

University of Arkansas, Fayetteville

ScholarWorks@UARK

---

Education Reform Faculty and Graduate  
Students Publications

Education Reform

---

2-16-2017

## Is Pre-Kindergarten an Educational Panacea? A Systematic Review and Meta-Analysis of Scaled-Up Pre-Kindergarten in the United States

Corey A. DeAngelis

University of Arkansas, Fayetteville, cadeange@uark.edu

Heidi Holmes Erickson

University of Arkansas, Fayetteville

Gary W. Ritter

University of Arkansas, Fayetteville

Follow this and additional works at: <https://scholarworks.uark.edu/edrepub>



Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Educational Leadership Commons](#), and the [Other Educational Administration and Supervision Commons](#)

---

### Citation

DeAngelis, C. A., Holmes Erickson, H., & Ritter, G. W. (2017). Is Pre-Kindergarten an Educational Panacea? A Systematic Review and Meta-Analysis of Scaled-Up Pre-Kindergarten in the United States. *Education Reform Faculty and Graduate Students Publications*. Retrieved from <https://scholarworks.uark.edu/edrepub/12>

This Article is brought to you for free and open access by the Education Reform at ScholarWorks@UARK. It has been accepted for inclusion in Education Reform Faculty and Graduate Students Publications by an authorized administrator of ScholarWorks@UARK. For more information, please contact [scholar@uark.edu](mailto:scholar@uark.edu).



UNIVERSITY OF  
ARKANSAS

College of Education & Health Professions  
*Education Reform*

## **WORKING PAPER SERIES**

### **Is Pre-Kindergarten an Educational Panacea? A Systematic Review and Meta-Analysis of Scaled-Up Pre-Kindergarten in the United States**

Corey A. DeAngelis

Heidi Holmes Erickson

Gary W. Ritter, Ph.D.

February 16, 2017

EDRE Working Paper 2017-08

The University of Arkansas, Department of Education Reform (EDRE) working paper series is intended to widely disseminate and make easily accessible the results of EDRE faculty and students' latest findings. The Working Papers in this series have not undergone peer review or been edited by the University of Arkansas. The working papers are widely available, to encourage discussion and input from the research community before publication in a formal, peer reviewed journal. Unless otherwise indicated, working papers can be cited without permission of the author so long as the source is clearly referred to as an EDRE working paper.

**Is Pre-Kindergarten an Educational Panacea? A Systematic Review and Meta-Analysis of Scaled-Up Pre-Kindergarten in the United States**

Corey A. DeAngelis  
Department of Education Reform,  
University of Arkansas  
[cadeange@uark.edu](mailto:cadeange@uark.edu)

Heidi Holmes Erickson  
Department of Education Reform,  
University of Arkansas  
[hh018@email.uark.edu](mailto:hh018@email.uark.edu)

Gary W. Ritter, Ph.D.  
Department of Education Reform,  
University of Arkansas  
[garyr@uark.edu](mailto:garyr@uark.edu)

February 16, 2017

**Acknowledgements**

The content of the report is solely the responsibility of the authors and does not necessarily represent the views of the University of Arkansas. Corresponding author is Corey A. DeAngelis, [cadeange@email.uark.edu](mailto:cadeange@email.uark.edu).

### **Abstract**

We synthesize the existing research and compute meta-analytic averages for the effects of scaled-up, publicly funded pre-kindergarten (pre-K) programs on student pre-kindergarten achievement in math and reading. Other systematic reviews of pre-K programs have focused on the effects for specific groups of students from various types of pre-K programs. We add to the literature by focusing on scaled-up pre-K often provided at the state level, which is of growing policy interest. Scaled-up programs are large state or district run programs that are available to a large portion of children before they enter kindergarten. We limit our analysis to state and districtwide pre-K programs in the United States from 2000 through 2016. In order to obtain the most accurate effect estimates, we restrict our analysis to experimental and quasi-experimental research designs with the highest internal validity.

We synthesize the short-term cognitive effects of pre-K and find large positive effects of scaled-up public pre-K programs on student pre-kindergarten test scores. In particular, we find that the overall effect on math scores is over a third of a standard deviation and the overall effect on reading scores is three-fifths of a standard deviation. This review is restricted to studies focused on short term results of pre-K programs; our search uncovered only one study rigorously assessing the impacts of scaled-up pre-K programs on student achievement after kindergarten. More research is needed on the sustained effects of pre-K as policymakers debate whether to expand or adopt such programs.

*Keywords:* early child education; public program evaluation; prekindergarten; preschool; systematic review; meta-analysis

### Introduction

Publicly-funded pre-kindergarten (pre-K) has been a highly politicized topic for many years. President Barack Obama<sup>1</sup> and many state governors have called for expanding preschool programs to all three and four-year-olds. Commonly cited goals of publicly funded pre-K programs are to reduce the considerable academic achievement gaps between advantaged and disadvantaged students and to provide access to quality early childhood education for all families (Slaby, Loucks, & Stelwagon, 2005). Although racial academic achievement gaps have declined since the 1970s, they are still present in the United States. In 2012, the white-black and white-Hispanic achievement gap on the National Assessment of Educational Progress (NAEP) was estimated to be between 50% and 90% of a standard deviation (Racial and Ethnic Achievement Gaps, 2016).

Two of the most well-known pre-kindergarten programs, the Perry Preschool Project and the Abecedarian Project, had significant positive effects for student academic achievement and other social benefits such as reduced crime rates and lower teen pregnancies (Campbell et al., 2002; Schweinhart, 2003; Schweinhart et al., 2005; Washington State Institute for Public Policy, 2013). Both of these programs were targeted to disadvantaged children, and several studies of these programs found a large return on initial investment (Schweinhart, 2003; Barnett & Masse, 2007). Others, however, have urged caution in generalizing the evaluation findings for these programs and have warned that these are “Cadillac” programs because their high costs and intensive services make them difficult to replicate on a large scale. For example, children in the Abecedarian Program received full time, high quality childcare with instruction and activities individualized to each child from birth to five years of age. Perry Preschool was a similar

---

<sup>1</sup> <https://obamawhitehouse.archives.gov/the-press-office/2014/10/02/remarks-president-economy-northwestern-university%22>

intervention; however, it was only available to three and four-year-olds. Nevertheless, while these findings are interesting, they are likely not as policy-relevant as many pre-K advocates would like to believe since most public pre-K programs are quite different from the Perry Preschool and Abecedarian Program models in cost and size.

Head Start is another commonly cited program that shows positive effects for student outcomes in the short-run (Puma et al., 2010; U.S. Department of Health & Human Services, 2010; Puma et al., 2012). Head Start was enacted by the U.S. government as part of President Lyndon B. Johnson's War on Poverty. Head Start has a preschool component, but it is a comprehensive program aimed at helping low-income families with early childhood care. Head Start also includes nutrition, health, and parental involvement services for families whose children are enrolled. While Head Start is a federal program, some state pre-K programs work in conjunction with Head Start to expand access to families in the state.

There are also many state and district pre-K programs. Forty-two percent of four-year-olds attend publicly funded pre-K, with 28% in state or locally funded pre-K, and 11% in Head Start (Yoshikawa et al., 2013). State and local programs vary in accessibility. Some programs are only available to disadvantaged families, while others are available to all families regardless of income level.

Barnett (2008) and Yoshikawa et al. (2013) review the literature on publicly funded pre-K and find positive effects for many student outcomes. The majority of the programs they review are targeted or semi-universal. Other systematic reviews focus primarily on the impacts that targeted pre-K programs have on student achievement (Gilliam & Ziegler 2001; Anderson et al. 2003; Brown & Scott-Little 2003; Camilli et al., 2010) or the effects that different programs, such as mentoring and certain curricula, within preschools have on student achievement

(Chambers et al., 2010). The previous systematic reviews on targeted preschool programs have found significantly large effects sizes on student academic outcomes; however, they do not limit their searches to experimental or quasi-experimental studies of universally available programs.

There is a smaller, but significant, section of literature that examines the effects of scaled-up public pre-K programs. There have been strong positive achievement effects found in evaluations of the Tulsa, Oklahoma and Boston, Massachusetts pre-K programs (Gormley and Gayer, 2005; Weiland and Yoshikawa, 2013), but mixed to negative achievement results in an experimental evaluation of Tennessee's pre-K program (Lipsey et al., 2013; Swain, Springer, & Hofer, 2015). Currently, no reviews of scaled-up pre-K literature exist, despite the increased focus on expanding access to preschool. We add to the literature by systematically reviewing the research on the effects of scaled-up public pre-K programs on student kindergarten math and reading test scores. We also present meta-analytic averages of the effects on math and reading achievement.

## **Methodology**

### **Review Strategy and Inclusion Criteria**

We established a set of inclusion criteria to guide our review. Setting criteria prior to searching the literature helped narrow our focus and facilitated an unbiased review of the current state of the literature. These criteria, listed below, reflect the overall goal that we had for the review: to evaluate what recent methodologically rigorous studies conclude about the academic impacts of state and districtwide scaled-up public pre-kindergarten programs on student kindergarten achievement.

We use the following inclusion criteria for our review:

- Quantitative studies using experimental or quasi-experimental<sup>2</sup> designs such as: Randomized Control Trials (RCTs), Randomized Field Trials (RFTs) and Regression-Discontinuity (RDD).
- Time-relevant, from 2000 to present
- Include a math or reading related outcome measure with at least one relevant effect size listed or include information to calculate an effect size
- Do not report solely on subgroup findings
- Published or unpublished studies
- Programs in the United States

The following are the types of studies that we exclude:

- Non-rigorous quantitative studies such as: pre-post and observational studies
- Qualitative Studies such as: case studies, journalistic accounts, and general inquiries
- Theoretical analyses
- Non-English language papers
- Pre-k programs outside of the United States

Using our criteria, we searched relevant databases such as Ebsco and ProQuest. All databases were accessed through the University of Arkansas Library.

The systematic review of database results was conducted in four stages. First, we reviewed titles; second, we reviewed abstracts; third, we reviewed the methods in each paper; last, we reviewed the entire article. At each stage, two reviewers evaluated the articles and coded whether the articles met the inclusion criteria and were thus kept for further review. We used a cautious approach and only discarded a study if both reviewers chose to discard. If there was disagreement between the reviewers at any stage of coding, the article was kept and moved forward to the subsequent review stage. This was a conservative review approach to systematically avoid incorrectly excluding any relevant study from consideration. For example,

---

<sup>2</sup> We define a quasi-experiment as a type of study that controls for unobservable characteristics through the research design.



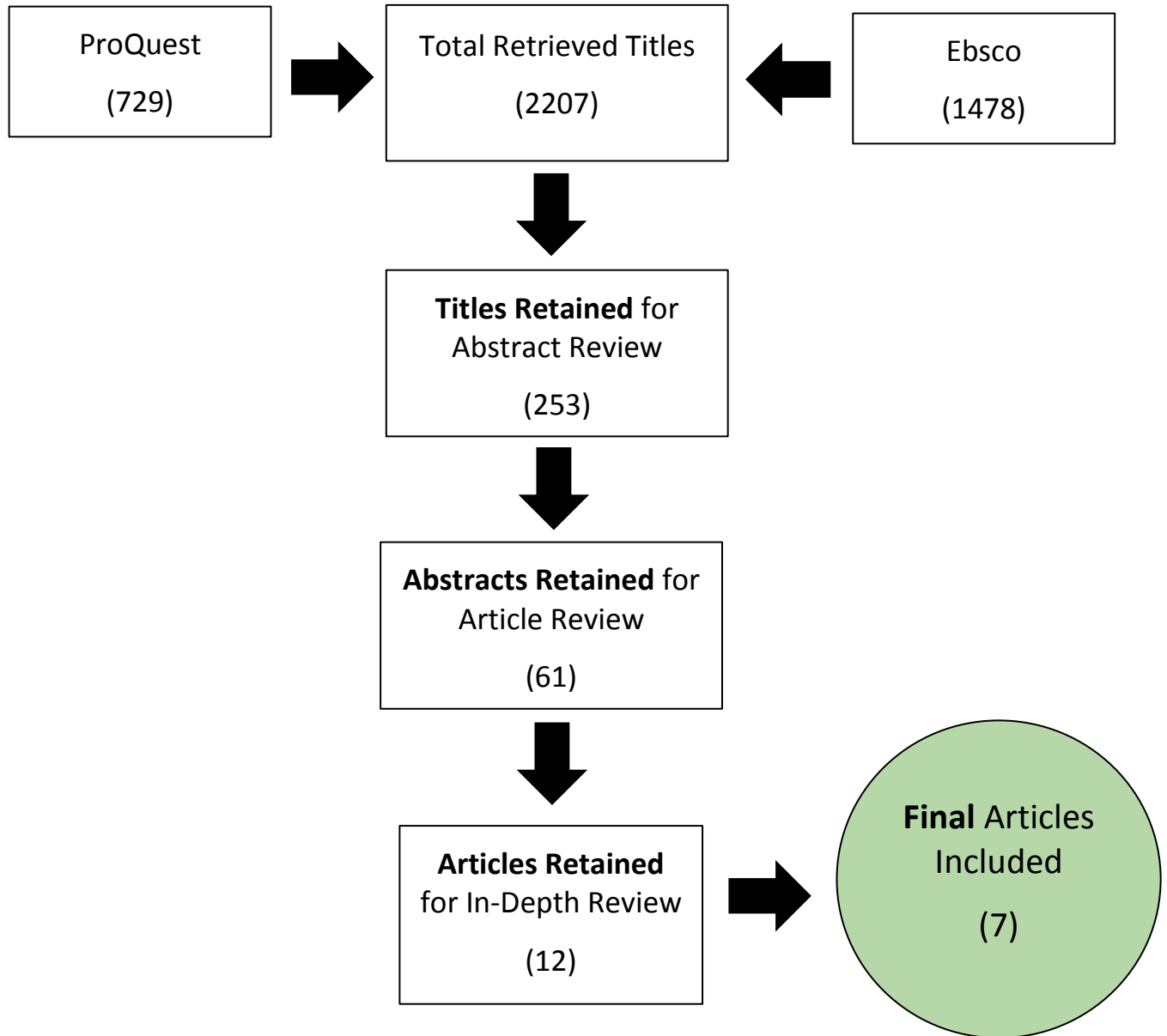
if both reviewers coded a title as meeting the inclusion criteria, the abstract was then reviewed. If the reviewers coded the abstract as meeting the inclusion criteria, then its methods were evaluated. This process was followed until all included articles were fully reviewed.

### **Results of Searches**

The initial Ebsco database search yielded 2,478 results with 1,478 unique articles, after duplicate titles were removed. After each of the 1,478 article titles were independently reviewed, we kept 142 study abstracts. Based on the review of the 142 abstracts, we retained 24 studies. After the review of the full articles, we kept four unique studies from the Ebsco database.

We further restricted our search in ProQuest because of the multiple irrelevant studies found in the Ebsco search. We excluded medical journals and specific foreign journals since none of the articles in such journals focused on evaluation of a pre-K program in the United States. The initial ProQuest database search resulted in 729 unique articles. After the initial review of the 729 titles, we retained 111 studies. Based on the review of the 111 abstracts, we kept a total of 37 studies. After the review of the full articles, we retained three unique studies from the ProQuest database. Figure 1 below shows a graphic display of our search and review process. For specific search terms refer to the Appendix.

**Figure 1: Graphic Display of the Search Results**



Articles were excluded for various reasons throughout each step of the review process. Many studies were excluded because they did not meet the inclusion criteria for relevance. For example, the following studies were excluded because they evaluated programs outside of the United States, Head Start, or specific aspects of pre-K programs:

- “Can We Improve Preschool Classroom Quality in Chile? A Cluster-Randomized Trial Evaluation of a Professional Development Program”
- “Do the Effects of Head Start Vary by Parental Pre Academic Stimulation?”
- “Effects of Coaching on Educators' and Preschoolers' Use of References to Print and Phonological Awareness during a Small-Group Craft/Writing Activity”
- “Effects of a Tablet-Based Mathematics Application for Pre-School Children”

Other articles were excluded because they were either descriptive in nature (Fram, Kim & Sinha, 2012; Whittenberg, 2013) or examined the effects of different pre-K programs in comparison to each other (Landry et al., 2009):

- Fram, M. S., Kim, J., & Sinha, S. (2012). Early care and prekindergarten care as influences on school readiness. *Journal of Family Issues*, 33(4), 478-505.
- Landry, S. H., Assel, M. A., Swank, P. R., & Anthony, J. L. (2009). An Experimental Study Evaluating a State Funded Pre-Kindergarten Program: Bringing Together Subsidized Childcare, Public School, and Head Start. *Society for Research on Educational Effectiveness*.
- Whittenberg, J. D. (2013). Brigance, reading scores, and student preschool participation: Predictors of future academic achievement.

Another common reason for excluding a study from the analysis was a failure to establish baseline equivalence of the outcome measure of interest. A majority of these studies were either simple before and after comparisons or observational ordinary least squares (OLS) regressions with control variables. Since observational studies such as these do not use research methods that can remove bias caused by unobservable characteristics, we excluded them. Selection bias is especially important in this type of program evaluation, since pre-K attendance is voluntary, and observational studies, by design, cannot distinguish program effects from the effects of pre-

existing differences between those who selected into the program and those who did not.

Explicitly, it is especially difficult for studies that use OLS regression to establish baseline equivalence of the outcome measure (test scores) for this particular research question. The main reason is that students do not have a record of test scores prior to entering pre-K.

The articles below were excluded because they did not meet the inclusion criteria for methodological rigor (Andrews, Jargowsky & Kuhne, 2012; Fitzpatrick, 2008) or did not report start of kindergarten academic effects of a pre-K program (Weiland, Eidelman & Yoshikawa, 2011). Artz and Welsch (2016) used a rigorous district fixed-effects analytic method; however, their unit of analysis was districts rather than students. We excluded non-student-level analyses since student-level data provide the most accurate estimates of the effects of pre-K on students.

- Andrews, R. J., Jargowsky, P., & Kuhne, K. (2012). The Effects of Texas's Targeted Pre-Kindergarten Program on Academic Performance (No. w18598). National Bureau of Economic Research.
- Artz, B., & Welsch D. (2016). The Impact of Publicly Provided Early Childhood Education Programs on District-Level Test Scores. *Contemporary Economic Policy*, 34(1), 89-106.
- Fitzpatrick, M. D. (2008). Starting school at four: The effect of universal pre-Kindergarten on children's academic achievement. *The BE Journal of Economic Analysis & Policy*, 8(1).
- Weiland, C., Eidelman, H., & Yoshikawa, H. (2011). A Regression Discontinuity Analysis of the Impact of “Building Blocks” in an Urban Public Prekindergarten Program and Associations between Fidelity-to-Curriculum and Child Outcomes. Society for Research on Educational Effectiveness.

Other articles were excluded because they were previous versions of updated studies already included in our review:

- Barnett, W. S., Lamy, C., & Jung, K. (2005). The effects of state prekindergarten programs on young children's school readiness in five states. New Brunswick, NJ: National Institute for Early Education Research.
- Weiland, C., & Yoshikawa, H. (2011). The Impact of an Urban Universal Public Prekindergarten Program on Children's Early Numeracy, Language, Literacy, and Executive Function Outcomes. Society for Research on Educational Effectiveness.

The most common analytic strategy employed in the included studies was the regression-discontinuity design (RDD). Most pre-K programs have a strict age cutoff determined by students' birthdays, generally prior to September 1st. The RDD is a rigorous quasi-experimental design used to identify causality since it is almost random chance that a student is born just before or just after the cutoff date that strictly determines program treatment. Studies report bandwidths from as large as 360 days around the cutoff to as low as a few days before and after the cutoff, with some studies reporting effect estimates at multiple bandwidths. RD designs are most useful if the sample sizes are large enough so that a narrow bandwidth around the cutoff date may be employed. However, in many studies without very large sample sizes, researchers are hesitant to limit the bandwidth because of the resulting decreases in sample size, and decreases in the power to detect a program effect. We address these issues in our results section below.

### **Summaries of Included Pre-K Programs**

Scaled-up pre-K programs vary by design and size, which can have a significant effect on the outcomes of each program. For this reason, we include a brief description of each program included in the final set of studies. There are ten different programs examined in the seven included studies (Gormley & Gayer, 2005; Wong et al, 2008; Lipsey et al., 2013; Bartik, 2013; Weiland and Yoshikawa, 2013; Hustedt et al, 2015; Hustedt et al, 2010). One study covers multiple programs and there are multiple studies that examine the same program (Table 3). However, we do not double count any studies that use the same data. A further discussion of studies included in our vote count and meta-analysis is provided in the results section.

Many of the included pre-K programs have similar characteristics. All programs are available for all four-year-olds, with only the New Jersey Abbott Preschool program and the

Arkansas Better Chance program available to three and four-year-olds. A majority of programs require teachers to be certified or have qualifications in early childhood education. Most programs also have a class size maximum of 20 students. A handful of the programs give preference to low income and at risk children and then open remaining seats to all other students. Data were not available for some of the program details. Table 1 below gives a description of each program's characteristics.

Many programs also used the same tests for their outcome measures. The most common tests used were the Peabody Picture Vocabulary Test (PPVT) and the Woodcock-Johnson (WJ), which has subsections such as: Letter-Word ID, Print Awareness, Comprehension, Applied Problems, Quantitative Concepts, and Calculation. Tennessee Voluntary Pre-K is the only program that does not report PPVT test scores. Table 2 below details what outcome measures each program uses. Following Table 2, we discuss in more detail the eight programs included in the review.

**Table 1: Pre-K Program Characteristics**

<b>Program</b>	<b>Level of Administration</b>	<b>Ages Served</b>	<b>Class Size</b>	<b>Teacher Qualifications</b>	<b>Enrollment Preferences</b>	<b>Students Served Per Year</b>	<b>Cost Per Student</b>	<b>Year Implemented</b>
Arkansas Better Chance Program (ABC)	State	3-4	20 max	Teacher w/BA, w/ assistant teachers	Yes	37% of 4-yr olds, 10% of 3-yr olds (2015)	NA	1991
Boston Public School Pre-K	District	4	NA	BA, MA w/in 5 yrs. teaching	No	34% of 4-yr olds (2008-09)	NA	NA
Kalamazoo, MI Ready 4s	County	4	requires <i>quality standards</i>		Yes	NA	\$4,500 max	2010
Michigan School Readiness Program (MSRP)	State	4	18	BA degree	Yes	19% of 4yr olds (2004-05)	\$5,000	1985
New Jersey Abbott Preschool	District	3-4	15 avg.	Certification plus aid	Yes	79% of Abbot 3-4yrs olds (2004-05)	\$10,361	1999
New Mexico Pre-K	State	4	20 max	No license, working toward	No	4,745 (2008-09)	\$3,124	2005
Oklahoma State Universal Preschool	State	4	20 max	BA with Certification	No	65% of 4 yr olds (2004-05)	\$6,100	1998
South Carolina's Early Childhood Programs	State	4	20	BA with training in early education	Yes	32% of 4yr olds (2004-05)	\$3,219	1984
Tennessee Voluntary Pre-K	State	4	20 max	Certification	Yes	NA	NA	2005
West Virginia Universal Pre-K	State	4	20 max	Certification	No	33% of 4yr olds (2004-05)	\$6,829	2002

*Notes: "NA" indicates that the information for the program was not publicly-available. Details are found in the respective papers description of the programs.*

**Table 2: Outcome Measures**

<b>Program</b>	<b>Woodcock-Johnson</b>	<b>PPVT Vocabulary</b>	<b>Pre-CTOPP Print Awareness</b>	<b>Other</b>
Arkansas Better Change (ABC)	Yes	Yes	Yes	No
Boston Public School Pre-K	Yes	Yes	No	REMA
Kalamazoo, MI Ready 4s	Yes	Yes	Yes	No
Michigan School Readiness Program	Yes	Yes	Yes	No
New Jersey Abbott Preschool	Yes	Yes	Yes	No
New Mexico Pre-K	Yes	Yes	No	Early Literacy Skills Assessment
Oklahoma State Universal Preschool	Yes	Yes	Yes	ECSI
South Carolina's Early Childhood Programs	No	Yes	Yes	No
Tennessee Voluntary Pre-K	Yes	No	No	Yes
West Virginia Universal Pre-K	Yes	Yes	Yes	No



**Arkansas Better Chance Program (ABC)**

The ABC program, enacted in 1991, is a statewide pre-K program available to at risk three and four-year-olds. The program is targeted to at risk children. Eligibility is primarily based on family income; students from households with income below 200% of the federal poverty line are given priority. Children in foster care, non-native English speakers, and those who have been in abusive situations are also eligible. The state contributes money and local governments match state funds. In 2015, 37% of Arkansas four-year-olds and 10% of three-year-olds were enrolled in the program. The program is primarily provided in public schools, but one-third of the students enrolled attend a nonpublic setting such as Head Start or private institutions. Arkansas requires that the lead teacher (a teacher over three classrooms) has a bachelor's degree and assistant teachers have an associate's degree. Class sizes cannot exceed 20 students and the staff-child ratio cannot exceed 10:1 (Hustedt et al., 2015). Hustedt et al. (2015) evaluated the effect of the ABC program on kindergarten test scores using an RDD.

**Boston Public School Pre-K Program**

Boston pre-K is universally open to all four-year-olds living in Boston. This is a full-day program run entirely in public schools. All schools use two very specific curricula: Opening the World of Learning (OWL) and Building Blocks. Teachers must have a bachelor's degree and are required to receive a master's degree within five years of teaching pre-K. They operate on the same pay schedule as public K-12 teachers and receive professional development and assistance from a curriculum coach. Weiland and Yoshikawa (2013) evaluated the effects of Boston public pre-K on kindergarten entrant tests using a RDD.

**Michigan School Readiness Program (MSRP)**

The Michigan School Readiness Program is for at-risk four-year-olds. Students must meet an income criteria as well as one additional risk factor from a list of 25. The program is half-day and lasts for at least 30 weeks per year. The program is provided through public, private, and Head Start facilities. The combined estimated spending from federal, state, and local sources is \$5,000 per student. Wong et al. (2008) evaluated the effects of MSRP on kindergarten entrant test scores from 2004-2005 using an RDD.

**New Jersey Abbott Preschool Program**

The Abbott Preschool Program was founded in 1999 and is administered and funded by the New Jersey Department of Education and Department of Human Services. It is universally available to all three and four-year-old students living in economically disadvantaged school districts. Low income students who live outside the district are also eligible to apply. It is a six-hour, 180-day program that also offers afterschool and summer courses. Classes have around 15 students and are staffed with a certified teacher and an assistant. The program costs approximately \$10,361 per student (Wong et al., 2008). Wong et al. (2008) evaluated the effects of the Abbott program on kindergarten entrant test scores from 2004-2005 using an RDD.

**New Mexico Pre-K Program**

New Mexico Pre-K is a universal statewide program. All four-year-olds are eligible for the program. Providers can be public or private, including Head Start contributors, faith-based centers, and universities. It is primarily funded by the state with some additional partnerships. Teachers are not required to be licensed when they start teaching, but they must be working towards a bachelor's degree and an early childhood education license. The class size cannot

exceed 20 students and the child to staff ratio cannot exceed 10:1. Hustedt et al. (2010) evaluated the effect of the New Mexico Pre-K program on kindergarten entrance tests using a RDD.

### **Ready 4s Pre-K Program: Kalamazoo, Michigan**

Ready 4s is a county-wide pre-K program that works with public and private providers. Families making below \$100,000 are eligible for tuition assistance. The amount of tuition assistance is based on the family income, with the highest amount at \$4,500 for the least advantaged families. The program is primarily funded through private donations from foundations, businesses, and individuals. Bartik (2013) evaluated the effect of the Ready 4s program on kindergarten entrant test scores using an RDD.

### **South Carolina's Early Childhood Program**

South Carolina has two pre-K programs: the Half-Day Child Development Program and the First Steps School Readiness initiative. Program eligibility is determined by the local district, but the state provides a list of risk factors such as household income levels at or below the poverty line. Public schools are the primary provider of the program, but other private and Head Start providers deliver services. The program lasts through the academic year and is half-day for five days a week. Funding from federal, state, and local sources is estimated around \$3,219 per student (Wong et al., 2008). Wong et al. (2008) evaluated the effects of South Carolina's pre-K program on kindergarten entrant test scores from 2004-2005 using an RDD.

### **Tennessee Voluntary Pre-K Program**

The Tennessee pre-K program started in 2005 and is targeted toward low-income children, but has a tiered admissions process; at-risk children receive priority, but any remaining seats are open to all four-year-olds. Local school districts operate the program, though it is

primarily funded by the state. Teachers are required to have a license, the student to staff ratio cannot exceed 10:1, and the class size cannot exceed 20 students.

Lipsey et al. (2013) evaluated the effects of Tennessee's pre-K program on kindergarten and first grade test scores using an experimental design. We only include the kindergarten test scores in our meta-analysis since we are evaluating the short-term effects of pre-K and this evaluation was the only one that examined academic effects past kindergarten.

### **Tulsa, OK and Oklahoma State Universal Preschool Program**

The Oklahoma pre-K program is available for all four-year-olds in every school district. It started in 1990 as a pilot program for disadvantaged children and was expanded to universal coverage for all four-year-olds within the state in 1998. The program funding comes from the state and the local districts, amounting to approximately \$6,100 per student (Wong et al., 2009). All participating teachers are required to have a college degree and be certified in early childhood education. Teachers also receive compensation equal to that of public school teachers, which is much higher than the pay given to most daycare providers. Group sizes are not to exceed 20 students and the child per staff ratio cannot exceed 10:1 (Gormley & Gayer, 2005).

Wong et al. (2008) evaluated the effects of the Oklahoma Universal Preschool program on school readiness from 2004-2005 using an RDD. Gormley and Gayer (2005) evaluated Oklahoma's Universal Preschool program specifically examining Tulsa Public Schools (TPS) from 2000-2001 using an RDD.

### **West Virginia Universal Pre-K**

The West Virginia pre-K program is universally open to all four-year-olds, regardless of parent income. Districts operate the program and determine if eligibility is first-come-first-served or by lottery. The program takes place in a variety of settings: public school, Head Start centers,

and private preschool institutions. The program lasts the entire school year, but individual districts vary from two to four days a week. Teachers are required to be certified. Class size cannot exceed 20 and the child to staff ratio cannot exceed 10:1. Total funding is approximately \$6,829 per student. Wong et al. (2008) evaluated the effects of the West Virginia pre-K program on kindergarten entrant test scores from 2004-2005 using an RDD.

Table 3 below provides a descriptive overview of the seven selected studies and eleven program evaluations. Specifically, the table provides the study citation, preschool program, study time period, analytical technique, outcome measures, and standardized effect sizes as reported in each study. Table 3 shows that all eleven program evaluations found significant positive results for kindergarten test scores.

**Table 3: Description of Selected Studies**

Citation	Intervention	Sample	Time Frame	Method	Outcome Measure	Effect Sizes	Math	Reading
<b>Bartik (2013)</b>	Ready 4s Pre-K program in Kalamazoo, MI	Total sample size: 220	2011-2013	RDD	Vocabulary (PPVT), Letter-word ID (Woodcock-Johnson Subset1), Math (Woodcock-Johnson Subset 10)	Kindergarten entry: Vocabulary: 0.74; Letter-Word ID: 0.23, null; Math: 0.97	Positive	Positive to null
<b>Hustedt et al. (2010)</b>	New Mexico Pre-K program	Treatment- 653 Control- 706	2008-2009	RDD	Kindergarten math (Woodcock-Johnson Tests of Achievement-3), literacy (Early Literacy Skills Assessment) and vocabulary (PPVT-III)	Kindergarten entry: math 0.37; literacy 1.3; vocabulary 0.24	Positive	Positive
<b>Hustedt et al. (2015)</b>	Arkansas Better Chance Program	Treatment-506 Control-395	NA	RDD	Kindergarten Math (Woodcock-Johnson Test of Achievement-3), Literacy (PPVT-III, Pre-CTOPP)	Kindergarten entry: Vocabulary: 0.28; Print awareness: 0.82; Math 0.33.	Positive	Positive
<b>Gormley &amp; Gayer (2005)</b>	Tulsa Public Schools' Universal Pre-K Program	Treatment- 2,276 Control- 1,284	2000-2001	RDD	Kindergarten Math (Woodcock-Johnson Test of Achievement-3), Literacy (PPVT-III,) Cognitive, Language Skills (ECSI)	Kindergarten entry: Language: 0.38, Cognitive: 0.39 (Treatment-on-the-Treated).	Positive	Positive to null
<b>Lipsey et al. (2013)</b>	Tennessee Voluntary Pre-K Program	Treatment-773 Control-303	2009-2011	RFT	Student test scores in kindergarten and Woodcock Johnson Achievement Battery	Start of Kindergarten: Applied Problems: 0.18, Quantitative Concepts: 0.27, Letter-Word: 0.41, Spelling: 0.29, Oral Comp: 0.09, Picture Vocab: 0.20	Positive	Positive

Citation	Intervention	Sample	Time Frame	Method	Outcome Measure	Effect Sizes	Math	Reading
<b>Weiland &amp; Yoshikawa (2013)</b>	Boston Public School Pre-K program	Total sample 2,018	2009-2010	RDD	Vocabulary (PPVT-III), Reading (Woodcock-Johnson Letter-word ID Subset), Math (Woodcock-Johnson Subset Applied Problem Subscale and REMA)	Kindergarten entry: Vocabulary: 0.44; Reading: 0.62; Math: 0.59	Positive	Positive
	Oklahoma State Universal Pre-K Program	Treatment-431 Comparison-407	2004-2005	RDD	Students Vocabulary (PPVT-II), Math skill (Woodcock-Johnson Test of Achievement), Print Awareness (Pre-CTOPP)	Kindergarten entry: Vocabulary: 0.28; Print Awareness: 0.42, null; Math: 0.34, null;	Positive to null	Positive to null
	New Jersey Abbott Preschool Program	Treatment-1177 Comparison-895	2004-2005	RDD	Students Vocabulary (PPVT-II), Math skill (Woodcock-Johnson Test of Achievement), Print Awareness (Pre-CTOPP)	Kindergarten entry: Vocabulary: 0.36; Print Awareness: 0.32; Math: 0.23	Positive to null	Positive
<b>Wong et al. (2008)</b>	West Virginia Universal Pre-K	Treatment-379 Comparison-341	2004-2005	RDD	Students Vocabulary (PPVT-II), Math skill (Woodcock-Johnson Test of Achievement), Print Awareness (Pre-CTOPP)	Kindergarten entry: Vocabulary: 0.16, null; Print Awareness: 0.92 Math: 0.06, null	Positive to null	Positive to null
	South Carolina's Early Childhood Program	Treatment-353 Comparison-424	2004-2005	RDD	Students Vocabulary (PPVT-II), Print Awareness (Pre-CTOPP)	Kindergarten entry: Vocabulary: 0.04, null; Print Awareness: 0.78	NA	Positive to null
	Michigan School Readiness Program (MSRP)	Treatment-485 Comparison-386	2004-2005	RDD	Students Vocabulary (PPVT-II), Math skill (Woodcock-Johnson Test of Achievement), Print Awareness (Pre-CTOPP)	Kindergarten entry: Vocabulary: -0.13, null; Print Awareness: 1.09; Math: 0.53	Positive	Positive to null

### Systematic Review Results

First, we perform a simple vote-counting analysis and place program evaluations into five different categories for math and reading. These five categories are: positive, positive to null, null, negative to null, and negative. If a program evaluation found either fully-positive or fully-negative results for a specific subject on kindergarten entrant tests, it would receive a *positive* or *negative* vote, respectively. However, if the significance of the results depended on the model used, the program evaluation received a *null to positive* or *null to negative* vote. If all results for a given program and subject are insignificant, the evaluation receives a *null* vote.

Based on this vote-counting procedure, we find seven full positive and three null to positive effects of scaled-up pre-K programs on Kindergarten math scores and five full positive and six null to positive effects for Kindergarten reading scores (Table 4). The results are quite clear: scaled-up preschool programs have positive impacts on test scores in the short-run. In addition, there do not seem to be significant differences between the effects on math and reading. However, only one study examined impacts of pre-K on student achievement after kindergarten (Lipsey et al., 2013). This study used an experimental design and found null to negative impacts of pre-K on student test scores in Tennessee. Note, however, this same evaluation found substantial positive effects on kindergarten test scores. In other words, we find that pre-K has large positive impacts on kindergarten achievement, but the subsequent effects are unclear, as the only long-term evidence indicates fadeout and a sign reversal.

The studies that found fully positive effects all used an RDD. One of those programs, the New Jersey Abbott Preschool program, had effects that were sensitive to changes in the RD bandwidth (Table 5). The other two programs, Boston Public Schools Pre-K and New Mexico



Pre-K programs had effects that were robust across different bandwidths. The Oklahoma and West Virginia effects, which were positive to null, were also sensitive across bandwidths.

This may be a concern, since the internal validity associated with the RDD is mostly tied to how small the bandwidth is around the cutoff. The larger the bandwidth, the larger the endogeneity concern (Lee & Lemieux, 2009). There are two possible explanations for why results are not robust across various bandwidths: 1.) the program did not actually have a statistically significant effect on student test scores. 2.) There is not enough power with the RDD to detect the actual effect using the tighter bandwidth because of smaller sample sizes. The second explanation may be more plausible in this case since the coefficients on the results are very similar across the different bandwidths used.

We should also take note of the appropriate implications of the results. The results of the all methods are relative to not being enrolled in the particular pre-K program. The results do not compare the program to receiving no preschool or child care. The counterfactual is business-as-usual, which includes a mix of various types of preschool programs as well as no preschool at all. Students who missed the birthday cutoff or lost a lottery could have enrolled in private pre-K, Head Start, participated in home based care, or received no substantial educational instruction.

**Table 4: Summary of Results of Systematic Review of Short Term Effects of Pre-K**

<b>Study</b>	<b>Program Assessed</b>	<b>Impact of Pre-K on Math</b>	<b>Impact of Pre-K on Reading</b>	<b>Sample Size</b>
Bartik, Timothy J. (2013).	Kalamazoo, MI	Positive	Null / Positive	220
Gormley, W. T., & Gayer, T. (2005).	Tulsa, OK	Positive	Null / Positive	2,298
Hustedt, J. T., Barnett, W. S., Jung, K., & Friedman, A. H. (2010).	NM	Positive	Positive	1,358
Hustedt J. T., Jung, K., Barnett, W.S., & Willians, T. (2015).	AR	Positive	Positive	901
Lipsey, M. W., Hofer, K. G., Dong, N., Farran, D. C., & Bilbrey, C. (2013).	TN	Positive	Positive	1076
Weiland, C., & Yoshikawa, H. (2013).	Boston, MA	Positive	Positive	2,018
Wong, Vivian C., Thomas D. Cook, W. Steven Barnett, and Kwanghee Jung. (2008).	OK	Null/ Positive	Null/ Positive	838
	NJ	Positive	Positive	2,072
	WV	Null/ Positive	Null/ Positive	720
	SC	NA	Null/ Positive	777
	MI	Positive	Null/ Positive	871

**Table 5: Summary of Regression-Discontinuity Results by Bandwidth**

Study	Program Assessed	365 Days or 12 months		180 Days or 6 months		90 Days or Fewer	
		Reading	Math	Reading	Math	Reading	Math
Gormley, W. T., & Gayer, T. (2005).	Tulsa, OK	Positive	Positive	Positive	Positive	Null	Positive
Wong, Vivian C., Thomas D. Cook, W. Steven Barnett, and Kwanghee Jung. (2008).	OK	Positive	Positive	Null	Null	Null	Null
	NJ	Positive	Positive	Positive	Null	Null / Positive	Null
	WV	Null / Positive	Positive	Null / Positive	Null	Null / Positive	Null
Bartik, Timothy J. (2013).	Kalamazoo, MI	Null / Positive	Positive	Null / Positive	Positive	N/A	N/A
Weiland, C., & Yoshikawa, H. (2013).	Boston, MA	Positive	Positive	Positive	Positive	N/A	N/A
Hustedt, J. T., Barnett, W. S., Jung, K., & Friedman, A. H. (2010).	NM	Positive	Positive	Positive	Positive	Positive	Positive
Hustedt J. T., Jung, K., Barnett, W.S., & Willians, T. (2015).	AR	Positive	Positive	Null / positive	Null	Null / positive	Null

### Meta-Analysis Results

The vote-counting results for the short-term effects of pre-K, as measured by kindergarten-entry achievement, show that out of the eleven results found, all show significant positive results in the short-run. However, this type of approach does not tell us much about the size of the effects. Consequently, it is beneficial to determine an overall effect size from the range of reported effects for policy implications.

To understand the effects of scaled-up pre-K, we calculated a meta-analytic average for the effects on kindergarten math and reading test scores separately. If a study reported more than one math score result, we combined them by taking an average of each effect size for an overall math score effect size. Similarly, if a study reported more than one reading score result, we combined them by taking the average of the effect sizes. For example, if a study reported both a PPVT score and a score for print awareness (Pre-CTOPP), we averaged the two scores to create a combined reading score.

In order to compute a meta-analytic effect size, we used results reported from the one-year bandwidth around each side of the birthday cutoff since this was used in every RDD study. We did not compute a meta-analytic average for smaller bandwidths because most studies did not give the information needed in order to compute an effect size.<sup>3</sup>

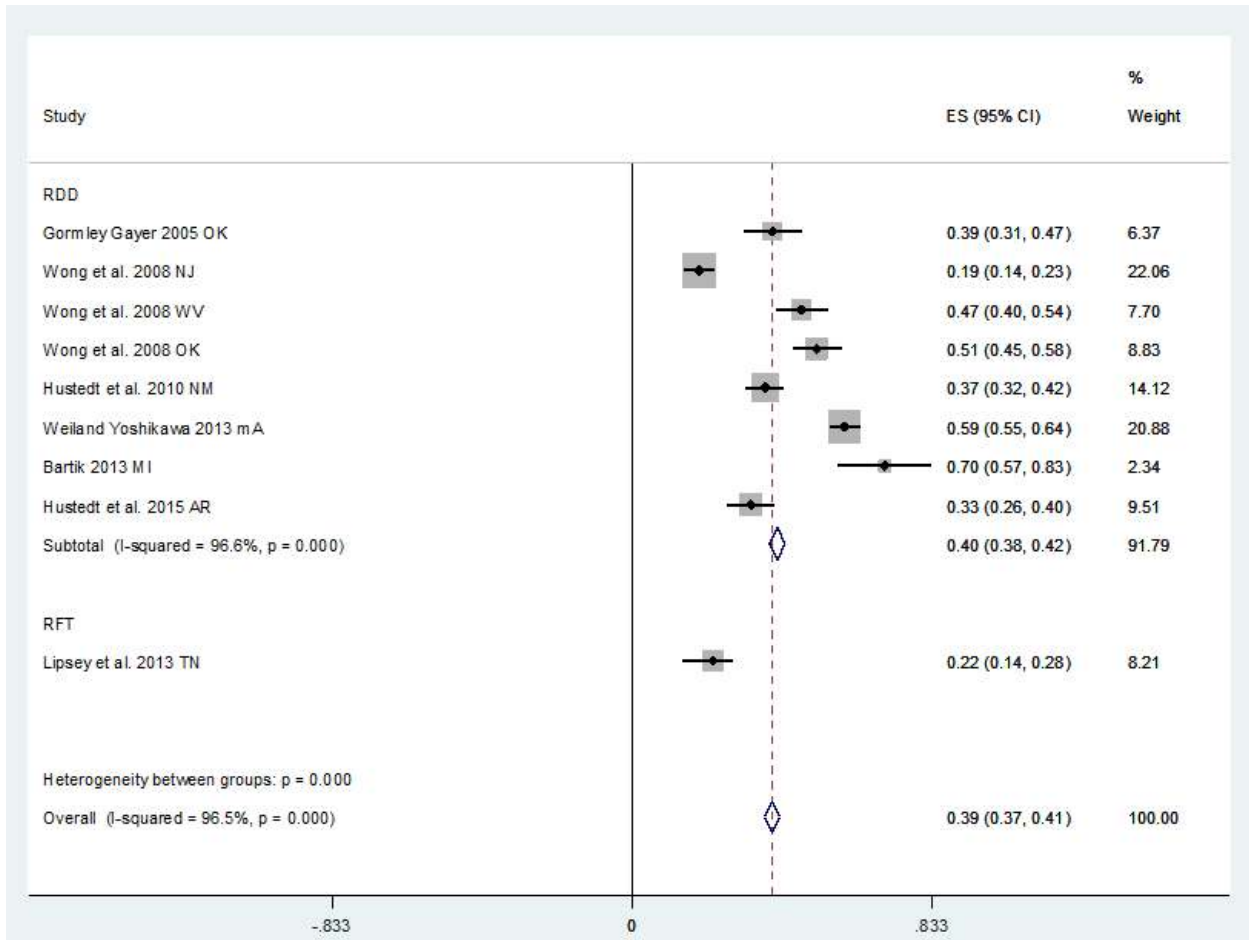
To compute the meta-analytic effect size, we use Glass' Delta. The effect sizes that we use are the difference in means between the treatment and control groups divided by the standard deviation of the control group (Glass, Smith, & McGaw, 1981). We use this method since the control group is the closest representation we have of the counterfactual.

---

<sup>3</sup> These studies did not report standard deviations for the sub-samples of students within the smaller bandwidth.

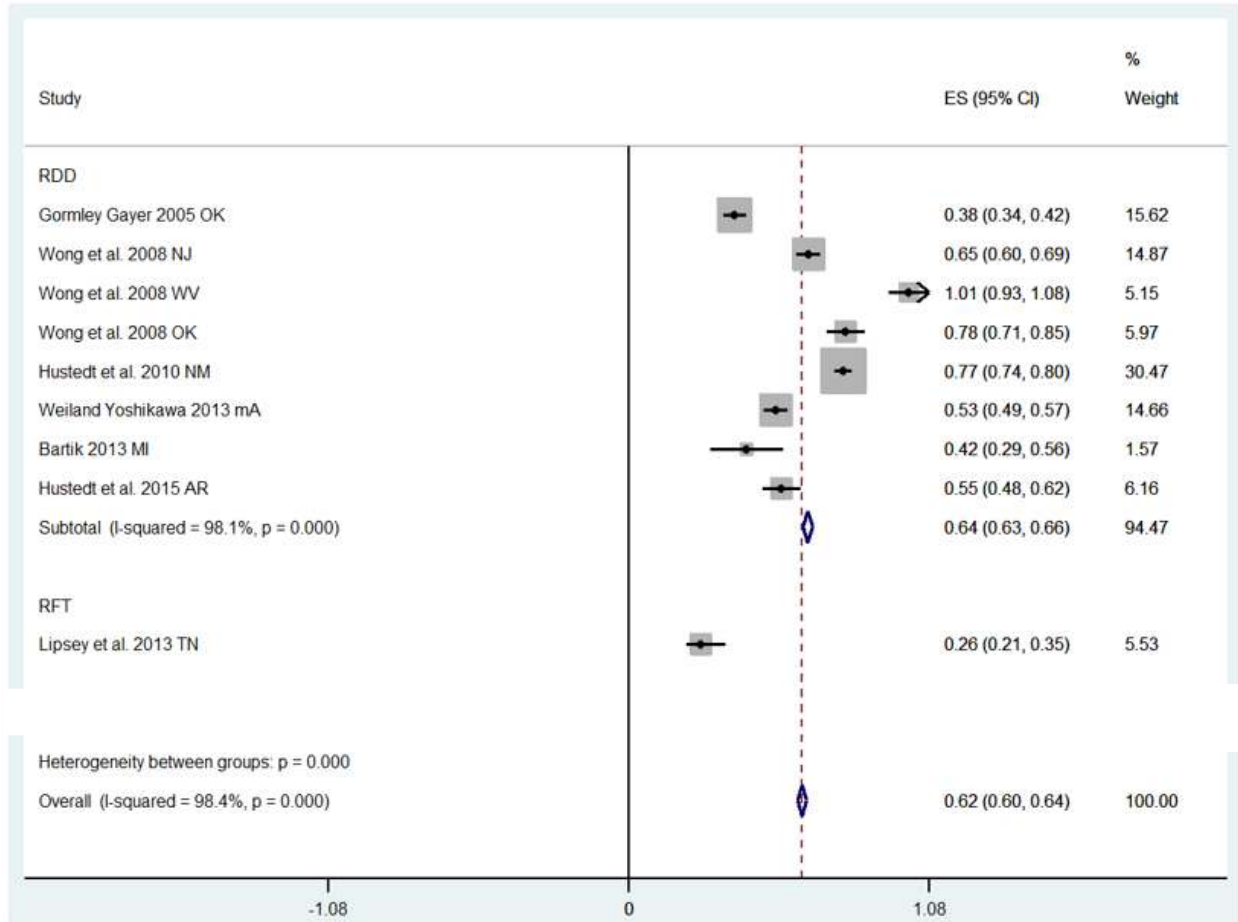
The meta-analytic effect size for math scores is 0.39 standard deviations (Figure 2) and 0.62 standard deviations for reading scores (Figure 3). Both effects are statistically significant at the 95% confidence level. The effect sizes ranged from 0.19-0.70 standard deviations for math and 0.26-1.01 standard deviations for reading. Scaled-up pre-K seems to have a larger effect on kindergarten reading test scores than math scores. Compared to the effects of other K-12 program evaluations, these all are quite large. A previous meta-analysis of the achievement effect sizes of education programs found that only five percent of the programs had an effect size over 0.39 standard deviations (Borman et al., 2003). Effect sizes should be interpreted in context. The average educational program effect size on standardized tests from experimental research for elementary school students range from 0.07 for broadly focused standardized tests to 0.44 for highly specialized tests; however, the magnitude of the effect size is likely to vary depending on the interventions targeted student population (Hill et al., 2007).

**Figure 2: Forest Plot Results for Kindergarten Mathematics Scores by Method**



*Note: All RDD estimates use a 12-month bandwidth on either side of the birthday cutoff date.*

**Figure 3: Forest Plot Results for Kindergarten Reading Scores by Method**



*Note: All RDD estimates use a 12-month bandwidth on either side of the birthday cutoff date.*

### Discussion and Implications

Certain policy questions are particularly ripe for interrogation via a systematic review of the evidence. There are several characteristics of these policy issues for which a careful review of the evidence would be useful to policymakers. First, systematic reviews can add value for questions that are properly addressed through quantitative analyses. When previous high-profile studies have reached apparently contradictory conclusions, reviews can be particularly helpful. Finally, the laborious efforts required for a systematic literature review are best exerted when a policy question is high on the agenda of policymakers.

The question of whether to expand pre-K programs certainly meets the conditions described above. The effectiveness of such programs are measured in terms of test scores, later schooling outcomes, or even later life outcomes, each of which can be counted and compared. At the same time, there have been high profile studies of pre-K programs generating headlines that tell opposing stories in recent years. For example, Puma et al. (2012), a random assignment study of the impacts of Head Start through grade three found “little evidence of systematic differences in children’s elementary school experiences through 3<sup>rd</sup> grade.” On the other hand, Nobel prize-winning economist James Heckman of the University of Chicago conducted a cost-benefit analysis of the famous Perry Preschool program and found that each dollar invested in pre-K would generate more than \$7 in return for society overall. This positive finding resonated so much that President Obama actually referred to it in his 2013 State of the Union Address by stating,<sup>4</sup> “Every dollar we invest in high-quality early education can save more than seven dollars later on – by boosting graduation rates, reducing teen pregnancy, even reducing violent crime.”

In the midst of this optimism surrounding early education, a 2015 random assignment

---

<sup>4</sup> [http://www.nytimes.com/2013/02/13/us/politics/obamas-2013-state-of-the-union-address.html?pagewanted=all&\\_r=1&](http://www.nytimes.com/2013/02/13/us/politics/obamas-2013-state-of-the-union-address.html?pagewanted=all&_r=1&)



study from Vanderbilt University upset the apple cart by finding that the Tennessee state-funded pre-K program did not show such positive results, and indeed appeared to have negative impacts at the end of first grade (Lipsey et al., 2013). In addition, with all of this confusion swirling, state policymakers across the country are making decisions about whether to support or even expand such early childhood education programs.

This is the context in which we present the findings of our meta-analysis of the effectiveness of scaled up early childhood education programs. The goal of our systematic review is to describe, with clarity, what we do know and what we do not know based on the current state of the research. Much of the debate and confusion, we believe, is due to a lack of agreement, or clarity, around the specific meaning of pre-K. In fact, all pre-K programs are not created equal; we learn from the experiences of some programs are not applicable to the potential success of other programs. For example, advocates of expanded pre-K sometimes cite the very positive results of the Perry Preschool project or the Abecedarian study, despite the fact that these were both small, intensive, and multi-year interventions that were delivered nearly 50 years ago.<sup>5</sup> For numerous reasons, it does not seem appropriate to extrapolate from these findings expectations for today's statewide or districtwide scaled-up pre-K programs. Nevertheless, some still rely on these two random assignment studies to make lofty claims on the cost effectiveness of pre-K expansion.

In the interest of clarity, we focus our review on the impacts of the scaled-up, publicly funded pre-kindergarten education programs. We focus on these sorts of programs because they have expanded in recent years and are now provided in settings such as Oklahoma, New Jersey, Arkansas, and Tennessee. Importantly, these are the types of programs that policymakers are

---

<sup>5</sup> <https://www.brookings.edu/research/does-pre-k-work-it-depends-how-picky-you-are/>

considering developing or expanding in districts or states across the country. While most would agree that small and expensive programs such as Perry Preschool have boasted clearly positive outcomes, there is no realistic push to bring such programs to scale.

Therefore, in this review, we ask the policy relevant question: what does the existing research reveal about the average effects of scaled-up (or universal), publicly funded pre-kindergarten (pre-K) programs on student kindergarten readiness in math and reading. While we would certainly like to provide information about longer term impacts of such programs (even through elementary school), the existing research base can only speak to these short term effects. Our findings are important and noteworthy; moreover, it is equally important that we articulate what we cannot conclude from this analysis.

*What do we know?* According to our review, the existing literature clearly shows significant academic benefits, in the short-term, of scaled-up pre-K in the United States. The results are unambiguous: these so-called universal pre-K programs have large positive impacts on math and reading test scores in the year following the intervention. In particular, our meta-analysis shows that access to scaled-up preschool is associated with over a third of a standard deviation increase in kindergarten math scores and over three-fifths of a standard deviation increase in kindergarten reading scores.

*What are the limitations?* While most researchers believe that regression discontinuity designs have strong internal validity, some have raised caution flags about the use of RDD evaluations in this context (e.g. Lipsey, Weiland, Yoshikawa, Wilson, & Hofer, 2015). Lipsey et al. (2015) point out that treatment and control conditions for pre-K evaluations are identified post hoc, when the outcome of interest is measured. Because of this, the estimates are treatment-on-the-treated (TOT), not intent-to-treat (ITT), as Gormley and Gayer (2005) noticed and explained.

In order for the TOT estimates to be unbiased, researchers must use the initial assignment of each student in the ITT sample to predict whether they actually received the pre-K treatment. In addition, researchers need to know whether the student complied with initial assignment and the outcome for all individuals whether or not they complied. Importantly, the attrition between the initial sample and the analysis sample can cause a serious threat to internal validity of estimates. Lipsey and colleagues also demonstrate that the direction and magnitude of the resulting bias is unknown. Consequently, future pre-K evaluations using the RDD methodology would be much stronger if researchers could obtain data on initial assignment, student compliance, and outcomes for the entire sample.

Finally, there is another pattern in our meta-analytic results which should also temper our optimism related to the magnitude of the positive effects: the six RDD studies found overall kindergarten effect sizes of 0.40 for mathematics and 0.64 for reading while the RFT study found smaller overall kindergarten effect sizes of 0.22 for mathematics and 0.26 for reading.

*What don't we know?* While the meta-analysis does give a better understanding of the overall short run effect of scaled-up pre-K and uncovers clearly positive results, it does not tell us anything about longer-term effects. This is an important point, as policymakers who choose to invest millions of dollars into state-funded pre-K programs may well expect that such investments will spur positive results beyond the kindergarten year. Unfortunately, the Nashville RFT study (Lipsey et al., 2013) was also the only evaluation to consider longer-term effects and indeed found that the positive effects did not persist beyond the kindergarten test scores. Surely, a single research study examining the possibility of persistent effects does not have the final say on the potential benefits of early childhood education. Nevertheless, it does provide a sober

reminder that we should not simply assume longer-term benefits based on the preponderance of evidence documenting the short-term positive effects of early childhood education.

*Where do we go from here?* In short, our review highlights and clearly documents the fact that there are strong reasons to believe that exposure to so-called universal pre-K programs will improve kindergarten readiness for students in both math and reading. We must now engage in further study to gain a better understanding on the extent to which these benefits persist throughout the years for students and can positively affect their educational trajectories. On this question, we have almost zero information.

As a result, policymakers and practitioners would benefit from additional longitudinal experimental research to determine the long-run impacts of scaled-up pre-K. As policymakers consider the implementation of new pre-K programs, or the expansions of existing ones, it is our sincere hope that such program growth is always accompanied by rigorous long-term evaluations. The knowledge base on this question has grown a great deal over the last 15-20 years; random assignment evaluations will likely be necessary to answer the next set of important questions. There appears to be a consensus, based on both new research on the importance of intellectual stimulation for very young children and on the positive short-term effects of pre-K programs, which early childhood education should be offered more broadly. As we will likely invest millions or billions of dollars across the country in pre-K over the next decade or so, it would be a shame if we did not also invest in rigorous evaluation so that we can know with more confidence the extent to which this sort of investment truly improves the lives of children.

## Appendix

### Ebsco Search

*(( pre school\* or preschool\* or pre-K\* or prekinder\* ) AND ( math\* or reading\* or literacy\* ) AND ( regression\* or experiment\* or RCT or RFT or random\* or quasi\*or quantitative\* ) AND ( effect\* or impact\* or outcome\* or result\* ) AND (( program\* or evaluation\* ))*

### ProQuest Search

*(all(preschool\* OR pre-K\* OR prekinder\* OR "pre school") AND all((math\* OR reading\* OR literacy\*)) AND all((random\* OR quant\* OR regression\* OR experiment\*)) AND all((effect\* OR result\* OR impact\* OR outcome\*)) AND all((program\* OR evaluation\*))) NOT (at.exact("News" OR "Case Study") NOT fdb(10000159 10000155 1007899 10000020) NOT ccl.exact("Nutrition" OR "Health care industry" OR "Health education" OR "Nursing" OR "Occupational psychology") NOT aloc.exact("Turkey" OR "Canada" OR "Saudi Arabia" OR "Campinas Brazil" OR "Chile" OR "Geelong Victoria Australia" OR "Germany" OR "India" OR "Jordan" OR "Kazakhstan" OR "Kosovo" OR "Malawi" OR "Melbourne Victoria Australia" OR "Ontario Canada" OR "Philippines" OR "Spain" OR "Sudan" OR "Sudan, Khartoum") NOT la.exact("SPA" OR "FRE" OR "GER" OR "AFR" OR "RUS" OR "TUR") AND yr(2000-2019))*

### References

- Anderson, Shinn, Fullilove, Scrimshaw, Fielding, Normand, & Task Force on Community Preventive Services (2003). The Effectiveness of Early Childhood Development Programs: A Systematic Review. *American Journal of Preventive Medicine*, 24(3), 32-46.
- Artz & Welsch (2016). The Impact of Publicly Provided Early Childhood Education Programs on District-Level Test Scores. *Contemporary Economic Policy*, 34(1), 89-106.
- Barnett (2008). Preschool Education and Its Lasting Effects: Research and Policy Implications. *Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit*.
- Barnett & Leonard (2007). Comparative Benefit--Cost Analysis of the Abecedarian Program and Its Policy Implications. *Economics of Education Review* 26, iss 1, pg: 113-125.
- Bartik (2013). Effects of the Pre-K Program of Kalamazoo County Ready 4s on Kindergarten Entry Test Scores: Estimates Based on Data from The Fall of 2011 And The Fall of 2012. *Upjohn Institute Working Paper*: 13-198.
- Borman, Hewes, Overman, & Brown (2003). Comprehensive School Reform and Achievement: A Meta-Analysis. *Review of Educational Research*, 73(2), 125-230.
- Brown & Scott-Little (2003). Evaluations of School Readiness Initiatives: What Are We Learning? SERVE's Expanded Learning Opportunities National Leadership Area Research Report.
- Camilli, Vargas, Ryan & Barnett (2010). Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development. *Teachers College Record* 112, no. 2, pg. 579-620.

- Campbell, Ramey, Pungello, Sparling, & Miller-Johnson (2002). Early Childhood Education: Young Adult Outcomes from the Abecedarian Project. *Applied Developmental Science* 6, iss. 1: pg. 42-57.
- Chambers, Cheung, Slavin, Smith, & Laurenzano (2010). Effective Early Childhood Education Programs: A Systematic Review. *Center for Research and Reform in Education*.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates. Inc CIT0006.
- Friedman-Krauss, Barnett, & Nores (2016). How Much Can High-Quality Universal Pre-K Reduce Achievement Gaps? *National Institute for Early Education Research*.
- Gilliam & Zigler (2001). A Critical Meta-Analysis of All Evaluations of State-Funded Preschool from 1977 to 1998: Implications for Policy, Service Delivery and Program Evaluation. *Early Childhood Research Quarterly*, 15(4), 441-473.
- Glass, Smith, & McGaw (1981). *Meta-Analysis in Social Research*. Sage Publications, Incorporated.
- Gormley & Gayer (2005). Promoting School Readiness in Oklahoma: An Evaluation of Tulsa's Pre-K Program. *The Journal of Human Resources*, 40(3), 533–558.
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the High Scope Perry Preschool Program. *Journal of Public Economics*, 94(1), 114-128.
- Hill, Bloom, Rebeck Black, & Lipsey (2008). Empirical Benchmarks for Interpreting Effect Sizes in Research. *MDRC Working Paper*.

- Hustedt, J. T., Barnett, W. S., Jung, K., & Friedman, A. H. (2010). The New Mexico PreK Evaluation: Impacts from The Fourth Year (2008-2009) Of New Mexico's State-Funded Prek Program. New Brunswick, NJ: National Institute for Early Education Research.
- Hustedt, J. T., Jung, K., Barnett, W. S., & Williams, T. (2015). Kindergarten Readiness Impacts of the Arkansas Better Chance State Prekindergarten Initiative. *The Elementary School Journal*, 116(2), 198-216.
- Lee, D. S., & Lemieux, T. (2009). Regression Discontinuity Designs in Economics (No. w14723). National Bureau of Economic Research.
- Lipsey, Mark W. (1990). *Design Sensitivity: Statistical Power for Experimental Research* (Vol. 19). Sage Publications.
- Lipsey, Mark W., Hofer, Dong, Farran, & Bilbrey (2013). Evaluation of the Tennessee Voluntary PreKindergarten Program: Kindergarten and First Grade Follow-Up Results from the Randomized Control Design. Peabody Research Institute.
- Lipsey, M. W., Weiland, C., Yoshikawa, H., Wilson, S. J., & Hofer, K. G. (2015). The Prekindergarten Age-Cutoff Regression-Discontinuity Design Methodological Issues and Implications for Application. *Educational Evaluation and Policy Analysis*, 37(3), 296-313.
- Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., ... & Ciarico, J. (2010). Head Start Impact Study. Final Report. *Administration for Children & Families*.
- Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., ... & Downer, J. (2012). Third Grade Follow-Up to the Head Start Impact Study: Final Report. OPRE Report 2012-45. *Administration for Children & Families*.



- Racial and Ethnic Achievement Gaps. (2016). Retrieved October 28, 2016, from <http://cepa.stanford.edu/educational-opportunity-monitoring-project/achievement-gaps/race/#second>
- Schweinhart, Lawrence J. (2003) Benefits, Costs, and Explanation of the High/Scope Perry Preschool Program. Paper presented at the Meeting of *the Society for Research in Child Development*.
- Schweinhart, Montie, Xiang, Barnett, Belfield, & Nores (2005). The High/Scope Perry Preschool Study Through Age 40. Excerpt from *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*: pp. 194-125.
- Slaby, R., Loucks, S., & Stelwagon, P. (2005). Why Is Preschool Essential in Closing the Achievement Gap?. *Educational Leadership and Administration*, 17, 47.
- Swain, W. A., Springer, M. G., & Hofer, K. G. (2015). Early Grade Teacher Effectiveness and Pre-K Effect Persistence. *AERA Open*, 1(4), 2332858415612751.
- United States Census Bureau. QuickFacts: Kalamazoo County, Michigan. <http://www.census.gov/quickfacts/table/SEX255214/26077>.
- U.S. Department of Health and Human Services. (2010). Head Start Impact Study: Final Report. Retrieved from <http://www.acf.hhs.gov/programs/opre/resource/head-start-impact-study-final-report>
- Washington State Institute for Public Policy. (2013). Model Early Childhood Education Programs: Meta-Analytic Results.
- Weiland, C., & Yoshikawa, H. (2013). Impacts of a Prekindergarten Program on Children's Mathematics, Language, Literacy, Executive Function, and Emotional Skills. *Child Development*, 84(6), 2112-2130.

Wong, Vivian C., Cook, Barnett, & Jung (2008). An Effectiveness-Based Evaluation of Five State Pre-Kindergarten Programs. *Journal of Policy Analysis and Management*, 27, no. 1: 122-154.

Yoshikawa, Hirokazu, Weiland, Brooks-Gunn, Burchinal, Espinosa, Gormley, Ludwig, Magnuson, Phillips, & Zaslow. (2013). Investing in Our Future: The Evidence Base on Preschool Education. Society for Research in Child Development.