prior to SIB development), where students may be held back in third grade if they are not meeting proficiency standards (Weyer, 2018). At least one grant winner decided to use this as an outcome measure, by equating the "cost" of third grade proficiency to a year of grade retention. However, existing cost-benefit analysis suggests that the economic consequences of grade retention go far beyond the cost of an additional year of education—there are also supplementary program costs and the cost of delayed earnings to both the student and taxpayer (Eide & Goldhaber, 2005). Additionally, there is evidence that grade retention in elementary school has no substantive

impact on high school graduation, while retention in later grades is actually associated with higher likelihood of dropping out of high school (Jacob & Lefgren, 2009). Taken together this evidence suggests there may be an undervaluation of avoidance of grade retention, but, these cost savings will not be realized until much further down the road. This would make it difficult for the government end payor to come up with cost savings to repay loans and loan premiums.

In considering potential payable outcomes, applicants and interviewees placed a considerable emphasis on kindergarten readiness and third grade achievement scores. These outcome metrics were almost universally proposed among grant applicants and winners. Additionally, there was a focus on social emotional learning outcomes, special education referral, and outcomes for English Language Learners. In some cases, social emotional learning outcomes were emphasized by the potential end payor:

> That [the district was] far less interested in academic readiness, [than] social and emotional readiness, and at one point we were talking about sort of the pricing outcomes, and they told us that they were prepared to pay double for a child who was socially and emotionally ready than they were for a child who was academically ready.

Applicants noted that their schools believed academic skills could be easily taught as long as students were able to follow school rules, anticipate the rhythms and structures of full-day school, and display learning behaviors such as listening and following directions.

> Their kindergarten teachers are most interested, you know, children who are … sort of compliant.

Applicants pursuing social emotional learning measures in their feasibility studies were largely frustrated by how to price outcome payments. First, partners must agree on appropriate measures of social emotional learning and behavior; is this an assessed outcome? Assessed by teacher or parent survey, or direct assessment? Does this include discipline referrals? One completed feasibility study anticipated that social-emotional kindergarten readiness would provide indirect

benefits through, for example, reduced involvement of behavioral technicians, and fewer classroom disruptions and removal for suspension. This feasibility study determined that the social emotional development outcome could be weighted at 55% of the total outcome payment, "costing" more than 3rd grade reading or math achievement.

Wrong Payor Problem

A major financial barrier to SIB feasibility highlighted by applicants is the "wrong payor problem," wherein it is unclear who ultimately sees cost savings, and therefore should be responsible for paying for outcomes. Social Impact Bonds are pay-for-outcome mechanisms where an additional fee is paid on top of principal investment. The additional funding comes, in theory, from government savings on future intervention or social services. One of the key questions in this framework becomes, which agency ultimately sees cost savings? In considering preschool, a state or city agency may have more incentive to subsidize a Head Start preschool classroom, where the cost is shouldered by the federal government, but savings are more likely to occur at the state or city level (in the form of reduced costs to local early schooling). More likely is the question of how to handle student mobility, where students receiving preschool in one district or school zone may easily move during their early elementary years, taking the savings benefit with them. Several applicants noted this as a political and financial barrier to SIB feasibility.

> Well, and then there's also the – they call it – I've heard the term wrong check book or wrong pocket problem. So if the state is the payer, are the savings actually state level savings or are they local district savings, or are they a combination?

Who Should Benefit?

Just three the applicants interviewed expressed concern regarding private sector benefit from social programming, which has become a key issue in the theoretical literature on the use of

SIB mechanisms (i.e., Warner, 2013). One interviewee who did express this concern, structured their comment around the additional cost of using the SIB mechanism:

> You know, in a traditional Pay for Success design, there is like an unavoidable leakage of money out of the system, right? Because you are paying a risk premium to the investors. They're generating interest on their loans for whatever it is.

Instead of hesitation around private sector profit, a commonly asked question was, *Who should benefit*? With the "who" not being public versus private sector, but which level of the public sector. This concern coincides with the "wrong payor problem" discussed previously. A particular concern across applicants was student mobility:

> Not to mention that in all of those conversations about, well, I'm not paying for kids that are leaving my district, none of them ever talk about the fact that kids are also coming into their district, and what were they getting somewhere else and who is paying for that.

Discussion

Moving forward, there was little suggestion that applicants from the sites that were interviewed will continue with SIB-financed preschool expansion projects. Applicants were aware of additional costs associated with SIB financing mechanisms. These payment schemes involve more actors than typical performance-based contracting, require more rigorous evaluation, as well as increased political capital to navigate relationships across public and private partners. The use of SIBs overall is on the rise in the U.S., with dozens of projects currently in the feasibility, contracting, or implementation phase. However, these are largely in areas such as recidivism, homelessness, and workforce development, which have a longer history with performance-based contracting mechanisms. Still, the next round of federally funded feasibility studies may provide solutions to some of the barriers encountered by the 2016 applicants. Interviewees for the current study indicated that they were continuing to pursue

financing and expansion options for their public preschool programs, outside of the SIB framework. Across the sample, there was universal support for public preschool, although program and finance mechanisms remain key questions.

> I think the argument for public preschool and treating it as a public good is probably even stronger than the argument for something like public higher education, and so I think from sort of a, you know, a public policy, economic development, child development, social justice and equity perspective, I think it's tough to argue against public preschool.

Research and Policy Implications

Caution should be taken in approaching the use of a SIB financing mechanism to support public preschool. The contracting mechanism of SIBs is complex and requires considerable cost-benefit accounting to show areas of government savings within a medium-term time horizon. Extensive cost benefit analysis can be an expensive, lengthy process for local governments considering entering into a Social Impact Bond or Pay for Success agreement. Still, award winners in the current study found the feasibility stage invaluable for conducting initial or updated evaluations of their current preschool programs. Award winners and those furthest along in the feasibility study process were those sites with established public-private partnerships in preschool finance and delivery. The importance of public-private partnerships was highlighted throughout interviews and will likely be a key factor as local and state programs expand. One way in which public-private partnerships are undertaken in delivery is within blended classrooms, which include seats for students from tuition-paying families, as well as students on (privately funded) scholarship, or using Head Start, state pre-k, or child care subsidy (public) funds.

Implementing "Universal" Preschool

Something potentially overlooked in research on public preschool is providing clear definitions of program elements. An important clarification in interview data was understanding what is meant by "universal pre-k" or the division between federal, state, and city/district programs. Here, the blending and braiding of funding streams is not only a hurdle for providers, but may alter how we interpret evaluations of public programs. For example, in a public preschool classroom with seats funded by Head Start, state pre-k, and tuition-paying families, which program elements are children most exposed to? Do these children benefit from mixed peer groups, beyond the benefits of Head Start standards and curricula? Ultimately, grant applicants were largely focused on increasing access, potentially through the use of mixed classrooms. One applicant describes their efforts to use blended classrooms to get closer to "universal" preschool provision.

> I think there's a contingent in [our area] that would want that to produce universal pre-K … just a vision perspective, that's what I would say is you know, is there a way that we could structure this that would result in universal pre-K. If not that, then what's the next step down? You know, do we expand to full day pre-K three… There's some opportunities I don't think we've explored of looking at our financial model around pre-K to look at how we might be able to do it in a better way, and potentially looking at like tuition paying families versus eligible families, and how to sort of integrate both into our pre-K classrooms and potentially have a financial model that supports expansion as a result of that. … I think in the back of our minds, like universal would be the ultimate goal.

Limitations and Future Directions

There are a number of considerations that limit the internal and external validity of the current study. The participation rate for interviews was approximately 60% of funded and non-funded applicants. Interviews were conducted with select individuals at each site and represent only a select piece of program development and implementation. In terms of external validity, sites selected to apply for the Federal grant and may not be representative of sites nationwide that are considering or have used SIB financed programs. While I have some quantitative data to

triangulate findings from interview data, the empirical data is selective based on 2016 grant applications. Finally, future studies will have to consider findings from the finished feasibility studies. At present, there are too few completed studies to warrant additional analysis. Taking these limitations into consideration, there is still an ample amount to be learned about challenges facing local preschool programs as well as the potential use of SIB financing for program expansion. Future research should consider take-up of preschool SIB or PFS programs, the extent to which implemented projects match feasibility study plans, and their ultimate success in supporting preschool expansion.

References

Anderson, J., & Phillips, A. (July 2, 2016). *What we learned from the nation's first social impact bond*. Retrieved from http://www.huffingtonpost.com/james-anderson/what-we-learned-from-the-_1_b_7710272.html

Bailey, D., Duncan, G.J., Odgers, C.L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness, 10*(1), 7-39. DOI: 10.1080/19345747.2016.1232459

Bassok, D., Gibbs, C.R., & Latham, S. (2018). Preschool and children's outcomes in elementary school: Have patterns changed nationwide between 1998 and 2010? *Child Development* DOI: 10.1111/cdev.13067

Berlin, Gordon L. 2016. Learning from experience: A guide to social impact bond investing. New York: MDRC.

Boller, K., & Maxwell, K. (2015). QRIS research: Looking back and looking forward. *Early Childhood Research Quarterly, 30(B),* 339-342. https://doi.org/10.1016/j.ecresq.2014.10.002

Burton, P. (July 2, 2015). *No success like failure: NY sees social impact bond pluses.* Retrieved from http://www.bondbuyer.com/news/regionalnews/ny-city-officials-social-impact-bond-big-plus-1077971-1.html

Caprara, G.V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P.G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11*(4), 302-306.

Charmaz, K. (2014). *Constructing Grounded Theory*, 2nd Ed. Thousand Oaks, CA: SAGE Publications Inc.

Chaudry, A., & Datta, A.R. (2017). The current landscape for public pre-kindergarten programs. In *The Current State of Scientific Knowledge on Pre-Kindergarten Effects*. Retrieved from: https://www.brookings.edu/wp-content/uploads/2017/04/duke_prekstudy_final_4-4-17_hires.pdf

Chaudry, A., Morrissey, T., Weiland, C., & Yoshikawa, H. (2017). *Cradle to Kindergarten: A New Plan to Combat Inequality*. New York, NY: Russell Sage Foundation.

City of New York, Office of the Mayor. (2012). *Mayor Bloomberg, Deputy Mayor Gibbs, and Corrections Commissioner Schriro announce nation's first social impact bond program* [Press release]. Retrieved from http://www.goldmansachs.com/what-we-do/investing-and-lending/impact-investing/case-studies/social-impact-bond-pdf

Clements, D.H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education, 38*(1), 35-70.

Cunha, F., & Heckman, J. (2007). The technology of skill formation. *The American Economic Review, 97*(2), 31-47.

Curran, Amy Cobb, 2017, Inroads to Innovation: State Adoption of Pay for Success Legislation. Chapman and Cutler, LLP, April.

Davidson, M.R., Fields, M.K., & Yang, J. (2009). A randomized trial study of a preschool literacy curriculum: The importance of implementation. *Journal of Research on Educational Effectiveness, 2*(3), 177-208.

Department for Communities and Local Government (DCLG). (2015). *Qualitative Evaluation of the London Homelessness Social Impact Bond: Second Interim Report*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/414787/Qualitative_evaluation_of_the_London_homelessness_SIB.pdf

Disley, E., & Rubin, J. (2014). *Phase 2 Report from the Payment by Results Social Impact Bond Pilot at HMP Peterborough*, Cambridge: RAND.

Dodd, J.A., & Moody, R. (2011). Outcomes, not process: Towards a new model for European funding in an age of austerity. *Journal of Contemporary European Research, 7*(1) 120-128.

Dugger, R., & Litan, R. (2012). *Early childhood 'pay-for-success' social impact finance: A PKSE bond example to increase school readiness and reduce special education costs*. A report of the Kauffman Foundation and Ready Nation Working Group on Early Childhood Finance Innovation.

Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428-1446. DOI: 10.1037/0012-1649.43.6.1428

Eide, E.R., & Goldhaber, D.D. (2005). Grade retention: What are the costs and benefits? *Journal of Education Finance, 31*(2), 195-214.

Elementary and Secondary Education Act of 1965, Public Law 114-95, 2015.

Fraser, A., Tan, S., Lagarde, M., & Mays, N. (2018). Narratives of promise, narratives of caution: A review of the literature on social impact bonds. Social Policy & Administration, 52, 4–28.

Friedman-Krauss, A.H., Barnett, W.S., Weisenfeld, G.G., Kasmin, R., DiCrecchio, N., & Horowitz, M. (2018). *The state of preschool 2017: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research. Retrieved from

http://nieer.org/wp-content/uploads/2019/02/State-of-Preschool-2017-Full-2-13-19_reduced.pdf

Greenblatt, Jonathan, and Annie Donovan, 2012, The Promise of Pay for Success. Community Development Investment Review.

Goldman Sachs. (2014, May 31). *GS Social Impact Fund* [Press release]. Retrieved from https://www.goldmansachs.com/insights/impact-investing/touts/fact-sheet.pdf

Gustafsson-Wright, E., Gardiner, S., and Putcha, V., 2015, The potential and limitations of impact bonds: Lessons from the first five years of experience worldwide. Washington, DC: Global Economy and Development at Brookings.

Haskins, R. (2017). Financing early childhood programs. In *The Current State of Scientific Knowledge on Pre-Kindergarten Effects*. Retrieved from: https://www.brookings.edu/wp-content/uploads/2017/04/duke_prekstudy_final_4-4-17_hires.pdf

Heckman, J.J., Moon, S.H., Pinto, R., Savelyev, P.A., Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool program. *Journal of Public Economics, 94*(2010), 114-128.

Heinrich, C. J., & Kabourek, S. (2018). Pay for Success Development in the U.S.: Feasible or Failing to Launch? *Manuscript under review.*

Hustedt, J.T., & Barnett, W. S. (2011). Financing early childhood education programs: State, federal, and local issues. *Educational Policy, 25*(1), 167-192. DOI: 10.1177/0895904810386605

Innocenti, M. (2015a). The facts: Behind Utah's social impact bond for early childhood education [Policy brief]. Retrieved from www.payforsuccess.org/sites/default/files/Utah%20Facts.pdf

Jackson, Edward T., 2013, Evaluating social impact bonds: questions, challenges, innovations, and possibilities in measuring outcomes in impact investing. *Community Development 44*(5), 608-616.

Jacob, B.A., & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics, 1*(3), 33-58.

Jenkins, J., Auger, A., Nguyen, T., & Yu, W. (2016). Distinctions without a difference? Preschool curricula and children's development. Working paper: Irvine Network on Interventions in Development.

Jenkins, J.M., & Duncan, G. J. (2017). Do pre-kindergarten curricula matter? In *The Current State of Scientific Knowledge on Pre-Kindergarten Effects*. Retrieved from: https://www.brookings.edu/wp-content/uploads/2017/04/duke_prekstudy_final_4-4-17_hires.pdf

Jolliffe, D., & Hedderman, C. (2014). *Peterborough Social Impact Bond: Final Report on Cohort I Analysis*. London: Ministry of Justice.

Karoly, L.A. (2016). The economic returns to early childhood education. *The Future of Children, 26*(2), 37-55.

Lieber, J., Butera, G., Hanson, M., Palmer, S., Horn, E., Czaja, C., Diamond, K., Goodman-Jansen, G., Daniels, J., Gupta, S., & Odom, S. (2009). Factors that influence the implementation of a new preschool curriculum: Implications for professional development. *Early Education and Development, 20*(3), 456-481. DOI: 10.1080/10409280802506166

Maier, Florentine and Meyer, Michael (2017) *Social Impact Bonds and the Perils of Aligned Interests*. Administrative Sciences, 7 (3). pp. 1-10.

McHugh, N., Sinclair, S., Roy, M., Huckfield, L., & Donaldson, C. (2013). Social impact bonds: A wolf in sheep's clothing? *Journal of Poverty and Social Justice, 21*(3), 247-57. http://dx.doi.org/10.1332/204674313X13812372137921

Parsons, J., Weiss, C., & Wei, Q. (2016). Impact Evaluation of the Adolescent Behavioral Learning Experience (ABLE) Program. New York, NY: Vera Institute of Justice.

Pay for Success Learning Hub. (n.d.) Retrieved from https://www.payforsuccess.org/

Phillips, D.A., Lipsey, M.W., Dodge, K.A., Haskins, R., Bassok, D., Burchinal, M.R., Duncan, G.J., Dynarski, M., Magnuson, K.A., & Weiland, C. (2017). Consensus Statement from the Pre-Kindergarten Task Force. In *The Current State of Scientific Knowledge on Pre-Kindergarten Effects*. Retrieved from: https://www.brookings.edu/wp-content/uploads/2017/04/duke_prekstudy_final_4-4-17_hires.pdf

Rotheroe, A., Joy, I, & Lomax, P. (2013). *Allia's Future for Children Bond: Lessons Learned*. London: New Philanthropy Capital. Retrieved from: http://www.thinknpc.org/publications/the-future-for-children-bond-lessons-learned.

Rudd, T., Nicoletti, E., Misner, K., & Bonsu, J. (2013). Financing promising evidence-based programs: Early lessons from the New York City social impact bond. New York City, NY: MDRC.

Sabol, T.J., Soliday Hong, S.L., Pianta, R.C., & Burchinal, M.R. (2013). Can rating pre-k programs predict children's learning? *Science, 341*(6148), 845-846.

Schinckus, C. (2017). Financial innovation as a potential force for a positive social change: The challenging future of social impact bonds. *Research in International Business and Finance, 39*(B), 727-736. http://dx.doi.org/10.1016/j.ribaf.2015.11.004

Sinclair, S., McHugh, N., Huckfield, L., Roy, M., & Donaldson, C. (2014). Social impact bonds: Shifting the boundaries of citizenship. *Social Policy Review, 26*, DOI: 10.1332/policypress/9781447315568.003.0007

Tse, Allison E. and Warner, Mildred E., 2018, The razor's edge: Social impact bonds and the financialization of early childhood services, Journal of Urban Affairs, DOI: 10.1080/07352166.2018.1465347.

U.S. Government Accountability Office. (2015). Pay for success: Collaboration among federal agencies would be helpful as governments explore new financing mechanisms (GAO-15-646). Washington, DC: Author.

Warner, M. E. (2013). Private finance for public goods: Social impact bonds. Journal of Economic Policy Reform, 16, 303–319.

Weiland, C. (2016). Launching Preschool 2.0: A road map to high-quality public programs at scale. *Behavioral Science & Policy, 2*(1), 37-46.

Weyer, M. (2018). A look at third-grade reading retention policies. *National Conference of State Legislatures* [Policy Brief]. Retrieved from http://www.ncsl.org/documents/legisbriefs/2018/june/LBJune2018_A_Look_at_Third_Grade_Reading_Retention_Policies_goID32459.pdf

Williams, James W. (2018). Surveying the SIB Economy: Social Impact Bonds, 'Local' Challenges, and Shifting Markets in Urban Social Problems. Working paper, York University, Department of Social Science.

White House. (2012). *Paying for success: The federal budget fiscal year 2012*. [Fact sheet]. Retrieved from https://www.whitehouse.gov/omb/factsheet/paying-for-success

White House. (2016). *Improving outcomes through pay for success financing* [Press release]. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/omb/budget/fy2016/assets/fact_sheets/improving-outcomes-through-pay-for-success.pdf

Appendix A: U.S. Preschool Pay for Success Applications

Preschool Pay for Success Project Location	Applicant/Project Leaser	Award Status (2016)
Austin, Texas	Austin Independent School District	Not funded
Clatsop County, Oregon	Clatsop County and Northwest Oregon Kinder Ready Collaborative	Funded
Cuyahoga County, Ohio	Cuyahoga County Office of Early Childhood	Funded
Durham, North Carolina	Durham County	Not funded
Greenville, South Carolina	Legacy Charter School	Funded
Las Vegas, Nevada	City of Las Vegas Department of Youth Development and Strong Start Academy	Not funded
League City, Texas	Clear Creek and Hitchcock Independent School District	Not funded
Mecklenburg County, North Carolina	Mecklenburg County Government and Charlotte-Mecklenburg Schools	Funded
Napa Valley, California	Napa Valley Unified School District and Napa County Office of Education (NCOE)	Funded
New York State	New York State Office of Children and Family Services and Council on Children and Families	Not funded
Pittsburgh, Pennsylvania	Office of Early Childhood, and Citiparks	Not funded
Racine County, Wisconsin	Higher Expectations for Racine County and Racine County Public Schools	Not funded
Rio Rancho, New Mexico	Shining Stars Preschool	Not funded
Santa Clara County, California	Santa Clara County Office of Education	Funded
State of Colorado	Colorado Department of Human Services, Office of Early Childhood	Not funded
State of Hawaii	Office of Hawaiian Affairs and Institute for Native Pacific Education and Culture	Not funded
State of Minnesota	Minnesota Department of Education and school districts	Funded
State of Oklahoma	Oklahoma Department of Education	Not funded
Ventura County, California	Ventura County Office of Education and First 5 Ventura County	Funded
West Sacramento, California	Early Learning Services for the City of West Sacramento and Universal Preschool for West Sacramento	Not funded

Appendix B: Study Instruments

Categories and Guidelines for Scoring PFS Feasibility Study Applications

- *Need for Project*
  - Applicants should clearly state and demonstrate the extent of the problem facing the Target Population using data and other relevant information.
- *Quality of the Preschool Program Design*
  - Applicants should identify clearly specified and measurable outcomes for the preschool program and explain how these outcomes can be achieved by the program.
- *Preschool PFS Partnership*
  - The quality of an existing Preschool PFS Partnership, including the history of the collaboration, or, if a Preschool PFS Partnership does not exist, the quality of the plan to form a Preschool PFS Partnership.
  - The extent to which the roles and responsibilities of members or proposed members of a Preschool PFS Partnership are clearly described and are appropriate and sufficient to successfully implement a PFS project.
- *Quality of the Work Plan*
  - The adequacy of the work plan to achieve the objectives of the proposed Feasibility Study project on time and within budget, including clearly defined responsibilities, timelines, and milestones for accomplishing project tasks on time.
  - The adequacy of procedures for ensuring stakeholder feedback in the operation of the proposed Preschool PFS Feasibility Pilot.
  - The extent to which the time commitments of the project director and team and other key project personnel are appropriate and adequate to meet the objectives of the proposed project.
- *Quality of the Project Leadership and Team*
  - The Secretary will consider the quality of the project leadership and team. The Secretary will consider the extent to which the applicant has the project and financial management experience necessary to manage the Preschool PFS Feasibility Pilot.
- *Budget Narrative*
  - The Secretary will consider the adequacy of resources necessary to complete the Feasibility Study, including any philanthropic or other resources that may be contributed toward the project. In determining the adequacy of resources, the Secretary will consider the extent to which the budget will adequately support program activities and achieve desired outputs and outcomes.
- *Competitive Preference Priority*
  - To meet this priority, an applicant must propose a Feasibility Study to evaluate if PFS is viable that would evaluate social and emotional or Executive Functioning Outcome Measures, or both. These potential outcome measures may be predictive of future school success, cost savings, cost avoidance, and other societal benefits, and may be appropriate to include in a PFS project.

Researcher Codebook for Qualitative Analysis of Feasibility Study Applications

- *Team Composition*
  - Applicant (name)
  - Project leader affiliation
  - Program delivery operator
  - Delivery method (public, private, or both)
  - Intermediary
  - Other Stakeholders
- *Setting Characteristics*
  - Setting (urban, suburban, rural)
  - Currently serving (number of preschool seats)
  - Anticipated need (anticipated number of new seats)
  - Existing resources: other
- *Target Population*
  - Eligibility criteria: age, poverty, English as a Second Language
- *Quality Elements*
  - Existing program expansion (Y/N)
  - Plan to develop new program (Y/N)
  - Research based curriculum (Y/N and name)
  - Research based quality indicators (Y/N and name)
- *Outcomes*
  - Finalized measures (Y/N)
  - Achievement measures (Y/N)
  - Social-emotional measures (Y/N)
  - Other school outcomes
- *Research Design*
  - Data sources
  - Methodology identified (Y/N)
  - Independent evaluator identified (Y/N and name)
- *Budget*
  - Amount requested
  - Amount received

Interview Guide

Section I: Questions for both winning and losing SIB feasibility pilot applications

Part A: Motivation

1. When did your organization first consider pursuing a Social Impact Bond as a means of financing your program and activities?
2. What reasons motivated you to develop a proposal for a Social Impact Bond feasibility pilot through the U.S. Department of Education?
3. Do you see the Social Impact Bond approach as a less risky approach to financing your program operations?
    a. How do you see your level of risk in comparison to the SIB investor in this project? In comparison to the SIB intermediary?

Part B: Grant Application and Planning

4. How much confidence did you have at the time you submitted your proposal that the proposed feasibility pilot could lead to a successful Social Impact Bond working arrangement? Please choose among the following:
    10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
5. What did (or do) you see as the strongest aspects of your Social Impact Bond feasibility pilot application?
6. What did (or do) you see as the weakest aspects of your Social Impact Bond feasibility pilot application?
7. What do you think is the most innovative or transformative feature of Social Impact Bonds?
8. How did you choose your evidence-based model for this Social Impact Bond project?
9. What other models did you consider? Are you confident in the evidence underlying your chosen model?
10. Do you plan to continue with this evidence-based model?

Part C: Partnerships

11. How did you identify your proposed project partners for the Social Impact Bond proposal?
12. Do the types of partners you considered or identified look any different than the collaborations you have formed in the past to conduct your work?
13. Did you identify a potential investor for your Social Impact Bond proposal? How did you go about this, or did an investor approach you to develop the proposal?
14. What do you see as most innovative about the Social Impact Bond approach to public-private partnerships (if anything)?
15. In what ways does (or would) a Social Impact Bond change the nature of your relationship to your project partners, compared to the typical ways you arrange contracts for service delivery or other program operations?
16. Did you select an independent evaluator for the Social Impact Bond project?
17. Have you used independent evaluators previously to assess the effectiveness of the work in your organization?
18. (If yes to both .16 and .17): Have you worked with this particular evaluator previously at your organization?

Part D: Preschool Program

19. Do you currently offer a public preschool program (aside from Head Start)?
    a. Where are preschool programs offered? (i.e., center-based, school-based, etc)
20. Of the students enrolling in public kindergarten in your area, what is the approximate proportion of students who have attended public preschool, Head Start, or private care?
21. Was there an existing public-private partnership to support/provide preschool prior to the Social Impact Bond project?
    a. What were the goals of that partnership?
22. Some national public preschool evaluations find that preschool intervention effects "fade out" by third grade.
    a. At this point, have you been able to track or determine any medium- or long-term outcomes?
    b. What do you see as the biggest challenge facing your current preschool program, in terms of its effectiveness towards long-term outcomes?
23. What is the estimated capacity of your current preschool program, in terms of classroom space, staff, and expenses?
    a. How was local capacity taken into account when undertaking the SIB feasibility proposal/study?

Section II: Questions for winning proposals only

Part E: Determining and Assigning Roles

24. Have you settled on an intermediary for carrying out your Social Impact Bond?
25. Have you settled on an investor for financing your Social Impact Bond?
26. Are you reaching out to populations that are otherwise less likely to be served (or more costly to serve)? How is this reflected in repayment terms for the SIB?
27. What types of roles are the intermediator and/or investor playing in executing the Social Impact Bond pilot activities?
28. What strategies have you used in the Social Impact Bond partnership to make key decisions, for example, in balancing stakeholder interests and authority over different aspects of the project?
29. How much influence does each Social Impact Bond project partner or stakeholder have in determining the following (note: please also indicate *which* partner(s) or stakeholder(s) have a role in these tasks):
    a. Which outcomes to measure
    b. The target population/eligibility criteria and number to serve
    c. Measures and methodologies for evaluation
    d. Terms of re-payment of investor(s) and timeline for re-payment
    e. Budget items
    f. Project deliverables?
30. Have any conflicts of interest or related problems emerged among the implementing partners?

Part F: Evidence and Evaluation

31. What types of evidence show that the proposed intervention could lead to government savings large enough to repay investors?

32. Did you conduct a cost-benefit analysis (CBA) to evaluate the viability of a SIB project?
    a. How did you determine the key cost/benefit components for the CBA?
    b. Is there a publicly available copy of the CBA that you could share with us?
33. Does the expectation for employing an evidence-based model limit (or support) the testing of new innovations?
34. Will you measure SIB program impacts with an experimental or quasi-experimental research design? How are the project deliverables linked to impact measures?
35. How will the intermediary/arbitrator determine how much should be repaid to the investor based on the results?
36. Have you established the terms of repayment in the contract, or have these provisions been discussed in the feasibility pilot?
37. At this point in the feasibility project, how much confidence do you have that your Social Impact Bond project will succeed? Please choose among the following:
    10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Section III: Questions for denied proposals only

Part G: Working Outside of the USDOE Grant

38. Having not been selected for the U.S. Dept of Education Social Impact Bond feasibility pilot, are you still pursuing a Social Impact Bond arrangement?
39. Are you pursuing your project/program goals through other means of collaboration or funding? (If so, what are you doing instead?)
40. Are you working to implement the evidence-based model included in your application? (Why or why not)?
41. How likely do you think it is that you will undertake a Social Impact Bond project in the future? Please choose among the following:
    10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Appendix C: Emergent Themes and Sample Codes

Thematic Category	Sample Focus Codes
Applicant and organization background information	Interviewee experience with SIB/PFS Applicant history with SIB/PFS Demographic information on district (/city/state)
Motivation to pursue PFS and Understanding of PFS structure	Technical understanding of SIB/PFS Experience with performance based contracting Using SIB/PFS as opportunity to innovate
Grant application and planning	Benefit of grant process Timeline of grant application and evaluation Difficulty with application
Motivation to expand preschool program	Public preschool as public good Demand for additional seats Belief in positive outcomes
Plans for preschool growth	Current plans to add seats Current plans to reform preschool __ (curriculum, teaching force, site expansion) Progress toward expansion (including financing)
Funding outside of PFS Feasibility grant	Existing financial capacity Legislative change Combination or "braided" funding
Program evaluation	Previous evaluation results Informal evaluation during application process Using PFS pilot program for evaluation
Data and evaluation methods	Availability of data Accessibility
Preschool outcomes	Measuring outcomes Long- vs. short-term outcomes Potential outcomes of interest
Evidence based or "quality" features of preschool program	Implementing evidence based changes Current features of preschool program Quality indicators
Capacity	Finding capacity (facility, instructional) for expansion Current capacity (facility, instructional) Capacity challenges to feasibility
Partnerships	Experience with contracting Primary partnerships: technical assistance Working with intermediary
Policies and politics	District (city/state) commitment to preschool Use of political capital District (city/state) changes to address project (funding, evaluation, data)

Appendix D: Sample Analysis Matrix

Preschool Financing Capacity and Goals

Quote	Notes
"We have removed demonstration projects because they've been implementing [the program] for a very long time, so they have a lot of funders, private funders, that support their early childhood program. And then we have programs that also braid and blend with the preschool special education fund, and they also braid and blend with childcare, Head Start, and general fund dollars. So we actually just published the guide on braiding and blending the fund." "..the private pay might be 100%. It could be just the copay if a parent has the subsidy…We also have something called expansion slots, which are half day slots that the state gives away. So they could have those state funded slots… they could have some funding for their Special Ed slots if Special Ed kids are also in [preschool]…Some sites get a little bit of money from the USDA…And I think that Head Start, they'll have a crazy quilt of funding, some of them have their own fundraising, etc." "There's some opportunities I don't think we've explored of looking at our financial model around pre-K to look at how we might be able to do it in a better way, and potentially looking at like tuition paying families versus eligible families, and how to sort of integrate both into our pre-K classrooms and potentially have a financial model that supports expansion as a result of that."	What do we mean when we talk about "state," "universal," or "public" preschool more generally? SIBs and PFS agreements are complicated, but in line with challenges programs already face.
"So we need to give this program 2,000 more dollars for every child that they serve…So we are not paying for basic operations. The idea is that while we have a target number of kids that they'll serve, that's because we want to give them enough money to improve quality for that number of kids. But it's not the main source of funding." "…There was maybe $3,000 here, $5,000 here…they were able to cobble together $2 million of the quality improvement fund, so that's kind of how we're moving forward."	In some instances, local agencies are trying to supplement existing programs that are in need of quality improvements or have struggled with sustainability
"Always intended to target with our dollars for preschool the population that doesn't qualify for either Head Start or state preschool. So we really wanted to serve the, you know, working poor population, the families that are just a little bit over income to qualify for a publicly funded space."	Or, targeting gaps in service provision