

RESEARCH

Open Access



The effects of public pre-kindergarten attendance on first grade literacy achievement: a district study

Michael Haslip*

*Correspondence:
haslip@drexel.edu
Drexel University, 3401
Market St, Philadelphia, PA
19104, USA

Abstract

This study investigated the link between public pre-kindergarten attendance and first grade literacy achievement in the United States. Participants ($N = 1056$; first grade children from one mixed-urban city in Virginia in 2012–2013) had either attended district-provided public Pre-K or had not attended formal or institutional preschool of any kind. Significant effects were found on each literacy measure in both the beginning and the middle of first grade (avg. $ES = .32$ and predicted gain of 13%). On average, Pre-K attendees were meeting reading benchmarks in the beginning and middle of first grade, while the no-preschool group was reading below the benchmarks during both time points. Findings strengthen the evidence base for the ability of public Pre-K to improve children's literacy achievement in first grade while also describing a transferable example of universal Pre-K at the district level which policy-makers and practitioners may wish to consider.

Keywords: Preschool, Pre-kindergarten, Longitudinal, Effect, First grade, Literacy, Reading

Background

Although some countries have significantly expanded public preschool, a major global trend remains the proliferation of private preschool due to government under-investment (Haslip and Gullo 2017; UNESCO 2015). In the United States, state-funded pre-kindergarten (Pre-K) operated by school districts serves more children than private preschool or federally funded Head Start programs (Barnett et al. 2011). Pre-K was the fastest growing preschool movement in the U.S. for 10 years reaching an enrollment of 1.4 million children by 2007 (Barnett et al. 2008) until enrollment growth stopped, primarily because of limited financial resources brought on by the 2008 financial crisis. In 2015, Pre-K attendance remained at 1.4 million children with 42 states providing public Pre-K programs (Barnett et al. 2016). State departments of education and research institutes have evaluated some of these state Pre-K programs to examine their effectiveness at improving school readiness and later grade school outcomes.¹

Understanding the ability of state-funded Pre-K attendance to have sustained impacts into grade school is important because public Pre-K is the largest provider of

¹ Public preschool has assumed various forms such as Head Start, targeted Pre-K with enrollment criteria, universal Pre-K, and comprehensive PK-3 programs.

preschool slots in the country, because Pre-K is seen as a potential contributor to closing the achievement gap between at-risk and other children (Frede and Barnett 2011), and because universal Pre-K, discarding low-income criteria, is expected to grow. Comprehensive and intensive preschool designed to assist vulnerable children (e.g., Abecedarian, High/Scope Perry, and Child–Parent Centers) produces sustained long-term impacts (Reynolds et al. 2010), but the potential for “typical” state-funded Pre-K (offered widely or universally by school districts) to sustain longer-term gains is less understood.

Most literature about the grade school impacts of attending public preschool have examined *targeted* Pre-K programs (e.g., income dependent) which do not permit universal enrollment. Given the debate about transitioning from targeted to universal Pre-K in the United States (Barnett 2010), there is a need for greater evidence to be drawn from existing *universal* programs when drawing conclusions about the ability of Pre-K attendance to impact student achievement over time. This paper operationally defines Pre-K as state-administered and state-funded public preschool operated by local school districts (not Head Start, private preschool, or comprehensive preschool interventions).²

Purpose and significance

The purpose of this study was to examine how attendance in a U.S. public school district’s Pre-K program affected children’s later literacy achievement in the beginning and middle of first grade. A related purpose is to contribute to our understanding of the sustainability of Pre-K impacts when the arena of operation is a school district which has scaled up public Pre-K to nearly universal enrollment.

A number of studies have examined the grade school impact of state-funded Pre-K attendance in the United States with positive, mixed, or negative results (e.g., Huang et al. 2012; Lipsey et al. 2015; Magnuson 2007b; Maloffeva et al. 2007). This study investigates a specific school district’s Pre-K program by including the entire population of a city’s public first grade cohort using a robust matching method with valid and reliable literacy measures across two time points. By presenting an example of universal Pre-K access in a mixed-urban setting, in a state that lacks a mandate for universal Pre-K, other states and districts may be interested in this example as they seek to “create a more coherent and uniform platform” (Pianta et al. 2009), promote a Pre-K-3 model (Pianta et al. 2009), and scale up Pre-K access at the district level.

Theoretical framework

Numerous theoretical perspectives inform the relationship between preschool and literacy achievement (Tracey and Morrow 2006). The current study examined the effect of attending public Pre-K on first grade literacy achievement from the perspective of Emergent Literacy Theory (EMT), which describes how reading-related behaviors gradually emerge before formal reading and writing is accomplished (Hall 1987). Using EMT, it is hypothesized that ongoing participation in a literacy-rich environment, such as the district Pre-K program, should correlate with higher literacy achievement in first grade. The school district in question created a literacy-focused environment using a teacher-centered language arts curriculum, described later.

² For a bibliography of these studies, see Child Care and Early Education Research Connections at <http://www.researchconnections.org/childcare/resources/32060/pdf>.

EMT was also used by Valenti and Tracey (2009) in a similar study about the effect of Pre-K attendance on first grade literacy achievement. As they quoted:

Emergent Literacy Theory underscores the finding that although many factors are important to children's reading success, including parents' education, occupation, and socioeconomic level, the quality of the literacy environment correlates most closely with children's early literacy ability (Tracey and Morrow 2006, p. 86).

Literature

Background to targeted and universal preschool

Vulnerable children are often resilient and tenacious in the face of adversity (Luthar 2003) but opportunities to develop to their full potential are limited. Disparity between low-income and middle-class children creates negative consequences (Knudsen et al. 2006) including an achievement gap at school entry, reduced cognitive development, a lower achievement trajectory in later grades, and increased delinquency and crime in later years. Early education has a significant impact on well-being, attainment of higher education, adult health and income in later life (Knudsen et al. 2006).

Publicly funded preschool, as an intensive intervention for children from poverty, has been extensively researched over several decades and can be an effective support for child development (Camilli et al. 2010). Support for public preschool has grown with the increasing understanding that intervention helps close the achievement gap, supports families to improve child health, facilitates emotional development, reduces criminality, and strengthens achievement (Temple et al. 2010). Public preschool also serves the society in a variety of ways, for example, by increasing productivity (Heckman and Masterov 2007).

Research suggests that only high-quality interventions are capable of making up the differences for vulnerable children (Sylva et al. 2011). Further, research shows greater effects of sustained participation in high-quality programs, calling for partnership among agencies and alignment of the PK-3 continuum (Reynolds et al. 2010). Toward these ends, researchers have sought to understand the relationship between Pre-K attendance and later school outcomes, beyond kindergarten entry (Magnuson et al. 2007a), to address the question of Pre-K quality, PK-3 alignment, and later school quality. A common goal is to help close the achievement gap by ensuring that vulnerable children, and all children, receive a high-quality early education from preschool through third grade and beyond.

Despite the importance of targeted preschool, which limits participation to children from low-income families or children with other risk factors, universal public preschool is needed for multiple reasons. Barnett (2010) summarizes the argument for a policy shift in the United States from targeted to universal public preschool as follows: (1) the stigma associated with "programs for the poor" may reduce participation, (2) as family incomes and income criteria both change in an uncertain economy, families in the income margins may lose access and cannot predict their future eligibility, (3) peer effects on learning may be greater under universal public preschool because of more heterogeneous classrooms (Mashburn et al. 2009) and teacher expectations may be greater in heterogeneous classrooms, (4) political support for preschool may increase when more higher-income

families are advocating to improve the quality of their children's public preschool education, (5) significantly more children from low- and moderate-income families will enroll, and (6) societal economic return is likely to be greater than additional costs.

The longitudinal effect of Pre-K attendance

Literature on the longitudinal effect of Pre-K attendance seeks to determine realistic expectations for the influence of Pre-K, to investigate the perceived quality of such Pre-K programs, to highlight challenges in Pre-K and K-12 alignment, and to evaluate later school quality in a framework of PK-3 education where the intent is to sustain preschool gains for disadvantaged children into elementary school (Reynolds et al. 2010). Findings from Pre-K effect studies examining grade school outcomes are relatively promising but still leave unresolved the question about the ability of typical public Pre-K attendance to significantly improve academic achievement beyond kindergarten readiness.

A number of studies have examined the correlation between public preschool attendance and early grade literacy achievement, with mixed results. For example, one study found that the main effect of a public preschool program was no longer statistically significant at the end of first grade (Huang et al. 2012). In another study, researchers found significant outperformance among preschool attendees in the middle but not the beginning of first grade (Valenti and Tracey 2009), and in another study no significant effects were found in first grade but sleeper effects appeared in third grade (Magnuson et al. 2007a). In a randomized control trial examining preschool attendance, no statistically significant differences were found at the end of kindergarten or first grade, and by the end of second and third grade Pre-K attendees underperformed children who received other forms of care (Lipsey et al. 2013, 2015).

There is a small body of evidence revealing positive longer-term gains in literacy following public preschool attendance. One study tracked a 2004–2005 cohort of preschoolers through fifth grade, finding effects on second grade receptive vocabulary, math and reading comprehension, and persistent effects in fourth and fifth grade language arts, math, and science (Frede et al. 2009). Finally, the Arkansas Better Chance Pre-K program tracked children from preschool through fourth grade (2005–2010), finding significant effects on (1) language at the end of kindergarten; (2) math, language, and literacy at the end of first and second grade; and (3) literacy at the end of third grade (Hustedt et al. 2008; Jung et al. 2013).

Key challenges face longitudinal Pre-K researchers, including differing data systems, different assessment measures, and varying state standards, among other issues (Hernandez 2012). Most Pre-K longitudinal studies have not measured or identified elements of process quality (i.e., interactions, learning activities, routines and materials related to the curriculum), hindering replicability among practitioners even when impactful programs are found. Baseline Pre-K process quality is rarely measured, making longer-term impacts difficult to accurately assess. The specific preschool curriculum that was used is also rarely identified. Furthermore, little attention has been given to affective outcome measures such as social and emotional learning. Methodological challenges also abound. Lack of adequate group equivalence through strong matched-pair designs has been a known concern (Maloffeva et al. 2007) with some recent studies responding to the call to employ robust matching methods (Hill et al. 2015; Moore et al. 2015). Studies matching

groups on demographic covariates alone are critiqued for lacking baseline achievement variables (Farran and Lipsey 2015). The need for random control trials remains acute. There is also a need to account for later school quality, for example, by measuring the amount and type of later instruction provided to treatment and comparison groups to better understand the variables influencing fade-out or catch-up effects.

The extent to which educators and policy-makers should expect Pre-K to contribute to longer-term child outcomes requires more and better evidence to solidify consensus. Most studies evaluate targeted or criteria-dependent Pre-K (e.g., low-income criteria) because universal Pre-K programs are rare. However, the debate about a national transition from targeted Pre-K to universal Pre-K should be informed by evidence drawn from more studies of universal Pre-K itself. As cited, there is also a need for studies to employ more robust matching methods to improve the quality of such evidence. The current study aims to help clarify these questions, and fill these gaps, by choosing a universal Pre-K program as its area of focus while employing a robust matching design to contribute to the evidence base about the ability of state-funded and district-run Pre-K to sustain children's literacy outcomes beyond school readiness.

Research questions

Two research questions guided this study: (1) Is district Pre-K attendance associated with a significant difference in letter–sound identification and word identification during the beginning and middle of first grade? (2) Is district Pre-K attendance associated with a significant difference in text-level reading ability during the beginning and middle of first grade? It was hypothesized that literacy achievement among Pre-K attendees would be higher during the beginning and middle of first grade, compared to the no-preschool group, for each literacy measure administered.

Methods

This retrospective cohort study used a causal comparative design and archival data to observe text-level reading ability and other literacy changes between the comparison and treatment groups based on the exogenous independent variable (Pre-K attendance). Data existed on who attended Pre-K, who attended other forms of preschool, and who had not attended preschool of any kind, as indicated by parent survey at kindergarten entry. The comparison group was limited to children who attended no preschool of any kind, while the treatment group contained only children who attended the district-run public Pre-K program. Propensity score matching to approximate baseline equivalence between the treatment and comparison groups was chosen because it can reduce selection bias to an acceptable minimum, permitting a degree of causal inference in the absence of random assignment (Stuart and Rubin 2004).

Context of the study

District setting

This study chose as its research setting a large pre-kindergarten program operated in a mixed-urban school district in Virginia. The school district serves about 30,000 students from Pre-K through high school. There are 24 elementary schools, 14 of which receive

Title I funding. In the early 2000s, the district converted four schools into early childhood centers to be used exclusively for the Pre-K program.

Fifty-four percent of students in the district are Black, 28% are White, 11% are Hispanic, and 7% are mixed or other races. About 58% of PK-12 students are economically disadvantaged. The first grade cohort in this study is slightly different: Black (51%), White (25%), Hispanic and/or Mixed (20%), and Asian (3%). Sixty-seven percent of first grade children in the city (2012–2013) received free (60%) or reduced-price (7%) lunches. Poverty rates among young children are typically higher than older children across the United States. Higher poverty rates correlate with higher mobility rates. This population is significantly mobile: 30% of all kindergartners in the district in 2011–2012 moved to a different school for first grade.

District Pre-K program and quality

The district Pre-K is a free, full-day preschool for 4-year-olds run by the public school system. Four early childhood centers combine to serve close to 2000 preschoolers per year, with slots available to about 99% of children who apply. There are no exclusion criteria for applicants. Centers range in capacity from ten classrooms serving 180 children to 34 classrooms serving 612 preschoolers. Eighteen students are assigned to each classroom, along with a certified teacher who is endorsed in early childhood education and an instructional assistant. All four centers teach the same curriculum. Funds are contributed from the Title I program, from the State, and from the school district's budget. Collaborative special education classrooms are provided at each site and ESL is available.

All open slots are awarded based on academic need following a prescreening test. Students with the lowest scores are placed on the top of the ranking system, moving down the ranked list as scores rise. No student is formally "rejected," although their name does not come to the top of the list for selection until students with lower academic readiness scores are selected. A student's place in the ranking changes with every screening, as more students' scores are added to the total pool of applicants. Nearly all high-scoring 4-year-olds are eventually admitted because of the large number of spots available in the Pre-K centers. Thirty preschoolers were unplaced for the start of the 2013–2014 academic year when the largest center reached capacity. The other three centers admitted all applicants. Students are zoned to attend an assigned center based on home address. Transportation is provided to and from school. Children must be 4 years old by September 30th and a city resident to be admitted.

Structural indicators are reported by the district to suggest overall Pre-K program quality. Structural features include a limit of 18 children per classroom; a certified teacher in every room with a trained teaching assistant (i.e., 2 years of college or passing an academic assessment); early childhood certification for every teacher to support developmentally appropriate practice; a principal at every early childhood center; and a shared standards-based curriculum across all centers based on the state preschool learning objectives. Structural quality is the same across all four centers. The length of the instructional day, and the number of minutes for various parts of the day, is likewise the same across all centers. Low teacher turnover exists in the Pre-K program. Out of at least 92 certified Pre-K teachers, just two applied to transfer out of Pre-K during the 2012–2013 year. Parity exists between the district's Pre-K and K-3 teachers' starting

salary, salary schedule, fringe benefits, and paid planning or professional development time, which is rare throughout the country (Barnett et al. 2016).

Pre-K schedule and curriculum

Instruction was designed in terms of units based on standards and objectives set by the state for public preschool. The day was seven and a half hours long, inclusive of a 30-min lunch, a 30-min free recess block, a 30-min structured physical education program, and two center rotation blocks: one for language arts and one for open centers in math, science, drama, and transportation. Children did not attend other resource classes (art, music, library, computer lab). Therefore, teachers did not have a planning block during the day. No daily snack or nap time was provided.

The language arts curriculum being used during the 2010 academic year was based on the Harcourt Trophies® Pre-K Program, which came to the city in 2004. Math was taught for a few days a week. Math- or science-related centers were rotated but there was no formal math curriculum. In 2010, the district's department of curriculum and instruction was not involved in the Pre-K program. Curriculum was divided into a series of themes and units. A daily literacy lesson plan was prepared for teachers covering each day of the week, as day 1, day 2, and so on. Center activities were suggested in the curriculum in the areas of literacy, writing, listening, math, science, art, dramatic play, manipulatives, water, sand, and computers.

Participants

The population consisted of all 2221 first grade students in the public school district during the 2012–2013 academic year. To arrive at the study's sample, 623 students were first removed from the dataset because they received a different form of preschool, such as Head Start or private providers, or had no preschool information reported on the parent intake survey. This resulted in a sample of 1598 students (treatment, $n = 1269$; comparison, $n = 329$). Students were then removed from the sample for the following reasons: (a) missing demographic data necessary for later matching, (b) transferred into the district (incorrect survey information about former preschool experience), and (c) a child was missing all literacy scores necessary for analysis. Removal for these reasons resulted in a final sample before matching of $T = 1197$; $C = 176$. Each comparison case was then matched to five treatment cases (1:5 fixed ratio) using the optimal propensity score matching technique, described below, resulting in a final post-match sample of 176 comparison cases and 880 treatment cases, or 1056 children overall. The matching process computationally excluded 317 treatment cases (from $T = 1197$ to $T = 880$) because propensity scores were not close enough to be matched into a set.

Matching

Children who attended district-run Pre-K were matched with similar children who did not attend any type of formal or institutional preschool, such as Head Start or private providers. Propensity score matching (PSM) was used to assign a single score to every participant representing their likelihood of receiving the treatment. The propensity score represents all covariates as a single summed score calculated for each child (Stuart and Rubin 2004). PSM allows group assignment to be completed by controlling for

a large number of covariates. This reduces the error introduced by selection bias and can approximate random assignment when large sample sizes are used (such as 1000 or more) and when matching takes place using a large number of covariates (usually 10 or more) (Stuart and Rubin 2004). The current study met these conditions.

The group sizes in this study were widely different from one another, so optimal matching calling the MatchIt package (Ho et al. 2011) in the R statistical analysis software on a one-to-many fixed ratio was performed, where each comparison unit was matched to several treatment units (1:5 ratio). Matching each treatment case to multiple comparison cases ensures that the matched sample preserves enough cases and is representative of the target population. It also allows sufficient statistical power to detect the treatment effect. Optimal matching was chosen because it aims to minimize the average distance between the treatment and comparison groups. Covariates were matched on all available demographic variables which included age, sex, race/ethnicity, SES, and disability.³ First grade disability status was included as one of the matching variables because it is rarely identified in preschool. As a result of fixed-ratio matching, 176 comparison cases were matched to 880 treatment cases ($176 * 5 = 880$; sample size of 1056).

Results of propensity score matching (Table 1) show that the average absolute standardized difference between the two groups was reduced from the pre-match level of 12.12% to the post-match average of 5.1%, which is sufficiently balanced to conduct causal comparative analyses. Only one covariate remains above the 10% threshold (Hispanic and White). If one or two covariates have an absolute standardized difference above 10% after matching (such as the Hispanic and White race variable at 14.94%), the two groups are still considered to be sufficiently balanced overall, and analysis can proceed so long as the average standardized difference is below 10% (Rosenbaum and Rubin 1983).

Data collection procedures

Demographic and academic achievement data were collected from the district central office following institutional review board approval and district research approval. Assessments were administered in the beginning and middle of the first grade year to all children. Testing took place through individual teacher–child conferences by the child’s lead teacher. Lead teachers received training in the proper scoring of each assessment by a reading specialist. Lead teacher qualifications included a minimum of a bachelor’s degree and a PK-3 or PK-6 teaching license.

Measuring early literacy skills

The Phonological Awareness Literacy Screening (PALS 1–3) was used to measure early literacy skills. The PALS is a criterion-referenced assessment administered fully twice a year (fall, spring) to each child in a one-on-one teacher conference. The district in this study also administers several PALS subtests at mid-year. The PALS identifies if children are at risk of reading difficulty or not, according to grade-level expectations (Invernizzi et al. 2003). PALS assessments (PALS-PreK, PALS-K, and PALS 1–3) are used by every

³ The proxy for SES was free, reduced, or full-price lunch. Disability included speech/language disability and learning disability. Ethnicity and race codes included Black, Asian, Mixed, Black and White, and Hispanic and White.

Table 1 Pre/post-match balance results

Pre-match Covariates	Comparison (N = 176)		Treatment (N = 1197)		Standardized difference	Absolute SD
	Mean	SD	Mean	SD		
Age	6.55	.57	6.52	.53	5.55	5.55
Black	41.00%	49.00%	57.00%	50.00%	− 31.50	31.50
Asian	5.00%	21.00%	3.00%	17.00%	9.03	9.03
Mixed	9.00%	29.00%	10.00%	30.00%	− 2.05	2.05
Black_white	5.00%	22.00%	4.00%	19.00%	6.14	6.14
Hispanic_white	10.00%	30.00%	5.00%	21.00%	21.58	21.58
Male	57.00%	50.00%	51.00%	50.00%	12.57	12.57
Speech	2.00%	15.00%	3.00%	18.00%	− 6.47	6.47
Learning	1.00%	11.00%	1.00%	11.00%	− .31	.31
Free	69.00%	46.00%	62.00%	49.00%	13.88	13.88
Reduced	3.00%	18.00%	9.00%	29.00%	− 24.21	24.21
Average						12.12%

Post-Match Covariates	Comparison (N = 176)		Treatment (N = 880)		Standardized difference	Absolute SD
	Mean	SD	Mean	SD		
Age	6.55	.57	6.54	.53	2.67	2.67
Black	41.00%	49.00%	46.00%	50.00%	− 9.61	9.61
Asian	5.00%	21.00%	4.00%	19.00%	4.58	4.58
Mixed	9.00%	29.00%	11.00%	31.00%	− 6.78	6.78
Black_white	5.00%	22.00%	4.00%	20.00%	3.75	3.75
Hispanic_white	10.00%	30.00%	6.00%	24.00%	14.94	14.94
Male	57.00%	50.00%	59.00%	49.00%	− 2.76	2.76
Speech	2.00%	15.00%	2.00%	15.00%	.77	.77
Learning	1.00%	11.00%	1.00%	9.00%	2.25	2.25
Free	69.00%	46.00%	66.00%	48.00%	6.77	6.77
Reduced	3.00%	18.00%	3.00%	18.00%	1.27	1.27
Average						5.10%

school in the state of Virginia and are reported to the state department of education. The PALS was appropriate because it measures phonological awareness, rhyme, and concept of word for emergent readers, in addition to reading ability, creating a comprehensive profile of literacy achievement.

PALS subtests included a 20-word spelling test, a sight-word recognition test (children were given 2 seconds each to identify 20 consecutively flashed words on a computer screen), and a letter–sound identification test (children identified the sounds of 26 letters.) The 20-word spelling test was worth 44 points in the fall and 52 points at mid-year when slightly more complex words were tested. Each correctly spelled feature within a word was awarded 1 point, and spelling the entire word correctly was awarded 1 point. For example, the word “chin” could earn 2 points: 1 point for the “ch” plus 1 point for the entire word.

The district also mathematically combined all the PALS subtest results to create one *sum score* for each child. The sum score was calculated by adding together all points awarded on the spelling, sight word recognition, and letter–sound tests, while excluding

the text-level reading scores. The sum score evens out variation present when analyzing a child's spelling, sight word recognition, or letter-sound identification scores. A perfect score on the Fall PALS sum score was 90 points: 44 points for spelling 20 words correctly, 26 points for identifying all 26 letter sounds, plus 20 points for recognizing 20 sight words. A higher sum score suggests better early phonological skills.

The PALS assessment has a high classification accuracy with an area under the curve (AUC) of .91, meaning that it accurately diagnoses children as either at risk or not at risk of reading difficulty. Internal consistency for the PALS 1–3 is good with Cronbach's alpha scores ranging between .79 and .93. Interrater reliability has been tested at .98 to .99 (Huang et al. 2012).

Reading

The Developmental Reading Assessment, 2nd Edition (DRA2) was used to measure reading ability. The DRA2 measures students' ability to read fiction and nonfiction texts (Beaver 2006). The test includes subscores for accuracy, fluency, and comprehension on a passage of text that has been read by the child to confirm an independent reading level. Reading levels begin with 1, 2, 3, 4, and then rise by twos as 6, 8, 10, into the 20s, and then rise by fours beyond level 30 as 34, 38, etc. The DRA2 is administered in a one-on-one conference between the teacher and child. Children read a leveled text, while the teacher records errors and notations on an observation form with the same text. The student's errors are divided by the number of words read to determine a rate of accuracy, with no errors being 100% accuracy. Reading rate (fluency) is measured by timing the speed at which a student finishes a passage, for text levels 14 and above. Comprehension is measured by students' oral responses to questions (below level 28) or through students' written responses for text levels 28 and above. A student passes a particular text level (considered an independent reader at that text level) when their accuracy, fluency, and comprehension scores all exceed a stated benchmark score for each of those three constructs. Student reading achievement is indicated by the ability to independently read and comprehend each proceeding text level in the DRA2 continuum, which includes texts for K-8.

Reliability analyses performed for the DRA2 include internal consistency (.50–.80 reliabilities between fluency and comprehension), passage equivalency (MANOVA used to show no significant differences), test-retest reliability (correlation coefficients above .90), interrater reliability (66–72% agreement), and expert rater reliability (McCarty and Christ 2010). The DRA2 was tested for validity using criterion-related validity (no significant difference with other tests: with .60–.70 correlations), construct validity (low correlation at .41 across subtests), and predictive validity (teacher ratings with DRA2 scores: coefficient .60–.63) (McCarty and Christ 2010).

Data analysis

To examine differences in spelling, sight word recognition, and letter-sound identification over time by the Pre-K attendance group for research question 1, three 2-between

and 2-within repeated measures ANOVAs were conducted—one for each related dependent variable (PALS Summed Score, PALS Spelling Score, PALS Letter Sounds). Follow-up independent samples *t* tests between groups at each time point were conducted when ANOVA results indicated significant differences over time by group. This was done to better understand group differences. Question 2 was investigated by completing two independent samples *t* tests looking for differences in text level (PALS and DRA2) by the Pre-K attendance group.

Further, Hedges' *g* effect sizes (ES) for each outcome were calculated using the mean difference divided by weighted and pooled standard deviation (Hedges 1981). Using the additional weight is recommended when the treatment and comparison groups are significantly different in size (Ellis 2010). Hedges' *g* is a more conservative statistic, appropriate for wider generalization. The effect sizes were then converted into predicted percentile gains using a conversion chart.⁴ For example, a .35 ES predicts a 14% gain: a first grade student scoring in the 50th percentile without Pre-K attendance would be predicted to score in the 64th percentile if they had attended Pre-K.

Results

Research question 1

To answer the first research question (Is district Pre-K attendance associated with a significant difference in letter–sound identification and word identification during the beginning and middle of first grade?), between-subject testing revealed a significant difference in mean scores over time depending on group, with treatment (Pre-K attendees) being significantly higher than comparison (no preschool) on all three measures ($p < .001$). See Table 2 for descriptive and inferential results. Literacy scores grew from fall to mid-year of first grade at a similar rate for both groups.

To better understand differences between groups at each time point, independent samples *t* tests were conducted on each measure. Hedges' *g* effect sizes and comparable percentile gain and change estimates⁵ between the treatment and comparison groups are calculated for PALS literacy measures (Table 3).

Research question 2

To answer the second research question (Is district Pre-K attendance associated with a significant difference in text-level reading ability during the beginning and middle of first grade?), multiple independent samples *t* tests were conducted for text-level dependent variables by group (Pre-K vs. no preschool). An independent samples *t* test revealed that the average Fall PALS text level was significantly higher among the Pre-K group ($M = 3.86$, $SD = 2.10$) than among the no Pre-K group ($M = 3.07$, $SD = 1.87$), $t(1035) = 4.59$, $p < .001$. Another independent samples *t* test revealed that the average Mid-Year DRA2 independent text level was significantly higher among the Pre-K group ($M = 11.37$, $SD = 6.35$) than among the no Pre-K group ($M = 9.36$, $SD = 5.67$), $t(1036) = 3.88$, $p < .001$ (Table 4).

⁴ Marzano, R. & Pickering, D. (2011). "Conversion of effect size to percentile gain." The highly engaged classroom. Marzano Research Laboratory. [Reproducibles]. Retrieved from http://soltreemrls3.s3-website-us-west2.amazonaws.com/marzanoresearch.com/media/documents/reproducibles/highly_engaged/conversionofeffectsize.pdf.

⁵ Ibid.

Table 2 Means, standard deviations, and analysis of variance for the effects of Pre-K and no preschool attendance on three dependent variables

	Fall		Mid-year		<i>F</i>	<i>P</i>	η^2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
PALS sum score							
Treatment (<i>n</i> = 857)	58.78	15.01	68.71	15.80	21.16	.000	.020
Comparison (<i>n</i> = 171)	52.81	16.19	63.06	17.12			
PALS spelling							
Treatment (<i>n</i> = 859)	18.99	9.28	27.71	10.81	17.61	.000	.017
Comparison (<i>n</i> = 171)	15.37	8.81	24.66	10.54			
PALS letter sounds							
Treatment (<i>n</i> = 858)	23.40	3.04	25.03	1.61	9.57	.002	.009
Comparison (<i>n</i> = 170)	22.69	3.94	24.53	2.58			

Table 3 *T* tests, effect sizes, and predicted percentile gains for first grade literacy outcomes as a result of Pre-K attendance

Literacy outcome	<i>t</i>	<i>p</i>	<i>g ES</i> ^a	Predicted percentile gain	Predicted percentile change
Fall PALS sum score	4.76	.001	.398	16	50th–66th
Mid-year PALS sum score	3.99	.001	.332	13	50th–63rd
Fall PALS spelling	4.72	.001	.395	16	50th–66th
Mid-year PALS spelling	3.15	.002	.261	10	50th–60th
Fall PALS letter sounds	2.75	.006	.228	9	50th–59th
Mid-year PALS letter sounds	3.193	.001	.266	11	50th–61st

^a Hedges' *g* effect size

Table 4 Effect sizes and predicted percentile gain and change for first grade text-level reading ability related to Pre-K attendance

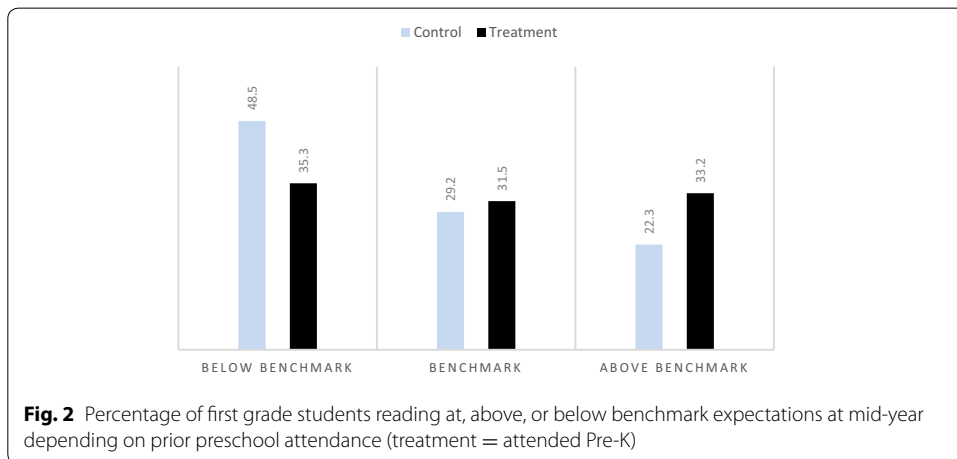
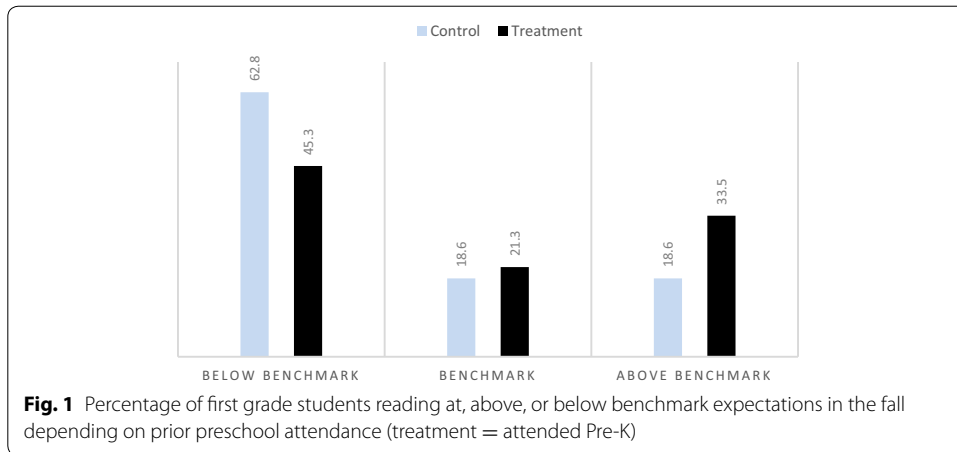
Literacy outcome	<i>g ES</i> ^a	Predicted percentile gain	Predicted percentile change
Fall PALS text level	.382	15	50th–65th
Mid-year DRA2 text level	.321	13	50th–63rd

^a Hedges' *g* effect size

Discussion

Children who attended district Pre-K began first grade reading nearly one full text level higher than the no-preschool group. Children who attended Pre-K are at less risk for reading difficulty than the children who did not attend any type of preschool. Children who did not attend any type of preschool began first grade reading nearly one text level below the expected benchmark. As a result, the no-preschool group would have been placed into below-benchmark reading groups in their classrooms and qualified for additional reading intervention more often than Pre-K attendees. Figure 1 shows fall reading scores aggregated by benchmark expectations: below benchmark, at benchmark, or above benchmark.

Figure 2 shows mid-year reading ability scores aggregated by benchmark expectations. Rather than showing a declining trend as text difficulty increases, as observed in the



no-preschool group, the Pre-K group shows a fairly even split across all three categories with about one-third of the students scoring below, at, and above benchmark. In fact, slightly more Pre-K students are reading above benchmark than on benchmark.

Identifying benchmark categories to help interpret student reading results overlooks the amount of growth an individual child has achieved regardless of category. Actual rate of growth over time (fall to mid-year) between the two groups appeared the same. However, children received up to twice as much literacy instruction in first grade after being identified as reading below grade level (intervention disproportionately included more comparison group children), so to say that the rate of growth between the two groups was equal, as the data imply, overlooks the fact that the amount and type of literacy instruction received in first grade were influenced by a student’s beginning first grade reading ability, which informed their placement into reading ability groups and determined whether or not they received additional intervention. If future studies could control for the number of minutes and type of literacy instruction provided, to address instructional differences based on reading ability groupings, then differences in reading trajectories as a result of Pre-K attendance could be more confidently identified.

The results of this study contribute meaningfully to the literature by suggesting that universal Pre-K, and not just criteria-selective Pre-K, can sustain literacy gains well into first grade when systematically pursued by a committed district. As mentioned previously, most evidence indicating that public preschool produces longer-term outcomes for children comes from criteria-selective programs (i.e., low income, learning disability, or other risk factors are required for admission). Since longer-term effects of universal public preschool have remained relatively unknown, policy-makers may have been reluctant to invest more significantly in universal preschool, choosing instead to maintain programs targeting at-risk populations, where the body of evidence clusters. The current findings suggest that universal public preschool is both viable and effective for all children, in addition to at-risk and minority children.

First grade children in the 2012–2013 cohort who previously attended the district's Pre-K program in 2010–2011 were predicted to score on average 9–16% higher than the no-preschool group in reading ability, spelling, sight word, and letter–sound identification with effect sizes ranging from .23 to .40. The average effect size across all eight first grade measures of reading and early literacy achievement is .35, translating into a 14% gain for the Pre-K group. When considering the focused approach to teaching literacy used by the Pre-K program in 2010–2011, these results may not be surprising. However, the magnitude of the effects is meaningful and their sustainability across the two time points is noteworthy.

It is also important to note that most children in the no-preschool group received additional literacy intervention in first grade. Despite additional reading intervention for the no-preschool group, the gains experienced by the Pre-K group continued to persist to the middle of first grade and the gap between the two groups narrowed very slightly. The district incurred a significant cost to provide later reading intervention to those children who did not attend preschool, yet the gap between the two groups persisted across the two time points. Universal Pre-K was effective in raising children's average first grade literacy scores to benchmark expectations, thus avoiding placement into additional reading intervention for the majority of Pre-K attendees. These findings suggest that Pre-K attendance not only places children on a benchmark trajectory for literacy achievement, but will also reduce reading intervention costs. Findings also revealed that twice as many Pre-K attendees were reading *above* expected first grade literacy benchmarks compared to the no-preschool group which is important because it implies that meeting benchmarks is not the only indicator of success. Many more children can and will perform *beyond* expectations when given an early start. Therefore, why would we limit public preschool access only to children with predetermined risk factors?

These results are more consistently significant than several related studies. Until now, the only other peer-reviewed study in Virginia was completed by Huang et al. (2012), a quasi-experimental study that also used the PALS literacy assessments to examine the Virginia Preschool Initiative (VPI). The main effect was not sustained through first grade but subgroup comparisons by race/ethnicity were significant. It is likely that the school district in the current study exerted more control over literacy instruction, as all preschool teachers attended the same professional development and implemented the same curriculum, compared to the State average represented in the VPI study. In another distinction, the current study examined a Pre-K program that did not have a low-income

criteria stipulation and is substantively universal (99% selection rate), whereas VPI uses income criteria, excluding a large population.

Other related studies revealed fluctuating first grade literacy achievement in relation to attending state-funded preschool. For example, Valenti and Tracey (2009) found significant outperformance in the middle but not the beginning of first grade and (Magnuson et al. 2007a) found significant effects in third grade but not first grade. In comparison, the current study finds meaningful and consistently significant results across the beginning and middle of first grade, for each literacy measure administered. This is important because it suggests that universal Pre-K contributes to more stable and consistently significant first grade literacy gains without hiding sleeper effects. As such, policy-makers may have more confidence that Pre-K attendance, as opposed to later literacy instruction or other factors, is responsible for the observed group differences.

The current study likely demonstrated better outcomes for first graders than previous research because (1) a truly no-treatment comparison group was secured and (2) a rigorous matching methodology was employed. In a randomized control trial of the Tennessee Voluntary Prekindergarten program, Lipsey et al. (2013) found no significant literacy or other differences at the end of first grade. However, forty-one percent of the children in the Tennessee study control group received other forms of care. The combination of a truly stay-at-home comparison group in the current study design, with the district's strong preschool literacy emphasis in 2010, likely contributed to the current study's significant first grade findings in contrast to the Tennessee results. This is significant because we need to know the effect of Pre-K attendance compared to no-preschool when considering if universal public preschool should be adopted, and this cannot be revealed if control or comparison groups include many children who attended other forms of preschool.

The current study also responds to the concerns raised by Maloffeva et al. (2007) that strong matched-pair designs have been mostly absent in the extant literature. Research into the Arkansas Better Chance and the New Jersey Abbot preschool programs have reported significant positive effects of public Pre-K on first grade literacy achievement (Frede et al. 2009; Hustedt et al. 2008; Jung et al. 2013), but these studies did not use matched-pair designs. Furthermore, in New Jersey it is unclear if the comparison group received other forms of preschool. The magnitude of the effects in the current study appear to be greater than those reported in Arkansas and similar to the New Jersey effects in terms of predicted percentile gains for first grade literacy following Pre-K attendance.

Recommendations for policy and practice

Worldwide, a deeper commitment has emerged to increase access to preschool, as seen in the sustainable development agenda (UNESCO 2015). "However, public preschool remains sparse globally and private providers proliferate" (Haslip and Gullo 2017, p. 11). Given the highly inequitable access to public preschool that exists worldwide, governments and policy-makers should consider the findings presented here as an additional impetus to expand public preschool. Likewise, it is recommended that policy-makers expand universal public preschool, rather than limiting public preschool to targeted or

income-dependent programs, because these findings show that even for children not at risk Pre-K makes a significant difference.

The present study examined a school district that provided a highly stable environment for children, as indicated in the very low teacher turnover rate among Pre-K teachers (2% in 2010), the presence of a certified teacher and trained assistant teacher in every Pre-K room, and parity between Pre-K and K-3 teacher salaries, benefits, and paid planning or professional development time, which is often not the case in state preschool programs in the US (Barnett et al. 2016). Therefore, it is recommended that policy-makers and school administrators establish a comparable support system by adopting similar structural features.

When considering all eight literacy measures together across the two time points (text-level reading ability, spelling, and the sum score of spelling, sight words, and letter-sound identification), this public Pre-K program in 2010–2011 had a meaningful impact on children's first grade literacy achievement in 2012–2013. Some might argue that improved literacy scores sustained into first grade is evidence of high Pre-K program quality and may therefore assume that replicating the Pre-K program described here is desirable. This study makes no such assumption. Data on important elements of quality were lacking to confidently evaluate the true or overall quality of the district Pre-K as administered in 2010–2011. For example, *process* quality measured by observing teacher-child interactions was lacking. We cannot claim or assume that developmentally appropriate practices were used to achieve the results reported. Nevertheless, the district Pre-K program had sufficient *structural* quality based on such indicators as teacher certification in early childhood education, appropriate teacher-child ratios, presence of assistant teachers, use of a formal literacy-based curriculum, and the adoption of state preschool objectives for learning.

Questions for future research include: Can these results be replicated or improved upon when utilizing a holistic, child-centered Pre-K curriculum supported by observed high-quality teacher-child interactions? Does a heightened focus on preschool literacy come at the expense of other cognitive and affective outcomes (e.g. mathematical thinking, scientific inquiry, creative expression, and social and emotional development)? Beyond literacy achievement, what other child outcomes are associated with universal Pre-K programs that use a holistic, child-centered curriculum supported by observed high-quality teacher-child interactions?

Limitations

Impact estimates could be inflated because a matched comparison group was used rather than random assignment, introducing unobserved selection bias. However, propensity score matching on a wide range of variables with a large sample size approximates group equivalence in the absence of random assignment (Stuart 2010), permitting a degree of causal inference. Nevertheless, certain confounding variables were unknown, such as parent's highest level of education (although this is related to SES) and home environment. Children's at-home literacy experiences and parent education influence literacy development (Yaden et al. 2000). Parents who tested and enrolled their child in public Pre-K may be different in other ways as well. For example, parents of Pre-K children may work more or be more assertive than the no-preschool group (Valenti and Tracey

2009). The comparison group could have been significantly more disadvantaged than the treatment group in ways that race, SES, and disability do not fully reveal. All available covariates were used in the matching process and the average absolute standardized difference between the two groups was reduced to 5% which is significantly below the 10% threshold proposed by Rosenbaum and Rubin (1983).

Results may not generalize to other cities, states, or regions in the United States or abroad because data were gathered from one mixed-urban school district. However, many other cities have similar characteristics with regard to diversity, income, and urban concentration, which may permit this study to serve as a model for similar districts considering universal Pre-K implementation. This study also accounts for some additional variance by reporting a more conservative, and therefore generalizable, effect size (Hedges' *g*).

Longitudinal research can present findings that quickly become irrelevant due to program change over time. However, this cohort of first grade children was enrolled in Pre-K in 2010–2011 when there was a significant focus on early literacy, which remains the case today in most public Pre-K programs around the U.S.

Finally, the literacy measures were administered by teachers introducing the possibility of teacher subjectivity. However, the DRA2 and PALS 1–3 tests have acceptable reliabilities and multiple measures of literacy were used across two time points in a relatively large sample ($n = 1056$) representative of all first grade teachers in the district. Results were also reported in the aggregate.

Conclusion

The results suggest that direct instruction in early literacy provided by a Pre-K program meeting recognized structural quality indicators continues to significantly and meaningfully impact children's first grade text-level reading ability, spelling, sight word, and letter–sound identification. On average, children attending a district-run universal public Pre-K program in a mixed-urban city in the United States were meeting expected literacy benchmarks in first grade, regardless of income or other risk factors, but children with no preschool experience were reading below expected benchmarks. Twice as many Pre-K attendees were also reading *above* expected benchmarks in first grade compared to students in the no-preschool group. This achievement gap persisted between the beginning and middle of first grade despite the no-preschool group receiving disproportionately more reading intervention. This study contributes to the literature by demonstrating that universal Pre-K attendance, examined using a methodologically rigorous matching method and a truly no-treatment comparison group, does have a significant and sustained impact across multiple measures of literacy achievement well into first grade.

Acknowledgements

The author wishes to thank Ning Rui and Toni Sondergeld for their helpful assistance. The reviewers also provided excellent feedback.

Competing interests

Not applicable.

Availability of data and materials

Raw data are held by the author.

Consent for publication

Not applicable.

Ethics approval and consent to participate

This study was reviewed and approved by the Institutional Review Board at Old Dominion University.

Funding

Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 27 July 2017 Accepted: 15 December 2017

Published online: 10 January 2018

References

- Barnett, W. S. (2010). Universal and targeted approaches to preschool education in the United States. *International Journal of Child Care and Education Policy*, 4(1), 1–12.
- Barnett, W. S., Carolan, M. E., Fitzgerald, J., & Squires, J. H. (2011). *The state of preschool 2011: State preschool yearbook*. New Brunswick: National Institute for Early Education Research.
- Barnett, W. S., Epstein, D. J., Friedman, A. H., Boyd, J. S., & Hustedt, J. T. (2008). *The state of preschool 2008: State preschool yearbook*. New Brunswick: National Institute for Early Education Research, Rutgers University.
- Barnett, W. S., Friedman-Krauss, A. H., Gomez, R. E., Horowitz, M., Weisenfeld, G. G., & Squires, J. H. (2016). *The state of preschool 2015: State preschool yearbook*. New Brunswick: National Institute for Early Education Research.
- Beaver, J. (2006). *Developmental reading assessment, second edition (DRA2)*. Lebanon: Pearson Learning.
- Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *The Teachers College Record*, 112(3), 579–620.
- Ellis, P.D. (2010). *Effect size FAQs*. Retrieved June 26, 2013, from <http://www.effectsizefaq.com>.
- Farran, D., & Lipsey, M. (2015). Expectations of sustained effects from scaled up pre-K: Challenges from the Tennessee study. *Evidence Speaks Reports*, 1(3). https://www.brookings.edu/wp-content/uploads/2016/06/Expectations-of-sustained-effects-from-scaled-up-preK-Tennessee-study_4.pdf.
- Frede, E., & Barnett, W. S. (2011). Why Pre-K is critical to closing the achievement gap. *Principal*, 90(5), 8–11.
- Frede, E., Jung, K., Barnett, W. S., & Figueras, A. (2009). *The APPLES blossom: Abbott preschool program longitudinal effects study: Preliminary results through 2nd grade*. Interim report. New Brunswick: National Institute for Education Research, Rutgers University. Retrieved from http://nieer.org/pdf/apples_second_grade_results.pdf.
- Hall, N. R. (1987). *The emergence of literacy*. Portsmouth: Heinemann Educational Books.
- Haslip, M. J., & Gullo, D. (2017). The changing landscape of early childhood education: Implications for policy and practice. *Early Childhood Education Journal*. <https://doi.org/10.1007/s10643-017-0865-7>.
- Heckman, J. J., & Masterov, D. V. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy*, 29(3), 446–493. <https://doi.org/10.1111/j.1467-9353.2007.00359.x>.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 106–128.
- Hernandez, D. J. (2012). PreK-3rd: Next Steps for State Longitudinal Data Systems. Policy to Action Brief (8). *Foundation for Child Development*.
- Hill, C. J., Gormley, W. T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly*, 32, 60–79.
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8), 1–28. <https://doi.org/10.18637/jss.v042.i08>.
- Huang, F. L., Invernizzi, M. A., & Drake, E. A. (2012). The differential effects of preschool: Evidence from Virginia. *Early Childhood Research Quarterly*, 27, 33–45. <https://doi.org/10.1016/j.ecresq.2011.03.006>.
- Hustedt, J. T., Barnett, W. S., & Jung, K. (2008). *Longitudinal Effects of the Arkansas Better Chance Program: Findings from Kindergarten and First Grade*. New Brunswick: National Institute for Early Education Research, Rutgers University.
- Invernizzi, M., Meier, J., & Juel, C. (2003). *PALS: Phonological Awareness Literacy Screening*. Charlottesville: University Printing Services.
- Jung, K., Barnett, W., Hustedt, J. T., & Francis, J. (2013). *Longitudinal effects of the Arkansas Better Chance program: Findings from first grade through fourth grade*. New Brunswick: National Institute for Early Education Research.
- Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences*, 103(27), 10155–10162.
- 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 (Research Report). Nashville, TN: Vanderbilt University, Peabody Research Institute. https://my.vanderbilt.edu/tnpreevaluation/files/2013/10/August2013_PRI_Kand1stFollowup_TN-VPK_RCT_ProjectResults_FullReport1.pdf.
- Lipsey, M. W., Farran, D. C., & Hofer, K. G. (2015). A randomized control trial of the effects of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade (Research Report). Nashville, TN: Vanderbilt University, Peabody Research Institute. https://my.vanderbilt.edu/tnpreevaluation/files/2013/10/NPKthrough3rd_final_withcover.pdf.

- Luthar, S. S. (2003). *Resilience and vulnerability: Adaptation in the context of childhood adversities*. Cambridge: Cambridge University Press.
- Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007a). The persistence of preschool effects: Do subsequent classroom experiences matter? *Early Childhood Research Quarterly*, 22(1), 18–38.
- Magnuson, K. A., Ruhm, C., & Waldfogel, J. (2007b). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, 26(1), 33–51. <https://doi.org/10.1016/j.econedurev.2005.09.008>.
- Maloffeva, E., Daniel-Echols, M., & Xiang, Z. (2007). Findings from the Michigan School Readiness program 6–8 follow up study. *High/Scope Educational Research Foundation*.
- Mashburn, A. J., Justice, L., Downer, J. T., & Pianta, R. C. (2009). Peer effects on children's language achievement during pre-kindergarten. *Child Development*, 80(3), 686–702.
- McCarty, A. M., & Christ, T. J. (2010). Test review: Beaver, JM, & Carter, MA (2006). The Developmental Reading Assessment—Second Edition (DRA2). *Assessment for Effective Intervention*, 35(3), 182–185.
- Moore, J. E., Cooper, B. R., Domitrovich, C. E., Morgan, N. R., Cleveland, M. J., Shah, H., et al. (2015). The effects of exposure to an enhanced preschool program on the social-emotional functioning of at-risk children. *Early Childhood Research Quarterly*, 32, 127–138.
- Pianta, R. C., Barnett, W. S., Burchinal, M., & Thornburg, K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and what we need to know. *Psychological Science in the Public Interest*, 10(2), 49–88. <https://doi.org/10.1177/1529100610381908>.
- Reynolds, A. J., Magnuson, K. A., & Ou, S. R. (2010). Preschool-to-third grade programs and practices: A review of research. *Children and Youth Services Review*, 32(8), 1121–1131. <https://doi.org/10.1016/j.childyouth.2009.10.017>.
- Rosenbaum, P., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 604–620.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1–21. <https://doi.org/10.1214/09-STS313>.
- Stuart, E., & Rubin, D. B. (2004). *Matching methods for causal inference: Designing observational studies*. Cambridge: Harvard University Department of Statistics.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2011). Pre-school quality and educational outcomes at age 11: Low quality has little benefit. *Journal of Early Childhood Research*, 9(2), 109–124. <https://doi.org/10.1177/1476718X10387900>.
- Temple, J., Arteaga, I., & Reynolds, A. (2010). *Longer-term effects of preschool for urban children: Results from the Chicago Longitudinal Study*. Minneapolis: University of Minnesota.
- Tracey, D. H., & Morrow, L. M. (2006). *Lenses on reading: An introduction to theories and models*. New York: Guilford.
- UNESCO. (2015). *Global Monitoring Report 2015: Education for All 2000–2015. Achievements and Challenges*. Paris: United Nations Education, Science and Cultural Organization.
- Valenti, J. E., & Tracey, D. H. (2009). Full-day, half-day, and no preschool: Effects on urban children. *Education and Urban Society*, 41(6), 17. <https://doi.org/10.1177/0013124509336060>.
- Yaden, D., Rowe, D., & MacGillivray, L. (2000). Emergent literacy: A matter (polyphony) of perspectives. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 425–454). Mahwah: Lawrence Erlbaum.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
