



A Systematic Review of Research Designs Used to Evaluate State and Locally Funded Pre-K Programs

Anamarie A. Whitaker
University of Delaware

Cara L. Kelly
University of Oklahoma

Tyler W. Watts
Columbia University

Mark W. Lipsey
Vanderbilt University

Juan Camilo Cristancho
University of California,
Irvine

Jade M. Jenkins
University of California,
Irvine

Caroline M. Botvin
Columbia University

Drew Bailey
University of California,
Irvine

Greg J. Duncan
University of California,
Irvine

Tania Maria Cruz
Cordero
University of Delaware

Melissa Stoffers
University of Delaware

Mindy Rosengarten
Columbia University

Expansions of state and locally funded pre-k programs in recent decades have been promoted by advocates citing strong research support. However, questions have been raised about how supportive the research base actually is. We conducted a systematic review of pre-k evaluation studies (from 2000-2021) to document their research designs and the direction of their reported impacts. Most of the 120 evaluation reports identified came from 91 studies that used designs lacking internally valid comparisons between pre-k attenders and comparison children. The direction of achievement effects assessed at the end of pre-k was overwhelmingly positive, regardless of research design. At the same time, the direction of longer-term achievement effects varied by research design. Specifically, 65% of the estimates from randomized evaluations were negative in direction, compared with between 16% and 25% of estimates based on other designs. Our findings point to an urgent need for more rigorous research on pre-k effectiveness, particularly from studies using randomized designs, to investigate longer-term effects.

VERSION: August 2026

Suggested citation: Whitaker, Anamarie A., Cara L. Kelly, Tyler W. Watts, Mark Lipsey, Juan Camilo Cristancho, Jade M. Jenkins, Caroline M. Botvin, Drew H. Bailey, Greg J. Duncan, Tania Maria Cruz Cordero, Melissa Stoffers, and Mindy Rosengarten. (2026). A Systematic Review of Research Designs Used to Evaluate State and Locally Funded Pre-K Programs. (EdWorkingPaper: 26-1545). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/ebc6-je24>

**A Systematic Review of Research Designs Used to Evaluate
State and Locally Funded Pre-K Programs**

Anamarie A. Whitaker
University of Delaware

Cara L. Kelly
University of Oklahoma

Tyler W. Watts
Teachers College, Columbia University

Mark W. Lipsey
Vanderbilt University

Juan Camilo Cristancho
Jade M. Jenkins
University of California, Irvine

Caroline M. Botvin
Teachers College, Columbia University

Drew Bailey
Greg J. Duncan
University of California, Irvine

Tania Maria Cruz Cordero
Melissa Stoffers
University of Delaware

Mindy Rosengarten
Teachers College, Columbia University

Abstract

Expansions of state and locally funded pre-k programs in recent decades have been promoted by advocates citing strong research support. However, questions have been raised about how supportive the research base actually is. We conducted a systematic review of pre-k evaluation studies (from 2000-2021) to document their research designs and the direction of their reported impacts. Most of the 120 evaluation reports identified came from 91 studies that used designs lacking internally valid comparisons between pre-k attenders and comparison children. The direction of achievement effects assessed at the end of pre-k was overwhelmingly positive, regardless of research design. At the same time, the direction of longer-term achievement effects varied by research design. Specifically, 65% of the estimates from randomized evaluations were negative in direction, compared with between 16% and 25% of estimates based on other designs. Our findings point to an urgent need for more rigorous research on pre-k effectiveness, particularly from studies using randomized designs, to investigate longer-term effects.

Keywords: Systematic Review, Pre-Kindergarten Programs, Early Childhood Education

A Systematic Review of Research Designs Used to Evaluate State and Locally Funded Pre-K Programs

Many policymakers, practitioners, advocates, and researchers have argued that a strong body of evidence establishes the importance of pre-kindergarten (pre-k) programs for children's short- and long-term development. Frequently cited evidence comes from two small-scale, researcher-designed studies—Perry Preschool and Abecedarian—which operated in the 1960s and 1970s. These random assignment studies demonstrated lifelong positive effects of preschool participation on children's educational attainment, physical and mental health, economic status (Hahn & Barnett, 2023), and a reduction in criminal activity (García et al., 2019). These studies, plus more recent evaluations of some modern-day preschool programs, have provided the evidence to support a substantial scale-up of state and local pre-k programs over the past 30 years. Advocates argue that these programs constitute a population-level intervention to address the consequences of poverty and inequality of educational opportunities prior to school entry (Yoshikawa et al., 2013; White House, 2024).

Given the potential benefits—and costs—of increasing investments in pre-k, policymakers, practitioners, and researchers have been highly motivated to understand the effects of pre-k programs on children. This has led to many evaluation studies, some of which have appeared in academic journals but many of which take the form of reports commissioned by municipal and state governments and conducted by evaluation firms. The research designs deployed in this literature are diverse. Importantly, designs vary in the extent to which they meet widely-accepted scientific standards for evidence of program effectiveness. Although this variation exists in any field of inquiry, the use and quality of such evidence warrants serious attention when over one-third of American 4-year-olds enroll in such programmatic interventions each year (Friedman-Krauss et al., 2026).

Pre-k evaluations often focus on comparisons of mean elementary school state test scores for children who did and did not attend the program. This is potentially worrisome because differences between attendees and non-attendees, even when adjusted for differences in family resources, can be affected by many other child-, family- and community-specific factors determining which children are and are not enrolled by their parents in the program. As a result, evaluations based on these kinds of comparisons may not reflect the magnitude and even direction of pre-k effects. With weak evidence, policymakers cannot be confident that their pre-k programs are supportive of the needs of children enrolled in them.

The State of Pre-K Evaluations

A primary focus of pre-k research is on the timing and persistence of pre-k effects. Impacts estimated at the end of the pre-k year often come from comparisons of indicators of school readiness (academic and social-emotional) for children who just made or just fell short of the birthday cutoffs defining eligibility. These kinds of studies typically show positive short-run impacts on school readiness and are often viewed as laying the groundwork for children to participate more fully and productively in elementary school instruction (Brooks-Gunn et al., 2016). The end-of-program effects of scaled-up pre-k programs have been widely studied (Camilli et al., 2010; Phillips et al., 2017) and evaluators generally report positive effects on academic and social-emotional outcomes across a variety of program contexts, subgroups of children, and measures (e.g., Abenavoli, 2019; Duncan & Magnuson, 2013; Gormley & Gayer 2005; Lipsey et al., 2018; McCormick et al., 2021; Wong et al., 2008).

Based on evidence from legacy studies, many stakeholders expect that favorable evidence of end-of-program learning advantages for their programs will lead to continuing gains in K-12 and beyond. Yet, accumulating evidence has led some researchers to be concerned that non-

legacy pre-k programs often produce effects that fade out across the elementary school years (Bailey et al., 2017). A small number of randomized evaluation studies have reported largely null or even negative effects on academic achievement and some behavioral outcomes measured in elementary school (e.g., Durkin et al., 2022; Peisner-Feinberg et al., 2020; Weiland et al., 2020; Weiland et al., 2025; Woodyard et al., 2025). Puzzlingly, one study reports null effects on achievement test scores during grade school but positive effects on high school graduation, SAT taking, and college attendance, and reductions in criminal behavior (Gray-Lobe et al., 2023; see Burchinal et al., 2024 for an interpretive review).

In contrast, a larger body of quasi-experimental research reports generally positive effects on academic and social-emotional outcomes during children's K-12 years, with effect magnitudes that are usually smaller than those observed at the end of pre-k (e.g., Amadon et al., 2022; Barnett & Jung, 2021; Early et al., 2019; Gormley et al., 2011; Gormley et al., 2018). In addition, a few quasi-experimental studies report positive effects after high school, for outcomes such as high school graduation (Courtney et al., 2023) and civic participation (Kitchens & Gormley, 2023). Taken together, studies examining longer-term outcomes appear to produce mixed results.

Mixed or inconclusive findings could indicate the need for more research to identify underlying causes of effect heterogeneity. But another possible culprit is the diversity of the evaluation research designs employed in this literature, with some designs much better able to provide unbiased estimates of short- and longer-run program impacts than others (Phillips et al., 2017). Moreover, researchers (e.g., Meloy et al., 2019) and policymakers (e.g., Obama, 2013) have made general assertions that questions of pre-k effectiveness have already been settled by

the empirical literature. But if most studies have used designs that do not provide strong internal validity, such assertions may be premature.

Our systematic review begins with an overview of the kinds of evaluation designs used in this literature, with a focus on their ability to provide unbiased estimates of program effects (i.e., internal validity). Next, we present results from our assembled corpus of pre-k effectiveness studies, beginning with the frequency of use of different research designs. We then assess whether different designs yield different results about the nature of pre-k program influences on children. Because many of the evaluation studies we reviewed failed to provide the information needed to calculate program effect sizes, we characterize each reported finding as positive, null, or negative for the outcomes examined, distinguishing between end-of-program and longer-term effects and between achievement-related and social-emotional learning and behavioral outcomes. In this way, we aim to construct an overall picture of the nature and quality of the methods used to study the effectiveness of pre-k programs and the extent to which these methods have produced similar or different results. To assess which studies have differential influence in the scientific discourse, we look at study citations within the broader literature. In closing, we discuss how this research literature continues to influence public discourse, and we consider strategies for improving the quality of the evaluation research in this area.

Research Designs for Investigating the Effects of Pre-K

Every study involves balancing myriad methodological considerations, the most fundamental of which is the basic design of the study. An evaluation study investigating pre-k effects on the participating children makes a causal inference: that a child's exposure to pre-k affected selected outcomes, and that child outcomes would have differed in the absence of that exposure. The field has well-established standards that describe research capable of supporting

credible causal inferences and differentiate stronger (more “rigorous”) research designs from weaker ones (e.g., What Works Clearinghouse, 2022). For the key causal assumptions of a pre-k evaluation study to be plausible, researchers need to ensure that exposure to pre-k varies while everything else that might influence the outcome remains constant for pre-k attenders and non-attenders. Only with this assumption can researchers be confident that pre-k exposure *caused* changes in an outcome. Of course, the same child cannot simultaneously participate and not participate in pre-k at age 4. Causally informative comparisons require that, on average, participants and nonparticipants are equivalent on all other characteristics—both measured and unmeasured—that affect both participation and outcomes (Cook et al., 2002; Holland, 1986).

Randomized and Quasi-Experimental Pre-K Evaluations

This demanding standard is best met by random assignment to pre-k participation. As the sample size increases, the likelihood of differences on any characteristic other than the pre-k condition diminishes, whether known or unknown, measured or unmeasured (Rubin, 1974). Most randomized studies of pre-k programs involve oversubscribed program sites where an administrator or researcher uses a lottery to determine which children are admitted to the program (e.g., Durkin et al., 2022; Gray-Lobe et al., 2023; Peisner-Feinberg et al., 2020; Weiland et al., 2020; Woodyard et al., 2025). Yet, very few randomized studies of state- or locally funded pre-k programs exist (Burchinal et al., 2024; Bruhn & Emick, 2023). Although randomized designs offer high internal validity, concerns about external validity include the possibility that families selecting into oversubscribed programs may benefit differently from attending pre-k programs than those selecting into undersubscribed programs (Weiland et al., 2024).

Most research on the effects of state- and locally funded pre-k programs uses nonrandomized designs. For example, researchers have used birthday eligibility cutoffs to compare similarly aged children who are just completing and just beginning their pre-k program year. However, this regression discontinuity design (RDD) does not allow for the examination of longer-term outcomes (the comparison group gets the treatment after the pre-k year; e.g., DeAngelis et al., 2020; Wong et al., 2008). Further, the design relies on assumptions about the comparability of successive cohorts of pre-k participants that may be questionable (e.g., differential attrition; Lipsey et al., 2015).

Variations on designs that compare pre-k participants and nonparticipants include those that use proxy variables for pre-k exposure (e.g., funding availability or years before and after pre-k became available in a catchment area). These studies examine differences in outcome measures before and after a change in pre-k supply in selected catchment areas and compare them with changes in scores observed over a comparable period in another area with more restricted pre-k availability (e.g., Cascio, 2023; Watts et al., 2023). These designs are often referred to as “difference-in-difference” (DD) studies. The challenge for researchers using this design is to isolate the difference made in the target outcomes by pre-k implementation from other influential factors co-occurring over the same period (e.g., other investments in young children and their families that may accompany initiatives to expand pre-k; this is referred to as the “parallel trends” assumption in the econometric literature—see Cunningham, 2021, chapter 9). An additional challenge for readers is that the large number of data-dependent decisions that go into these designs creates room for researcher discretion in analytic decisions. In the presence of a consensus socially desirable finding (i.e., pre-k helps children in the short- and long-term), such discretion leaves more room for publication bias than a well-designed RCT. In economics,

such studies are more likely to show evidence of selective reporting and/or publication than RCTs (Brodeur et al., 2020). Unfortunately, we cannot open this “file drawer.” But quasi-experimental studies with imprecise estimates relative to their standard errors are plausibly upwardly biased.

Nonequivalent Comparison Group Pre-K Evaluations

More common are nonrandomized designs known as *nonequivalent comparison group designs* because equivalence between treatment and control groups on all relevant variables cannot be assumed (May & Collier, 2023). In these designs, outcomes are compared for children who attended pre-k and children who did not. A common example is the use of administrative data on pre-k attendance and 3rd grade reading scores to assess whether children who had attended a pre-k program outperformed children who did not attend. Control variables often include basic demographic variables known to be related to academic performance (e.g., socioeconomic status, gender). These types of easy, low-cost studies are common. For example, the Arkansas Department of Education has funded an annual report comparing pre-k and non pre-k attenders on elementary and middle school test scores and grade retention conditional on free- and reduced price-lunch status (Arkansas Research Center, n.d.).

In some cases, researchers attempt to mitigate the potential for selection bias in nonequivalent comparisons with some form of regression control or statistical matching to equate the participant and nonparticipant groups on relevant characteristics (e.g., covariate or propensity score controls in regression analyses; Amadon et al., 2022; Barnett & Jung, 2021). For example, researchers evaluating New Jersey’s Abbott Preschool Program used propensity score analysis with a host of child and family variables (e.g., child race/ethnicity, maternal education, parental employment) available from state administrative data to examine differences

in achievement and grade retention for participants compared with nonparticipants (Barnett & Jung, 2021). The rigor of these nonrandomized designs for causal inference depends completely on the degree to which the relevant confounding variables are matched or controlled. In rare cases, researchers have access to achievement and other child and family measures taken before entry into pre-k (Cook & Steiner, 2010; Dong & Lipsey, 2018). If the pre-k participant and nonparticipant groups already differ on early achievement before their differential exposure to pre-k, they are quite likely to also differ on their achievement trajectories afterwards, irrespective of that pre-k exposure. Importantly, pretest measures capture a wide variety of contextual, family, and child factors that may influence both selection into pre-k and later achievement in ways that adjusting only for measured family or contextual characteristics cannot.

Most researchers are aware of the threats to internal validity posed by nonequivalent comparison-group designs and may even present such designs as “descriptive” in purpose. However, readers generally view such studies as attempting to convey causal information (e.g., Haber et al., 2022). Nonequivalent comparison group design studies play a major role in the pre-k evaluation literature, likely because of the ease of data collection: administrative data available to school districts can be used to compare students by pre-k attendance, obviating extra testing as well as experimenter manipulation of a pre-k slot. In general, the threat to causal inference here is *positive bias*, whereby children and families with *higher potential outcomes* (e.g., achievement, conscientiousness) *are more likely to participate in pre-k*. Taken to scale, a plurality of nonequivalent comparison group designs in the pre-k literature would systematically alter the quality of the evidence base and lead to incorrect scientific conclusions. As we show in the paper, these studies represent a substantial proportion of the pre-k literature—55% of the studies included in our review of program (i.e., treatment-comparison) effects.

Designing an effective pre-k study that meets the highest standards for causal inference is difficult and therefore rare. Random assignment requires assembling a group of parents prior to pre-k enrollment who are willing to accept the equivalent of a coin flip to determine whether their child can attend pre-k. Identifying and assembling such parents would be logistically very difficult, and few parents—even if on a waitlist—are likely to relinquish their decision-making on this matter and accept a fully random process. Similarly, collecting adequate baseline measures prior to pre-k application or enrollment requires identification in advance of families expected to enroll their children and comparable families not expected to, which is also very difficult to arrange. It would further require identification and measurement of the largely unknown variables that characterize parental decision-making about pre-k participation. As a result of such substantial practical obstacles, most studies of the effects of pre-k fall short of these standards. Still, several pre-k RCTs exist. Whether the high frequency of low-quality causal evidence in the pre-k evaluation literature crowds out resources from additional high-quality evidence is an interesting question outside the scope of this review.

Present Study

We sought to describe and characterize the designs of studies purporting to evaluate publicly funded pre-k programs. While rigor is a comprehensive construct that can vary across studies, depending on implementation, circumstances, and the role of chance, we use well-established methodological standards to code the corpus of pre-k evaluation studies. Random assignment designs, along with well-executed regression discontinuity designs, have high internal validity (What Works Clearinghouse, 2022). Designs with no comparison group and nonequivalent comparison designs without pretests, on the other hand, are known to be vulnerable to many potential threats to internal validity (e.g., Cook et al., 2002).

We also assess whether research using the different designs produced different conclusions about impacts. Statistical effect sizes would be a good way to compare results, but this body of research includes a large proportion of studies reported outside formal journal publication, that lack the statistical details required to derive effect size estimates. With few exceptions, however, the *direction* of the effects (positive or negative) can be determined and indicate the extent to which the reported findings show favorable pre-k effects on the respective outcome measures (Borenstein et al., 2021). The statistical significance of the reported effects was not universally reported, but we include it when available.

Finally, we also wanted to determine the extent to which the various kinds of preschool evaluation studies are cited. We view this as an exploratory measure of how the combination of methodological rigor and the study's findings influences the broader knowledge base. Thus, our systematic review sought to answer three overarching research questions: 1) To what extent are pre-k evaluation studies leveraging designs that offer credible and causally informative estimates of pre-k effectiveness? 2) Do the direction and statistical significance of the reported effects vary by research design, type of outcome, and outcome timing? 3) Which research designs are the most influential in the academic literature?

Method

Defining State or Local Pre-K

We defined pre-k as a state- or locally funded program explicitly labeled “pre-kindergarten” that served 4-year-old children. We thus excluded programs exclusively serving infants and toddlers (although a pre-k program may be housed in the same location as infant and toddler care) and evaluations of Head Start, which is a federally funded program. These exclusion decisions were made a priori by our research team, as we were interested in focusing

our review solely on public pre-k, which currently constitutes the major source of ECE expansion in the U.S. (Moran et al., 2026). The pre-k programs being evaluated could be full- or half-day, universally available or targeted to specific families/children, operating on a full- or school-year calendar, located in a school or community-based setting, and have various quality components (e.g., teacher degree requirements, teacher-child ratios, etc.).

Literature Search Protocol

For our literature search, we attempted to generate an exhaustive list of pre-k evaluation reports that were available online. We divided the search into two main stages: an initial search and a follow-up search. The initial search was conducted in accordance with the NIEER *State of Preschool Report* (Friedman-Krauss et al., 2020), which identified 44 states (plus Washington, D.C. and Guam) that funded pre-k programs in 2019-2020. For each state/US territory that met NIEER's criteria for state-funded pre-k, we attempted to locate all possible evaluations published from 2000 to September 2021.

First, we searched Google and Google Scholar using a combination of terms to reflect the age group of interest (e.g., preschool, pre-k, pre-kindergarten), the name of the state/US territory, and the type of information we sought (e.g., evaluation, evidence, report). For any state that had a pre-k program with a specific name (e.g., New Jersey's Abbott Preschool Program), we employed a similar search strategy whereby we combined the name of the program with the term "evaluation" and "research." Additionally, we explored each state's Department of Education and Health and Human Services (or equivalent) websites for evaluations as well as state government archives for legislative reports. Finally, we searched for programs operating in the three largest cities in each state/US territory using the same search strategies laid out above.

Next, we examined aggregate resources to cross-check our initial search and capture any reports we may have otherwise missed. During this follow-up search, we applied the snowballing method to locate reports from the reference sections of relevant policy briefs, systematic reviews, and meta-analyses. We also used websites and online research libraries (i.e., *Research Connections*) to find public reports on state and local early childhood programs that may be pre-k programs. Overall, this literature search yielded 376 relevant reports.

Determining Report Eligibility

Reports were included in our systematic review if they met the following criteria: a) presented original research findings; b) provided evidence explicitly or implicitly addressing pre-k effects; and c) included child outcomes. For this initial screening process, the inter-rater reliability between coders was 91.4%. All discrepancies were noted, discussed, and resolved by at least one of the primary researchers on the study. Of the 376 preliminary reports, 92 (24.5%) did not pass this initial screening. Figure 1 describes the screening. More information on the coding criteria for this first phase, along with a complete list of reports that passed the eligibility round of coding, is provided in the supplementary material.

Research Design Coding Procedures

A total of 284 reports were included in the second phase of coding. During this coding process, each report was reviewed to identify the research design used to assess pre-k effects and classify it within six categories: (1) no comparison group; (2) nonequivalent comparison with no pretests; (3) nonequivalent comparison with pretests; (4) regression discontinuity design; (5) difference-in-difference; (6) random assignment or lottery-based (see Supplemental Table 1 for more details). A small number of reports presented results from more than one design ($n = 30$; 10% of total sample); for those reports, we coded the design into the most rigorous of the

categories represented. The initial inter-rater agreement between coders for the design coding was 76%, with the discrepancies between coders then resolved by a primary study researcher.

Pre-K Program Effect Coding Procedures

The third stage of coding revolved around reported program effects, with a slightly different methodology. Initial coding of reports was conducted by one or two research assistants, with extensive support and review from the primary study researchers. All reports meeting the inclusion criteria were first coded for whether main effects were reported (yes/no) for the entire sample of pre-k participants (the no comparison group category was handled separately; details below). Reports were coded “no” if findings were only reported for subgroups (e.g., multilingual learners, children receiving special education services, and gender). We also did not include any reports examining moderators (e.g., pre-k program quality) when the main effect of pre-k was not provided.

In many cases, the reports from which program effects were coded were from the same study, which we grouped into a “parent” source study. We defined a study as an evaluation of the same pre-k program by the same research team, for one or more cohorts of participants, with research designs and data collections similar to those used for the same or closely related programs. Studies were considered distinct if they used data from the same pre-k program or program cohorts yet involved different designs (e.g., no comparison group descriptive analyses conducted by a government agency vs. university researchers using an RDD). Similarly, evaluations of the same pre-k program were coded as different studies if they involved separate data collection efforts and were separately reported (e.g., Boston pre-k evaluation from Weiland et al. (2020) compared with the Boston evaluation from Gray-Lobe et al. (2023)). Conversely, annual reports (or near-annual reports) on new cohorts of pre-k attendees from the same auspice

(e.g., state Departments of Education) were coded as a single study. Further, reports using the same evaluation data analyzed by different research teams (e.g., RCT of Tennessee's Voluntary Pre-K Program) represent a single parent study. When appropriate, we provide results at the report and study level.

Our second stage coding designated each report a design category with the highest internal validity. In this third stage of coding, the pre-k effects associated with the highest internal validity design from each study were coded. For example, if a study included reports that fell into the random assignment category and others from a nonequivalent comparison with no pretests, only the random assignment results were coded. Lastly, we coded only the most recent state and/or city report for each study when states or cities reported annual results across several years. We checked the most recent report against previous reports to confirm that no outcomes were excluded.

Together, these criteria resulted in a total of 120 reports from 91 studies retained in our review for the pre-k program effect portion of the study (research question 2). Each of these reports was then coded for the following information: outcome construct, timing of measurement, direction of effects, and reported statistical significance of the effects.

Outcome Construct

The construct represented by each outcome measure was coded as academic-focused (grade level retention, math skills or achievement, reading and language skills or achievement, academic-other), social-emotional learning/behavior-focused (suspension or expulsion, executive functioning, social emotional learning) or other (attendance, college, and other¹). Grade level retention includes whether a child was retained in a grade and/or the percentage of children who

¹ Special education status was coded; however, given questions about whether this is a child-specific characteristic or one that can be directly affected by pre-k, we decided to omit this construct from our effects coding analysis.

were retained. Suspension/expulsion includes the frequency of suspensions and/or expulsions as well as disciplinary infractions. College includes college prerequisites and college enrollment. For the academic-other category, the specific construct was noted. In addition to the outcome construct, we reported whether the outcome was a direct assessment, teacher report, parent report, self-report, observation (researcher), or from administrative data, and recorded the name of the associated measure. More information on these categories is provided in the supplementary materials.

Timing

The timing of the measurement of each outcome was coded into one of 20 possible categories oriented to the respective school grade year (e.g., end of pre-k, fall of kindergarten, first grade, fourth grade, etc.). We also included the option for multiple grades or school levels (e.g., elementary or middle school) because some reports provided composite results that spanned multiple years. For our analyses, we defined end-of-program effects as measured immediately following pre-k and longer-term effects as measured in the spring of kindergarten and beyond².

Direction of Effect

Statistical effect sizes that encode both the direction and magnitude of treatment-control differences would have provided the best representation of pre-k effects. However, the standard deviations for the outcome variables required to calculate those were not reported in many of the studies. In such circumstances, the direction of the effect, whether the outcome measure favors the treatment or control group, can be used to assess the nature of the findings in a systematic

² Note that some of the research designs included in our review do not allow for the measurement of both end-of-program and longer-term outcomes. For example, no end-of-pre-k effects were reported for difference-in-difference designs, and no longer-term effects can be reported for age-cutoff RDDs.

review (Borenstein et al., 2021). The direction of the effects was almost universally evident in the studies reviewed and was coded as positive, negative, or null. Null was only coded if the outcome measures had the same value for the relevant comparison group (e.g., same means, precisely zero regression coefficient). Direction of effects was reverse-coded when necessary towards the favored or expected result on the outcome measure. For example, if those who attended pre-k were less likely to be retained in a grade than those who did not, it is a favorable outcome for the treatment condition even though it was numerically negative in direction.

Statistical Significance of Effect

We also coded whether each effect was reported as statistically significant, although that was often not reported. While vote counting statistical significance is not a good indication of the nature of the effects in a body of research (Hedges & Olkin, 1980), comparing the proportion of statistically significant effects with the proportion tested but nonsignificant provides some useful information beyond the direction-of-effect proportions. In particular, statistical significance is a function of both sample size and the magnitude of the effect. Similar results for direction-of-effects and statistical significance (yes/no), when those can be compared, provides some assurance that the direction of effects results do not stem from a surfeit of very small differences in studies with very small samples. In coding statistical significance, we did not make any independent judgment about the appropriateness of the statistical significance reported (e.g., the p -value threshold, the statistical test used). We simply coded whether the researchers designated an outcome as statistically significant. Outcomes for which the report did not provide a significance level were coded as “not reported.”

Coding Reliability Check

Because the statistical significance and effect direction coding were done by one coder, a second research assistant independently coded reports across all research design categories, except the large category of nonequivalent comparisons with no pretests. For that category, 25% of the reports were coded as a reliability check. Of these double-coded reports, there was 82% agreement. In cases of a disagreement regarding direction of effect and/or significance, the coders then discussed and resolved via mutual agreement.

Citation Count

To generate the citation count measure, two coders independently searched for each included report through Google Scholar. Citation counts were extracted by both coders on the same day (June 19th, 2024) to ensure that the search engine was pulling from a consistent body of literature. First, the exact report title was entered into Google Scholar and the number of citations was recorded. For non-scholarly works and reports that were not immediately found, the search was conducted again using variations of the report's title (e.g., a report entitled "Assembly Bill 563: Nevada early childhood education (ECE) program: Building a foundation for school readiness and success in k-12 and beyond: FY 2010-11 evaluation report" was searched for using "Building a foundation for school readiness and success in k-12 and beyond: FY 2010-11" or "Building a foundation for school readiness and success in k-12 and beyond"). If the same report spawned multiple entries on Google Scholar, citations from these entries were aggregated to reflect the total number of times the report's contents were cited. Moreover, report citation counts were recorded as "true zero" if they appeared on Google Scholar with no observable citations, and "missing" if the report was not located following multiple search attempts. Final citation counts were cross-checked by the two coders, and any existing discrepancies were resolved.

Analysis

Research Design

We conducted simple descriptive analyses—counts and percentages—to determine pre-k evaluation report year and type (e.g., journal article, research brief, presentation). Similar descriptive statistics were computed to describe research designs.

Effect Direction and Statistical Significance

Analyses with these data focused on whether the direction and significance of effects differed by the research design. First, we examined the proportion of outcomes for which the direction of effects was positive or negative (very few were exactly null; these were included with the negative ones). Direction was represented as a binary value, 1 when the outcome measure favored the treatment group (positive) and 0 when it favored the control group or neither (negative or null). Analysis investigated whether the proportion of positive effects for various groups of studies was greater or less than the .50 probability expected under the null hypothesis. These analyses were conducted for the studies in each research design category (excluding the no comparison group category), by outcome construct (achievement-related vs. other), and by the timing of the outcome measurement (end-of-pre-k vs. spring of kindergarten and beyond).

Similar analyses were conducted for statistical significance when significance test results were reported. Binary codes differentiated reported significance (coded 1) from reported nonsignificance (coded 0). The few studies with no direction of effects, i.e., identical outcomes for treatment and control groups, were also coded as nonsignificant on the logic that, on their face, it was mathematically impossible for such results to be statistically significant. Studies for

which the statistical significance of the effects was not reported or evident from precisely null results were omitted from these analyses.

The number of effects for these analyses varied widely across the source reports, and there were varying numbers of reports for the different studies. To better balance the potential distortion in descriptive comparisons stemming from these uneven numbers, the proportions of positive and negative effects were first determined within each parent study with the means of those proportions across studies used in further analysis, thus making study the unit of analysis irrespective of the number of reports or effects for each study. For a more refined analysis, the binary values were used as dependent variables in multilevel regression models with effects nested in reports and reports nested in studies. These models allowed for statistical significance testing of the differential proportions of positive effects across the different research design categories.

Results

Overall, this systematic review included data coded from 284 reports about pre-k effectiveness. Table 1 describes the types of reports included and the years in which they were released. Most of these documents were identified as research reports (76%), followed by academic journal articles (15%). Overall, we observed an increase in the availability of study findings over time, with 46% of studies made public in or after 2015.

Designs Used to Evaluate Pre-K Programs

Table 2 provides counts and percentages of studies for each research design category in the effect coding sample (120 reports from 91 studies). More than half of the studies were coded as using nonequivalent comparison with no pretest designs (55%). The same pattern holds true at the report level. In contrast, only 4% of the studies used a randomized design to evaluate a pre-k

program. Further, the five reports on randomized designs drew from only four studies, including two lottery-based evaluations of the Boston pre-k program, one RCT of Tennessee's Voluntary Pre-K Program, and one RCT of North Carolina's pre-k program. Not included in the analysis are the 72 reports from 10 studies using the no-comparison group research method, representing 25% of reports and 10% of studies in the full review sample.

Direction of Pre-K Effects

Table 3 reports the number of effects coded for direction and the number of parent studies by construct (achievement vs. other), timing (end of pre-k vs. longer-term follow-up), and type of design. Overall, direction was coded for 901 effects. A total of 297 end of program effects were reported (spring of pre-k or fall of kindergarten) along with 604 longer-term effects (spring of kindergarten and beyond). Most of these effects (61%) are from nonequivalent comparisons with no pretests, and only 16% are from random assignment designs. Most of the effects coded are for achievement outcomes (75%).

Table 4 shows the direction of the effects (favors pre-k vs. favors control/neither) for the achievement outcomes broken down by the timing of outcome measurement and the design category. For these descriptive comparisons, the proportions of the positive and negative/null effect sizes are first determined for each study to allow each study to contribute equally to the overall mean values presented. (Supplemental Table 2 provides a more detailed breakdown that shows the small number of null effects as a separate category).

Two notable differences appear in Table 4. First, the overall proportion of favorable end-of-pre-k achievement effects was generally higher across the different designs than those measured in later years, consistent with the fade-out of achievement effects often reported in pre-k studies. Second, end-of-pre-k achievement effects were consistent across the designs in the

proportion favorable. By contrast, for longer-term outcomes, the proportion of favorable achievement effects was lower for randomized designs (35%) than for the other three designs (75 to 84%). Further information on the direction of effects can be found in Supplemental Tables 2 and 3.

Figures 2 and 3 summarize key data from Tables 3 and 4. For each design category, the full span of the bar across the x-axis represents the number of studies in that category. These show, for example, how many nonequivalent comparisons with no pretests, and how few random assignment studies appear in the body of research on pre-k effects. The proportions of positive and negative/null effects from Table 4 are then superimposed on those bars to show the relative differences in the findings across the design categories. These describe, for example, the rather consistent positive effects for end-of-pre-k outcomes and the much greater contrast with randomized designs for longer-term effects.

The multilevel regression analyses put a finer point on the differences visible in Figures 2 and 3. For achievement outcomes at the end of pre-k (51 studies, 208 effects), the omnibus test of differences in direction of effects across the design categories showed no significant variation ($F=1.92$ (3, 47), $p=.139$), which indicates no significant differences by design category for end-of-pre-k achievement effects. Further, there were no significant differences in effect direction for any of the individual design types compared against random assignment as the reference category. These results show the same pattern of effects as the descriptive analysis and further demonstrate the finding of largely positive effects across designs for end-of-pre-k outcomes.

For longer-term achievement effects (50 studies, 468 effects), the multilevel regression analyses also yielded results like those observed in the descriptive analyses. The omnibus test of differences across the four designs that addressed longer-term effects found significant variation

($F=5.34$, (3, 52), $p=.003$). In particular, the nonequivalent comparisons without pretests and difference-in-difference designs had significantly higher proportions of positive effects than the random assignment reference category. (Supplemental Tables 4a and 4b provide more detail for the results of these regression analyses).

Statistical Significance of Effects

Of the 901 effects coded for direction, statistical significance was available for 754 (84%) and, of those, 377 (50%) were reported to be significant. Statistical significance is not a very good indicator of the magnitude of an effect because it is highly dependent on sample size. In this context, however, the pattern of significant positive and negative findings does provide some check on the possibility that the direction of difference results reported above involves a large proportion of trivially small differences that exaggerate the extent to which the different designs yield meaningfully different pre-k effect estimates. (Supplemental Table 3 provides a detailed breakout of the reported statistical significance findings.)

Table 5 displays the direction of only the effects reported as statistically significant, broken out by research design, outcome, and timing. While the number of effects is much smaller, the pattern for the proportions favoring pre-k or the control conditions is much the same as that in Table 4 for all effects. Most notably, the significant effects for longer term outcomes show the same contrast between random assignment and the other designs with the proportions favoring pre-k in the randomized designs considerably smaller for both achievement and other outcomes. The multilevel regression analyses indicated that the difference was statistically significant for the contrast with no pretest designs and difference-in-difference designs (Supplemental Tables 5a and 5b).

Additional Lottery-Design Pre-K Studies

After the conclusion of our systematic review report screening, a handful of additional lottery-designed pre-k evaluations were published. Because our original sample included so few studies in this design category, we coded the effects reported in these documents. The list of these reports and the table showing the direction of effects are in Supplemental Table 6.

These results are strikingly similar to our main findings: While only 3 effects were found for the end of treatment, all were in the expected positive direction. For longer-term outcomes, only 32% (out of 62 effects) of the achievement effects and 25% (out of 8 effects) of the other outcomes were in the positive direction.

Citation Counts of Pre-K Studies

A total of 158 reports from 79 studies were found on Google Scholar. After compiling the Google Scholar citation data for each report, we found that on average, an age cutoff RDD paper was cited the most frequently, but with wide variation across reports (average of 130.2 times, SD = 267.4; Table 6). Similarly, the age cutoff RDD category in total had over 5,200 citations (Figure 4), which was the highest citation count of any of our research categories. In contrast, the no comparison research design group had the lowest average citation count (9.2; SD = 10.5) and the fewest overall citations, with under 260.

Additional Analyses

Because we observed a large number of studies in the no comparison group category, we provide some additional information on their design features. First, we coded each report ($n = 74$) for the type of comparison being made. A total of 53 reports compared pre- and post-test scores of children attending pre-k, and 44 reports compared scores for children attending pre-k to a norm reference, target scores/percentages, or quartile percentages. Categories are not mutually exclusive; reports may have fallen into either or both of the above categories. Next, we coded a

random sample of 25% of the reports included in this category. Of the 19 reports coded, 18 reported a positive direction of effects at the end of pre-k or the fall of kindergarten. Of the 123 effects reported, 61% represented growth in pre-k children's academic skills (e.g., math, literacy, and language). A total of 66 statistically significant findings were reported at the end of pre-k or fall of kindergarten, and the majority (95%) were positive effects. Only two reports included longer-term effects.

Discussion

In our systematic review, we attempted to determine the extent to which different research designs are used to evaluate pre-k programs and whether the reported end-of-program and longer-term effects vary by design. Our results indicate that most pre-k evaluations use designs with internal validity that is too low to support confident conclusions about causal program impacts. Most notably, at the time of our review, only four pre-k studies used an RCT or lottery design, and two of those studied the same program. Furthermore, these designs were not among the most frequently cited in the academic literature. The nonequivalent comparison designs were cited more than the lottery or RCT evaluations, suggesting our most rigorous studies are not the most influential. We find significant differences across research designs in whether results are positive for longer-term academic outcomes. Specifically, randomized designs result in significantly more negative effects than weaker research designs. Our results are consistent with the inclusion of the most recent lottery-designed pre-k evaluations.

The overwhelming use of nonequivalent comparison group designs (May & Collier, 2023) reflects, in part, the challenges of rigorous evaluations of pre-k programs, including the costs involved in administering academic and social-emotional measures prior to the start of the pre-k year, and in following children into later school settings. Further, pre-k is likely the first

educational program many children participate in that is state or locally funded, thus establishing baseline equivalence is particularly difficult given the lack of prior administrative data on participants. Indeed, many state reports and academic journal articles report matching pre-k and non-participants on demographic characteristics from administrative data in early elementary school. Yet, as our results suggest, such designs do not produce longer-term effects that are similar to effects reported from studies that ensure baseline equivalence through randomization.

In discussions about this report with the research community, we have heard the argument that state governments require evaluations, so we should not be surprised by the existence of these studies, as this is “simply how the government works.” We see this view as needlessly pessimistic. Of course, it is unrealistic to expect excellent well-powered RCTs to be redone each year for each program, and ongoing descriptive work on program quality inputs and processes is likely useful. However, our review raises an important question: whether a research design may be so weak that a study is in danger of doing more harm than good to evidence-based policymaking. As our review has documented, the quality of evaluations is far too low to be useful but is widespread enough to cause concern by way of their opportunity costs (i.e., the time evaluators could be dedicating to the more difficult work of obtaining credible causal estimates; the information crowded out by dozens of reports that fail to do so). For example, by conducting a no-comparison or nonequivalent comparison-group study to “show” that pre-k is effective, rather than an implementation study of a program's quality, stakeholders can miss key weaknesses in their programs, inequities across sites or populations served, or other programmatic challenges that frequently occur during massive scale-up. Such pressure for research could lead stakeholders to produce “evaluation” data without meaningfully engaging in whether and how the data inform our knowledge of pre-k effectiveness and supporting children’s

well-being. Further, given the current federal research climate, questions remain about whether and how future pre-k evaluations will change (e.g., even fewer lottery or RCT studies without federal funding) and what this may mean for the state of pre-k evidence on children's short- and long-term development.

Researchers have identified key limitations and challenges, along with possible solutions, in using lottery-based or random-assignment designs to evaluate pre-k programs (Weiland et al., 2024). We believe these concerns are important and the potential solutions provided (e.g., additional data checks to determine how a lottery sample may differ from overall program sample, using first choice lotteries) are necessary to consider when designing evaluations. However, without strong internal validity, studies using representative samples of pre-k attendees provide little basis for expanding policies, given the very strong likelihood of selection bias in the results. If the question of interest is causal, as the pre-k effectiveness question clearly is, then discussions about external validity are only worth having after internal validity has been achieved (Campbell & Stanley, 1963). Moreover, the lottery design studies reviewed here were not conducted in laboratories with highly controlled treatments that have little basis in policy reality. Rather, they were large-scale evaluations of real-world programs serving the populations of students the programs intended to target (e.g., Lipsey et al., 2018).

Pre-k studies using RDDs constitute another category of evaluations with strong internal validity and may be easier to implement. Our citation counts results also suggest these designs are influential in the pre-k literature—they constitute the most cited research design. This is encouraging given the strong internal validity of the design. However, because these designs rely on child birthdate as the forcing variable, they are limited to investigating immediate program outcomes (e.g., Wong et al., 2018), and end-of-program impacts appear to be non-controversial

given that almost all study designs find positive effects at the end of preschool. Thus, to make progress on the longer-term effects of pre-k, which are essential to document for cost-benefit analyses, other designs are necessary.

Our results highlight mixed findings on academic outcomes in the longer run and point to the need for more research in this area. Specifically, we echo calls from others (Abenavoli, 2019; Burchinal et al., 2024; Whitaker et al., 2026) for more rigorous longitudinal research on pre-k programs, which is especially important to test potential mediators and other theories like the sustaining environments hypothesis. The use of lottery-based admissions to such programs provides an important avenue for future research, as other scholars have noted (Bruhn & Emick, 2023) and begun to capitalize on (e.g., Woodyard et al., 2025). If demand for pre-k remains strong and enrollment continues to increase (Friedman-Krauss et al., 2026), there is strong potential for lotteries and for researchers to use administrative data to examine pre-k effectiveness at the end of the program and in the longer term. These lottery designs have the potential to follow children beyond K-12 and examine college, career, and well-being outcomes, much like Gray-Lobe et al. (2023) examined with Boston data.

Finally, we hope that our results will serve as a motivating case for researchers who conduct ECE evaluations to strive for quality over quantity. As our review reflects, the field is not lacking in less internally valid studies that provide, at best, descriptive information on children who attend public pre-k programs. They are not likely to advance public knowledge. Instead, we sorely need high-quality studies with strong internal validity that can answer the question at the crux of research-based advocacy for public pre-k expansion: do pre-k programs benefit children in the long-term? Some might react to this argument with skepticism, arguing instead that the level of evidence that we currently have is good enough for continued policy

advocacy (Phillips et al., 2024). After all, most children attend second grade, and we do not have *any* high-quality evaluations of second grade attendance. Why must we put such an onerous research burden on public pre-k?

This rhetorical argument could be used to justify additional public spending on *any* program. The answer is that we are not aware of any ongoing academic policy debates that hinge on whether second grade provides potentially life-altering benefits in the long-term. As long as researchers continue to support advocacy for expanding public pre-k, the validity of the research literature will remain under scrutiny. Millions of government grant dollars have been spent to evaluate the efficacy of public pre-k because advocates have claimed that investments will pay off in the long-term (e.g., Heckman, 2006; Council of Economic Advisers, 2023)—a claim that warrants serious empirical investigation. Despite this substantial investment, the field has yet to definitively answer fundamental questions, and the continued preponderance of weak research designs will not help us move our understanding forward. If we want to make strong claims about the importance of pre-k that would convince a reasonable skeptic, then we need stronger evidence.

Conclusion

Our large-scale systematic review of research designs used to evaluate publicly funded pre-k programs, which includes 120 reports from 91 studies, demonstrates that most pre-k effectiveness evaluations use designs with weak internal validity. Further, those designs produce significantly different, more positive longer-run effects in terms of the direction of effect compared with RCT and lottery-based studies. The significant difference in longer-run direction of effects requires the pre-k and broader education field to seriously consider what this means for our knowledge base on pre-k effectiveness. Much of the justification for expansion of pre-k

programs is based on early demonstration programs (e.g., Perry Preschool and Abecedarian) that had consistent effects on participants' development into adulthood (Burchinal et al., 2024). Much more limited evidence exists on the longer-run effects of modern-day, scaled-up programs. More rigorous research on current pre-k programs, coupled with additional information that allows investigation of different pre-k models can help researchers make sense of seemingly contradictory findings (Whitaker et al., 2026). Results from our study should give pause to researchers championing nonequivalent comparison group research designs as causal and to practitioners and policymakers considering implementing sweeping practice and policy changes based on such results.

References

- Abenavoli, R. M. (2019). The mechanisms and moderators of “fade-out”: Towards understanding why the skills of early childhood program participants converge over time with the skills of other children. *Psychological Bulletin*, 145(12), 1103-1127. <http://dx.doi.org/10.1037/bul0000212>
- Amadon, S., Gormley, W. T., Claessens, A., Magnuson, K., Hummel-Price, D., & Romm, K. (2022). Does early childhood education help to improve high school outcomes? Results from Tulsa. *Child Development*, 93(4), e379-e395. <https://doi.org/10.1111/cdev.13752>
- Arkansas Research Center. (n.d.). *Research*. Arkansas Research Center, University of Central Arkansas. <https://arc.arkansas.gov/arcweb/research.html>
- Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness*, 10(1), 7-39. <https://doi.org/10.1080/19345747.2016.1232459>
- Barnett, W. S., & Jung, K. (2021). Effects of New Jersey’s Abbott preschool program on children’s achievement, grade retention, and special education through tenth grade. *Early Childhood Research Quarterly*, 56(1), 248–259. <https://doi.org/10.1016/j.ecresq.2021.04.001>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to meta-analysis* (2nd ed.). Wiley.
- Brooks-Gunn, J., Markman-Pithers, L., & Rouse, C. E. (2016). Starting early: Introducing the issue. *The Future of Children*, 3-19.
- Brodeur, A., Cook, N., & Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, 110(11), 3634-3660.
- Bruhn, J., & Emick, E. (2023). *Lottery evidence on the impact of preschool in the United States: A review and meta-analysis*. (No. 2023.20). Blueprint Labs. https://blueprintcdn.com/wp-content/uploads/2023/12/MIT-Blueprint-Discussion-Paper_Preschool-Research.pdf
- Burchinal, M., Whitaker, A., Jenkins, J., Bailey, D., Watts, T., Duncan, G., & Hart, E. (2024). Unsettled science on longer-run effects of early education. *Science*, 384(6695), 506-508.
- Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, 112(3), 579-620.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*.
- Cascio, E. U. (2023). Does universal preschool hit the target?: Program access and preschool impacts. *Journal of Human Resources*, 58(1), 1-42. <http://www.nber.org/papers/w23215>
- Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference* (Vol. 1195). Houghton Mifflin.
- Cook, T. D., & Steiner, P. M. (2010). Case matching and the reduction of selection bias in quasi-experiments: The relative importance of pretest measures of outcome, of unreliable measurement, and of mode of data analysis. *Psychological Methods*, 15(1), 56-68. <https://doi.org/10.1037/a0018536>
- Council of Economic Advisers. (2023). Chapter 4: Investing in young children's care and education. In Economic report of the President (pp. 125–152). U.S. Government Publishing Office. <https://www.govinfo.gov/content/pkg/ERP-2023/pdf/ERP-2023.pdf>
- Courtney, J. R., Garcia, J. T., Rowberry, J., Eckberg, N., Dinces, S. M., Lobaugh, C. S., & Tolman, R. T. (2023). Measuring impact of New Mexico prekindergarten on standardized

- test scores and high school graduation using propensity score matching. *International Journal of Child Care and Education Policy*, 17(1), 1-18. <https://doi.org/10.1186/s40723-023-00112-9>
- Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press.
- DeAngelis, C. A., Holmes Erickson, H., & Ritter, G. W. (2020). What's the state of the evidence on pre-K programmes in the United States? A systematic review. *Educational Review*, 72(4), 495-519. <https://doi.org/10.1080/00131911.2018.1520688>
- Dong, N., & Lipsey, M. W. (2018). Can propensity score analysis approximate randomized experiments using pretest and demographic information in pre-k intervention research? *Evaluation Review*, 42(1), 34-70. <https://doi.org/10.1177/0193841X17749824>
- Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives*, 27(2), 109-132.
- Durkin, K., Lipsey, M. W., Farran, D. C., & Wiesen, S. E. (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Developmental Psychology*, 58(3), 470-484. <https://doi.org/10.1037/dev0001301>
- Early, D. M., Li, W., Maxwell, K. L., & Ponder, B. D. (2019). Participation in Georgia's pre-k as a predictor of third-grade standardized test scores. *AERA Open*, 5(2), 1-16. <https://doi.org/10.1177/2332858419848687>
- Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G., & Gardiner, B. A. (2020). *The state of preschool 2019: State preschool yearbook*. National Institute for Early Education Research. https://nieer.org/sites/default/files/2023-09/YB2019_Full_Report.pdf
- Friedman-Krauss, A., Barnett, W. S., Hodges, K. S., Garver, K., Weisenfeld, G. G., Duer, J., & Siegel, J. (2026). *The state of preschool 2025*. National Institute for Early Education Research. <https://nieer.org/sites/default/files/2026-04/2025%20NIEER%20Yearbook.pdf>
- García, J. L., Heckman, J. J., & Ziff, A. L. (2019). Early childhood education and crime. *Infant Mental Health Journal*, 40(1), 141-151. <https://doi.org/10.1002/imhj.21759>
- Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-k program. *Journal of Human Resources*, 40(3), 533-558.
- Gormley, W. T., Phillips, D., & Anderson, S. (2018). The effects of Tulsa's pre-k program on middle school student performance. *Journal of Policy Analysis and Management*, 37(1), 63-87. <https://doi.org/10.1002/pam.22023>
- Gormley Jr, W. T., Phillips, D. A., Newmark, K., Welti, K., & Adelstein, S. (2011). Social-emotional effects of early childhood education programs in Tulsa. *Child Development*, 82(6), 2095-2109. <https://doi.org/10.1111/j.1467-8624.2011.01648.x>
- Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2023). The long-term effects of universal preschool in Boston. *The Quarterly Journal of Economics*, 138(1), 363-411. <https://doi.org/10.1093/qje/qjac036>
- Haber, N. A., Wieten, S. E., Rohrer, J. M., Arah, O. A., Tennant, P. W., Stuart, E. A., ... & Fox, M. P. (2022). Causal and associational language in observational health research: A systematic evaluation. *American Journal of Epidemiology*, 191(12), 2084-2097
- Hahn, R. A., & Barnett, W. S. (2023). Early childhood education: Health, equity, and economics. *Annual Review of Public Health*, 44, 75-92. <https://doi.org/10.1146/annurev-publhealth-071321-032337>
- Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782), 1900-1902. <https://doi.org/10.1126/science.1128898>

- Hedges, L. V., & Olkin, I. (1980). Vote counting methods in research synthesis. *Psychological Bulletin*, 88(2), 359–369. <https://doi.org/10.1037/0033-2909.88.2.359>
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), 945–960. <https://doi.org/10.1080/01621459.1986.10478354>
- Kitchens, K. E., & Gormley, W. (2023). From preschool to politics: Early socialization in Tulsa. *Early Childhood Research Quarterly*, 62, 259–274. <https://doi.org/10.1016/j.ecresq.2022.09.004>
- Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, 45, 155–176. <https://doi.org/10.1016/j.ecresq.2018.03.005>
- 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. <https://doi.org/10.3102/0162373714547266>
- May, H., & Collier, Z. K. (2023). Nonequivalent comparison group designs. In H. Cooper, M. N. Coutanche, L. M. McMullen, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology: Research designs: Quantitative, qualitative, neuropsychological, and biological* (2nd ed., pp. 629–651). American Psychological Association. <https://doi.org/10.1037/0000319-028>
- McCormick, M., Weiland, C., Hsueh, J., Pralica, M., Weissman, A. K., Moffett, L., ... & Sachs, J. (2021). Is skill type the key to the preK fadeout puzzle? Differential associations between enrollment in preK and constrained and unconstrained skills across kindergarten. *Child Development*, 92(4), e599–e620. <https://doi.org/10.1111/cdev.13520>
- Meloy, B., Gardner, M., & Darling-Hammond, L. (2019). *Untangling the evidence on preschool effectiveness: Insights for policymakers* (Research brief). Learning Policy Institute. https://learningpolicyinstitute.org/media/302/download?inline&file=Untangling_Evidence_Preschool_Effectiveness_BRIEF.pdf
- Moran, C., Rosengarten, M. L., Cha, E., Jenkins, J. M., Dodge, K. A., Friedman-Kraus, A., Barnett, S. W., & Watts, T. W. (2026). 23 Years of State Pre-K Expansion: How States Balance Investment, Access, and Quality in the Modern Era. (EdWorkingPaper: 26-1541). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/ex9a-ps25>
- Obama, B. (2013, February 12). *Remarks by the President in the State of the Union address*. The Obama White House Archives. <https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/remarks-President-state-union-address>
- Peisner-Feinberg, E., Kuhn, L., Zadrozny, S., Foster, T., & Burchinal, M. (2020). *Kindergarten follow-up findings from a small-scale RCT study of the North Carolina Pre-Kindergarten Program*. The University of North Carolina at Chapel Hill.
- Phillips, D., Bassok, D., Clements, D., Dodge, K., Duke, N., Espinosa, L., Gibbs, C., Johnson, A., Morris-Perez, P., Stipek, D., Yoshikawa, H., Zaslow, M., & Weiland, C. (2024). Unsettled but actionable. *Science*, 384(6695). <https://www.science.org/doi/10.1126/science.adn2141>
- Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., & Weiland, C. (2017). Puzzling it out: The current state of scientific knowledge on pre-k effects: A consensus statement. *The current state of scientific knowledge on pre-kindergarten effects*, 19–30.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688–701.

<https://doi.org/10.1037/h0037350>

- Watts, T. W., Jenkins, J. M., Dodge, K. A., Carr, R. C., Sauval, M., Bai, Y., Escueta, M., Duer, J., Ladd, H., & Muschkin, C. (2023). Understanding heterogeneity in the impact of public preschool programs. *Monographs of the Society for Research in Child Development*, 88(1), 7-182. <https://doi.org/10.1111/mono.12463>
- Weiland, C., McCormick, M., Duer, J., Friedman-Krauss, A., Pralica, M., Xia, S., ... & Mattera, S. (2024). The mixed-delivery pre-k opportunity gap? Differences in demographics, quality, and children's gains in community-based versus public school programs across five large-scale systems. *Early Childhood Research Quarterly*, 68, 247-259.
- Weiland, C., Unterman, R., Dynarski, S., Abenavoli, R., Bloom, H., Braga, B., Faria, A.-M., Greenberg, E., Jacob, B. A., & Lincove, J. A. (2024). Lottery-based evaluations of early education programs: Opportunities and challenges for building the next generation of evidence. *AERA Open*, 10(1), 1-23. <https://doi.org/10.1177/23328584241231933>
- Weiland, C., Unterman, R., Shapiro, A., Staszak, S., Rochester, S., & Martin, E. (2020). The effects of enrolling in oversubscribed prekindergarten programs through third grade. *Child Development*, 91(5), 1401-1422. <https://doi.org/10.1111/cdev.13308>
- Weiland, C., Wu, T., Unterman, R., Shapiro, A., Lightner, S., Staines, T., & Taylor, A. (2025). *Impacts of oversubscribed Boston pre-k programs through middle school*. (EdWorkingPaper: 25-1194). Annenberg Institute at Brown University. <https://doi.org/10.26300/4n08-s025>
- What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Whitaker, A. A., Burchinal, M., Jenkins, J., M., Bailey, D. H., Watts, T. W., Duncan, G. J., & Hart, E. R. (2025). *Why are preschool programs becoming less effective?* *Journal of Policy Analysis and Management*, 45(1). <https://doi.org/10.1002/pam.70031>
- The White House. (2024, March 7). Remarks by President Biden in State of the Union Address. <https://www.whitehouse.gov/briefing-room/speeches-remarks/2024/03/08/remarks-by-president-biden-in-state-of-the-union-address-3/>
- Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, 27(1), 122-154. <https://doi.org/10.1002/pam.20310>
- Woodyard, H., Sass, T., & Fazlul, I. (2025). Assessing the benefits of education in early childhood: Evidence from a Pre-K lottery in Georgia. *Economic Inquiry*, 63(3), 663-680.
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., ... & Zaslow, M. J. (2013). Investing in our future: The evidence base on preschool education. *Society for Research in Child Development*. <http://www.srce.org/policy-media/policy-updates/meetings-briefings/investing-our-future-evidence-base-preschool>

Table 1*Characteristics of Reports and Studies Included in the Systematic Review*

	Study Count	Percentage	Report Count	Percentage
Document Type*				
Journal article	25	22.73%	42	14.79%
Research report/brief/abstract/executive summary	78	70.91%	215	75.70%
Presentation	5	4.55%	6	2.11%
Conference proposal	5	4.55%	6	2.11%
Dissertation/thesis	7	6.36%	7	2.46%
Fact sheets/infographics	1	0.91%	1	0.35%
Other	7	6.36%	7	2.46%
Publication Year				
2000-2004	8	7.27%	45	15.85%
2005-2009	19	17.27%	74	26.06%
2010-2014	25	22.73%	76	26.76%
2015-2021	51	46.36%	80	28.17%
Unknown	7	6.36%	9	3.17%

Note. Reports are products (e.g., published papers in academic journals, “gray literature,” and publicly available presentation slide decks from studies). Studies are a systematic evaluation or analysis of whether a pre-k program is effective and may be represented more than once in the document type analysis. We concluded our systematic review in September 2021; thus, data for this year (2021) is incomplete as more reports may have been made available in the later months.

Table 2*Distribution of Design Types*

Category	Studies	Reports	Effects
Nonequivalent comparison with no pretest	50 (55%)	76 (63%)	549 (61%)
Nonequivalent comparison with child pretest data	6 (7%)	7 (6%)	58 (6%)
Age-cutoff regression discontinuity	25 (27%)	26 (22%)	126 (14%)
Difference-in-difference	6 (7%)	6 (5%)	24 (3%)
Random assignment	4 (4%)	5 (4%)	144 (16%)
Total	91	120	901

Note. Studies are a systematic evaluation or analysis of whether a pre-k program is effective. Reports are products (e.g., published papers in academic journals, gray literature, and publicly available presentation slide decks from studies).

Table 3*Distribution of Design Types by Outcome Construct and Timing*

	End of Program		Longer-Term	
	Studies ^a	Effects	Studies ^a	Effects
Achievement Outcomes				
Nonequivalent comparison with no pretests	18 (35%)	52 (25%)	38 (76%)	350 (75%)
Nonequivalent comparison with pretests	6 (12%)	38 (18%)	2 (4%)	15 (3%)
Age-cutoff regression discontinuity^b	25 (49%)	106 (51%)	-	-
Difference-in-Difference	0	0	6 (12%)	24 (5%)
Random assignment	2 (4%)	12 (6%)	4 (8%)	79 (17%)
Total	51	208	50	468
Other Outcomes (e.g., behavioral, SEL)				
Nonequivalent comparison with no pretests	18 (62%)	54 (61%)	18 (86%)	93 (68%)
Nonequivalent comparison with pretests	1 (3%)	5 (6%)	0	0
Age-cutoff regression discontinuity^b	8 (28%)	20 (23%)	-	-
Difference-in-Difference	0	0	0	0
Random assignment	2 (7%)	10 (11%)	3 (14%)	43 (32%)
Total	29	89	21	136

Note. Researchers coded each individual report for the most rigorous design included, excluding the no comparison group studies. We define “longer-term” as any time point after the end of pre-k/fall of kindergarten. Academic outcomes include those representing subject matter mastery. “Other” outcomes include those representing social emotional characteristics, behavioral choices, and the like. Column percentages are within each outcome type (e.g., academic, other outcomes). ^a “Studies” indicates the number of studies that reported effects in the respective outcome category. Many studies produced effects in more than one of those categories and those studies are thus counted more than once. ^b Longer-term outcomes cannot be examined in age-cutoff regression-discontinuity designs.

Table 4*Distribution of Design Types by Outcome Type, Timing, and Direction of Effects*

Achievement Outcomes						
	End of Program			Longer-Term		
	N of Studies (Effects)	Effects Favoring Pre-K	Effects Favoring Control/Neither	N of Studies (Effects)	Effects Favoring Pre-K	Effects Favoring Control/Neither
Nonequivalent comparison with no pretests	18 (52)	88.2%	11.8%	38 (350)	84.0%	16.0%
Nonequivalent comparison with pretests	6 (38)	73.3%	26.7%	2 (15)	75.0%	25.0%
Age-cutoff regression discontinuity	25 (106)	97.5%	2.5%	-	-	-
Difference-in-Difference	0	-	-	6 (24)	81.7%	18.3%
Random assignment	2 (12)	90.0%	10.0%	4 (79)	34.7%	65.3%
Other Outcomes (e.g., Behavioral, SEL)						
Nonequivalent comparison with no pretests	18 (54)	77.9%	22.1%	18 (93)	73.5%	26.5%
Nonequivalent comparison with pretests	1 (5)	80.0%	20.0%	0	-	-
Age-cutoff regression discontinuity	8 (20)	68.8%	31.2%	-	-	-
Difference-in-Difference	0	-	-	0	-	-
Random assignment	2 (10)	100.0%	0.0%	3 (43)	39.0%	61.0%

Note. To avoid distortions stemming from the different number of effects reported in different studies, the effect proportions within each study were first computed, then those proportions were averaged over the studies contributing effects in the respective category.

Table 5*Distribution of Design Types by Outcome Type for the Direction of the Statistically Significant Effects*

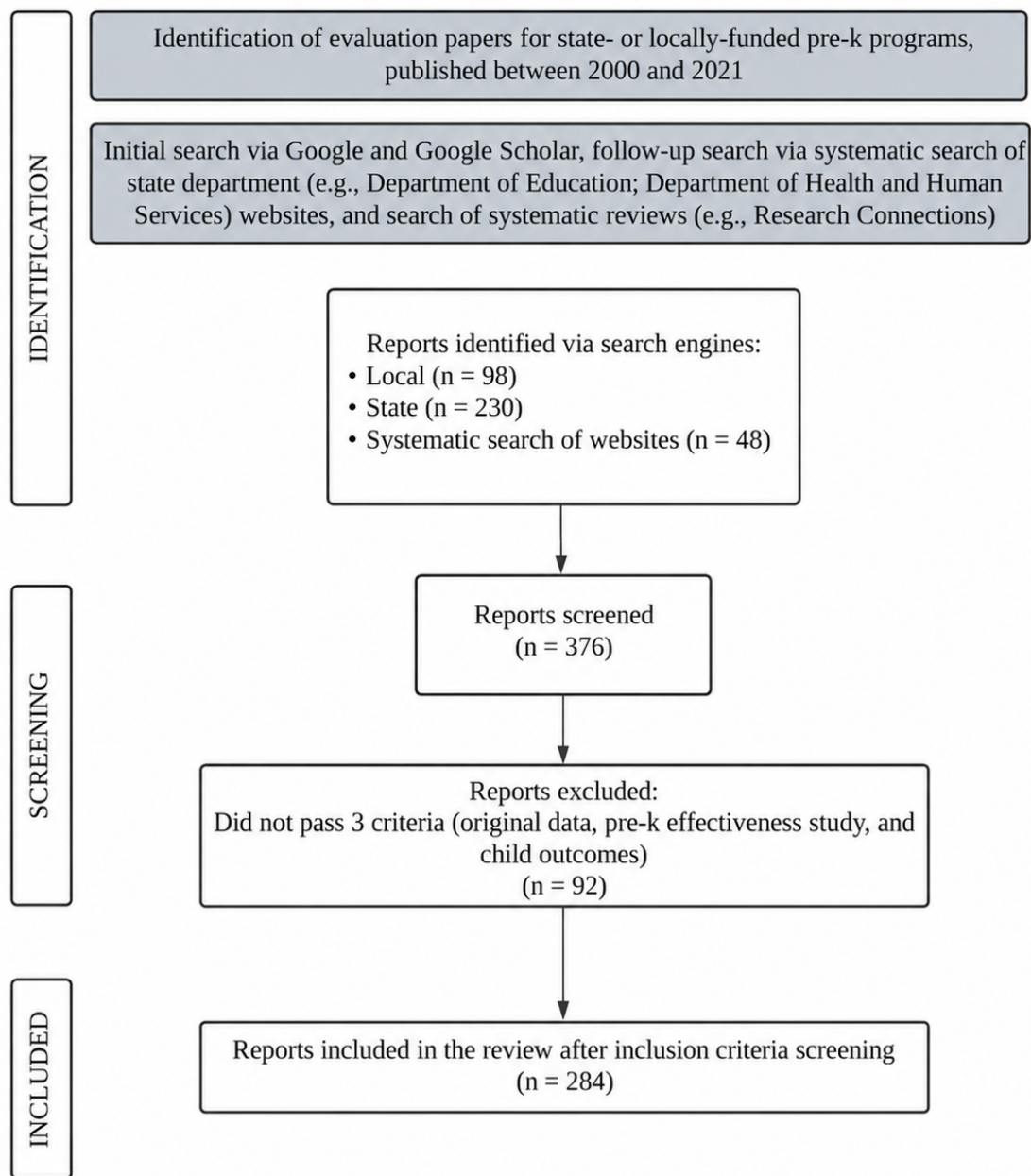
	Achievement Outcomes					
	End of Program			Longer-Term		
	N of Studies (Effects)	Significant Effects Favoring Pre-K	Significant Effects Favoring Control	N of Studies (Effects)	Significant Effects Favoring Pre-K	Significant Effects Favoring Control
Nonequivalent comparison with no pretests	11 (29)	90.9%	9.1%	28 (169)	93.9%	6.1%
Nonequivalent comparison with pretests	2 (5)	100.0%	0.0%	0	-	-
Age-cutoff regression discontinuity	24 (86)	100.0%	0.0%	-	-	-
Difference-in-Difference	0	-	-	6 (14)	83.3%	16.7%
Random assignment	2 (8)	100.0%	0.0%	1 (5)	20.0%	80.0%
Other Outcomes (e.g., Behavioral, SEL)						
Nonequivalent comparison with no pretests	11 (22)	81.8%	18.2%	9 (17)	100.0%	0.0%
Nonequivalent comparison with pretests	0	-	-	0	-	-
Age-cutoff regression discontinuity	5 (11)	100.0%	0.0%	-	-	-
Difference-in-Difference	0	-	-	0	-	-
Random assignment	1 (1)	100.0%	0.0%	2 (10)	50.0%	50.0%

Note. To avoid distortions stemming from the different number of effects reported in different studies, the proportions of statistically significant positive and negative effects within each study were first computed, then those proportions were averaged over the studies contributing effects in the respective category.

Table 6*Average Citation Counts from Google Scholar by Research Design Category*

Category	Reports		
	N	Mean	SD
1. No comparison group	28	9.2	10.5
2. Nonequivalent comparison with no pretest	64	26.0	46.1
3. Nonequivalent comparison with child pretest data	4	106.8	142.2
4. Age-cutoff regression discontinuity	40	130.2	267.4
5. Difference-in-difference	11	98.3	125.9
6. Random assignment	11	64.9	82.2
Total	158	59.2	151.0

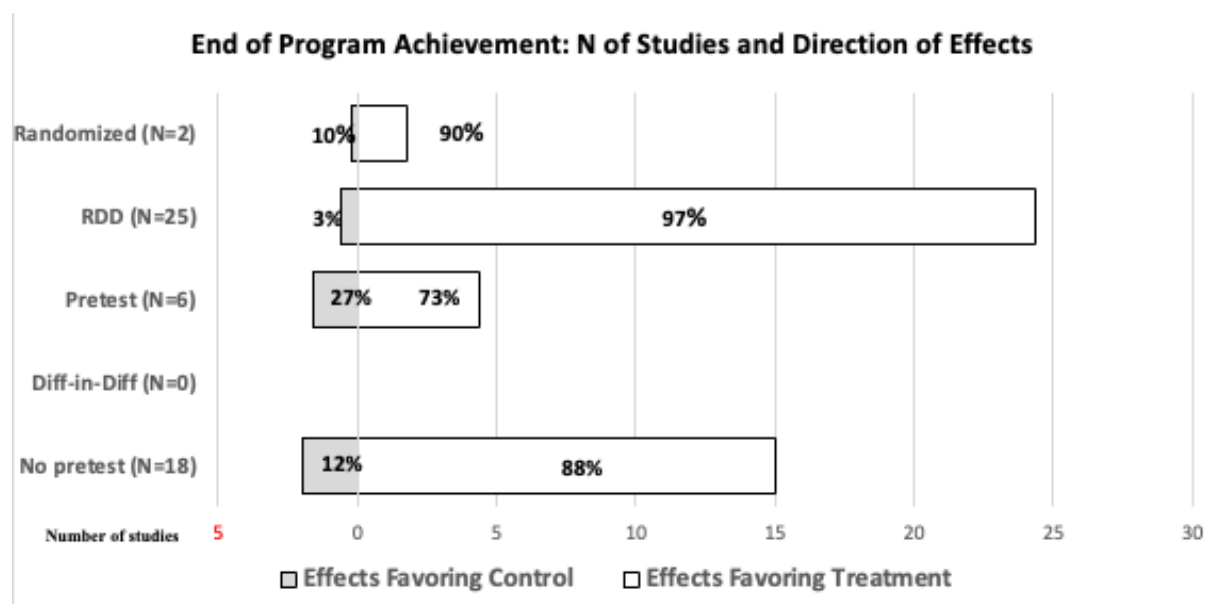
Note. 126 reports did not have a Google Scholar presence as of June 19, 2024.

Figure 1*Flowchart of Screening Methodology*

Note. Local and state refer to the administrative level of the pre-k program (e.g., a state or locally funded program).

Figure 2

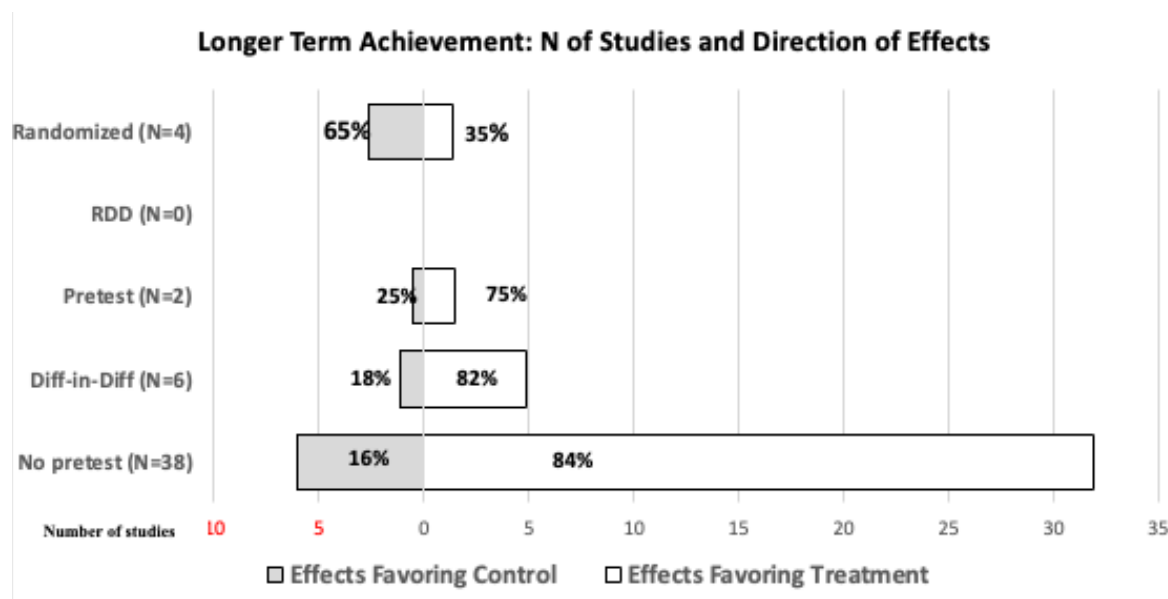
Number of Studies and Direction of Effects for Achievement Outcomes at the End of Pre-K



Note. The number of studies in each design category is indicated by the total length of the associated bar. The proportion of effects favoring the treatment and control groups within those studies are then superimposed on each bar. Effect proportions are based on the means of the respective within study proportions.

Figure 3

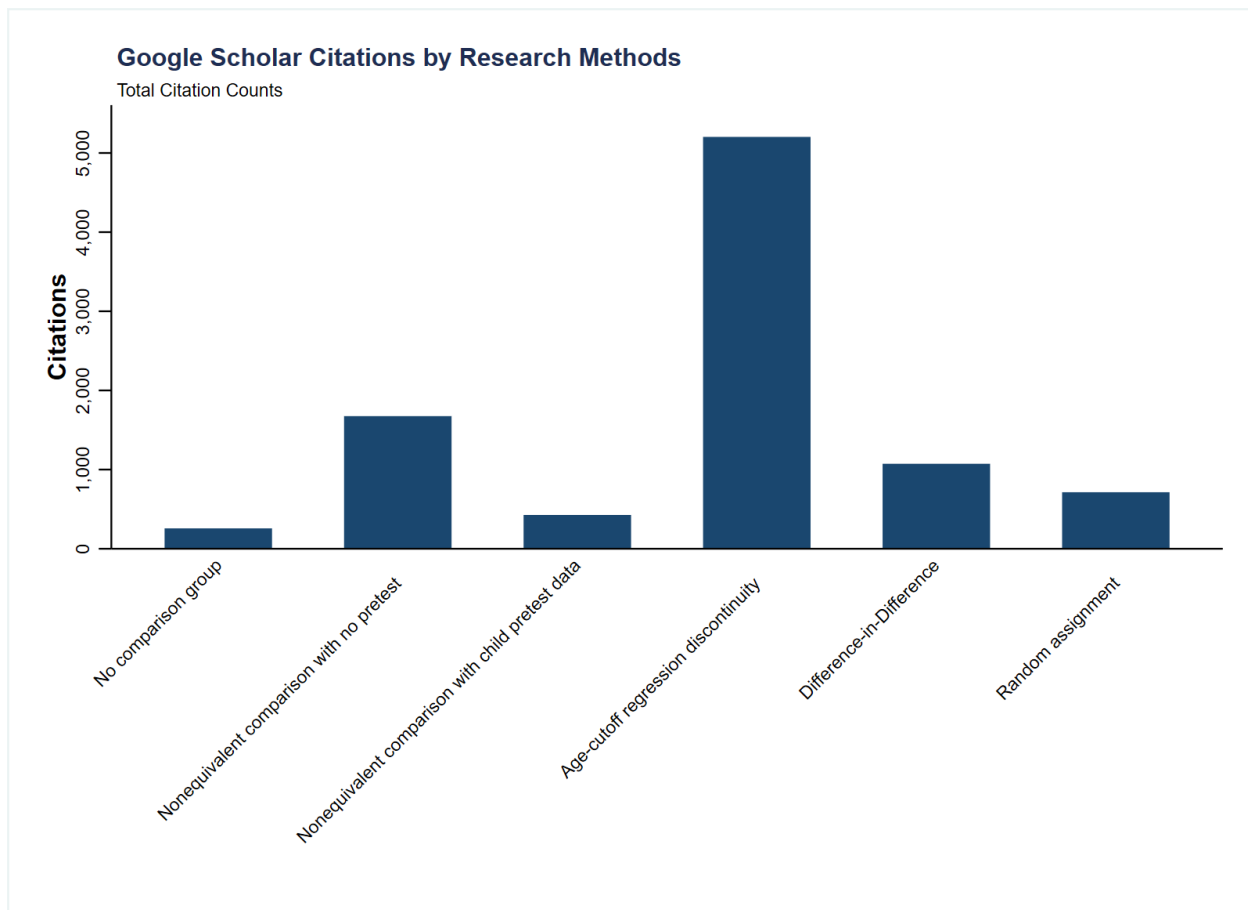
Number of Studies and Direction of Effects for Achievement Outcomes After the End of Pre-k



Note. The number of studies in each design category is indicated by the total length of the associated bar. The proportion of effects favoring the treatment and control groups within those studies are then superimposed on each bar. Effect proportions are based on the means of the respective within study proportions. Pre-k RD designs do not allow for follow-up past the end of pre-k.

Figure 4

Summary of Citation Counts from Google Scholar by Research Method



Note. Figure represents 158 reports. 126 reports did not have a Google Scholar presence as of June 19, 2024.

**A Systematic Review of Research Designs Used to Evaluate
State and Locally Funded Pre-K Programs**

Supplementary Materials

A Systematic Review of Research Designs Used to Evaluate State and Locally Funded Pre-K Programs

Training Coders to Reliability

Two senior researchers trained four doctoral students on the study's coding process. Initially, the four graduate students and two senior researchers coded 10 reports and discussed the alignment and misalignment in their codes. The initial coding process was iterative to accurately develop codes and ensure the accuracy of the coding process. After discrepancies were discussed, the remaining reports were randomly divided among the four graduate students into six sets of 25-45 reports. Each report was coded by two graduate students, and discrepancies were discussed with two senior researchers. The coding process occurred over a four-month period, and the senior researchers and graduate students met at least twice a month to discuss the coding process. Once all reports were coded, the two senior researchers discussed any remaining coding discrepancies and assigned unresolved discrepancies to other senior researchers on the team to ensure the reports were coded as accurately as possible. After all discrepancies were resolved, a master dataset was created that contained all the reports that did and did not pass the eligibility criteria.

Information about the Coding Process

Determining Eligibility

During step one of the coding process the coders determined eligibility of the report to be included in the study by coding whether the report presented original research findings, was a pre-k effectiveness study, and presented child outcomes. Each of the 376 reports was screened by two trained coders, who reviewed each report and dichotomously coded (yes/no) for each criterion.

The criterion concerning original research was designed to exclude literature reviews, commentaries, meta-analyses, research briefs that summarized reports, and other work that did not present original findings. The next criterion, which assessed whether the report was about pre-k effectiveness, was created to include studies attempting to examine the effectiveness of publicly funded preschool programs that are not Head Start. By and large, this refers to state-funded (e.g., Tennessee Voluntary Pre-K) or locally-funded (e.g., Boston Preschool) preschool programs that have an academic focus. For this criterion, pre-k enrollment, participation, or funding were generally compared against a control condition (or examined children's growth over time) that represents the absence of state-funded preschool. Effectiveness studies could use comparison groups that included children who did not attend pre-k or comparison groups could include children who attended alternative child care programs (e.g., Head Start, informal childcare, etc.). Effectiveness studies could also use statistical approaches (e.g., regression control; propensity scores) to generate comparisons. Finally, effectiveness reports could include experimental or quasi-experimental designs (e.g., regression discontinuity; difference-in-differences; lotteries).

The final criterion in the initial set of codes were about whether the report included child outcomes. This inclusion criterion encompassed measures of child functioning, abilities, or behaviors that were conducted at the child level. We further expanded the child outcomes definition to include the following outcomes: child test scores, grade retention, problem behavior, executive function, and absences. Reports were required to contain child outcomes to be included. We included reports that focused solely on child outcomes, as well as those that included child outcomes plus measures of classroom quality, teacher outcomes, or parent

outcomes; however, we did not include reports that solely evaluated classroom quality, teacher outcomes, or parent outcomes. You can find a full version of the reports we included in our study in Supplemental Table 7.

Coding Design Categories

If reports passed step one coding regarding eligibility, the reports were then coded for research design. The four graduate students coded each of the following categories as either yes or no: no comparison group/descriptive study, comparison group/regression/propensity scores with no baseline pretest data, comparison group/regression/propensity scores with child pretest data, birthdate regression discontinuity, non-birthdate regression discontinuity, diff-in-diff/quasi-experimental catch all, RCT/lottery, and no clear determination category. Detailed information about each of the categories can be found in Supplemental Table 1. Note, these category names differ slightly from the main text given we renamed the categories to be modeled after accepted terms used to describe research designs/methods. After the coding process was complete, the category titles were renamed for ease of understanding. Additionally, the non-birthdate regression discontinuity category was removed because none of the reports used this design.

Research assistants double-coded each report in this phase of coding. If coders had questions or were unsure about aspects of the report, they either made extensive notes about the report to discuss with the research team or contacted the primary study researchers for clarification about the report's research design.

In addition to coding for research design, each report was coded for document type and publication year. Document type included whether the report was a journal article, research report/brief/abstract/executive summary, presentation (e.g., by/to the state's Department of Education), conference proposal (submissions, research plans, or completed research), dissertation/thesis, fact sheet/infographic or some other kind of document. The other category encompassed documents that did not fit in any of the previous categories.

Significance and Direction of Effect Coding

Type of Outcome Reporting

Direct assessments include both standardized assessment (e.g., PPVT) as well as state-level standardized tests (e.g., Tennessee Comprehensive Assessment Program). Observation includes if a researcher was observing a behavior or outcome. Administrative data was used for items like special education placement, grade level retention, and high school graduation that are reported at the state- or school-level.

Results

Supplemental Table 1

Design Coding Strategy

Design Categories (Coded as Yes/No)	Description
No comparison group or descriptive study	Study did not include a comparison group of children. A comparison group is defined as a set of children that did not attend the state-funded pre-k program, and the comparison group can include alternative forms of formal childcare (e.g., Head Start, community based, private, church, etc.). More information on comparison groups is provided in the definition section. Instead, pre-k effects were assessed by comparing the outcomes of students who went to the pre-k program with their own pre-treatment outcomes or others in the pre-k program.
Comparison group/ regression/propensity scores with no baseline pretest data	Study contained comparison group, as defined before (i.e., comparison group can include Head Start, community based, private, church, etc.). The study does not have child-level baseline pretest (e.g., academic or social emotional skill assessments) data, but it may include demographic control variables taken during kindergarten or elementary school. The study could use regression or propensity scores to adjust comparisons between the “pre-k” group and the comparison group, but no pre-test measures were used.
Regression/propensity scores with child pretest data	Study used regression or propensity scores with baseline child-level pretest data to adjust nonrandomized comparisons between the pre-k group and the comparison group(s). Pretest data includes measures of <u>child skills or competencies collected at the start of pre-k</u> (or before). Pretests can include measures of achievement, social emotional skills, or executive functioning; however, it is important to note the pretest measure must match the domain of the outcome variable(s).
Regression discontinuity	Study used regression discontinuity as the main research design with child birth date as the forcing variable.

Difference-in-difference/ Quasi-experimental catch all	Study design was either difference-in-difference or another quasi-experimental design that is not captured above. Many studies at the county or state level (in other words, not child level) fall in this category, such as pre-k expansion studies. These studies claim to use some source of “exogenous” variation in access to pre-k to estimate the effectiveness of a state or local pre-k program.
RCT/Lottery	Study included child-level randomization to pre-k or control. This can include researcher-run randomization or program-run randomization through an admissions lottery.
No clear category determination	The study does not fit in any of the above categories, or it is unclear which category it falls under.

Note. Coding was recorded in a spreadsheet. Coders indicated yes or no for each of the design categories, guided by the descriptions. Comparison group was further defined to include groups that were constructed statistically (e.g., regression, propensity scores). The comparison group definition was created to serve as exclusion criteria designed to exclude studies that simply examine children who attended preschool without attempting to compare their outcomes with other children.

Supplemental Table 2*Differences in Direction of Effects by Design*

	Study Level ^a				Effect Level ^b			
	N of Studies	Favors PreK (% Row)	Favors Neither (% Row)	Favors Control (% Row)	N of Effects	Favors PreK (% Row)	Favors Neither (% Row)	Favors Control (% Row)
All Outcomes								
Nonequivalent comparison with no pretests	50	81%	1%	18%	549	77%	1%	22%
Nonequivalent comparison with pretests	6	75%	2%	23%	58	60%	5%	35%
Age-cutoff regression discontinuity	25	95%	-	5%	126	94%	-	6%
Difference-in-Difference	6	82%	-	18%	24	79%	-	21%
Random assignment	4	48%	1%	51%	144	47%	1%	52%
Total	91	83%	1%	16%	901	74%	1%	25%
End of Pre-K Achievement Outcomes								
Nonequivalent comparison with no pretests	18	88%	-	12%	52	83%	-	17%
Nonequivalent comparison with pretests	6	73%	2%	24%	38	53%	5%	42%
Age-cutoff regression discontinuity	25	97%	-	3%	106	97%	-	3%
Difference-in-Difference	-	-	-	-	-	-	-	-
Random assignment	2	90%	-	10%	12	92%	--	8%
Total	51	91%	-	9%	208	85%	1%	14%

Supplemental Table 2 (continued).

Post Pre-K Achievement Outcomes								
Nonequivalent comparison with no pretests	38	84%	1%	15%	350	78%	1%	21%
Nonequivalent comparison with pretests	2	75%	6%	19%	15	73%	7%	20%
Age-cutoff regression discontinuity	-	-	-	-	-	-	-	-
Difference-in-Difference	6	82%	-	18%	24	79%	-	21%
Random assignment	4	35%	-	65%	79	34%	1%	65%
Total	50	79%	1%	19%	468	71%	1%	28%
End of Pre-K Other Outcomes								
Nonequivalent comparison with no pretests	18	78%	2%	20%	54	76%	2%	22%
Nonequivalent comparison with pretests	1	80%	-	20%	5	80%	-	20%
Age-cutoff regression discontinuity	8	69%	-	31%	20	75%	-	25%
Difference-in-Difference	-	-	-	-	-	-	-	-
Random assignment	2	100%	-	-	10	100%	-	-
Total	29	77%	1%	22%	89	79%	1%	20%

Supplemental Table 2 (continued)

	Post Pre-K Other Outcomes							
Nonequivalent comparison with no pretests	18	74%	1%	25%	93	72%	2%	26%
Nonequivalent comparison with pretests	-	-	-	-	-	-	-	-
Age-cutoff regression discontinuity	-	-	-	-	-	-	-	-
Difference-in-Difference	-	-	-	-	-	-	-	-
Random assignment	3	39%	1%	60%	43	44%	2%	54%
Total	21	69%	1%	30%	136	63%	2%	35%

a. The different direction of effects categories are proportioned across effects within each study, then averaged across studies.

b. The proportions of the different direction of effects categories are computed across the full set of effects irrespective of different numbers within different studies.

Differences in Direction and Statistical Significance of Effects When Reported

	Study Level ^a					Effect Level ^b				
	N of Studies	Favors PreK & Significant (% Row)	Favors PreK & Not Significant (% Row)	Favors Control or Neither & Not Significant (% Row)	Favors Control & Significant (% Row)	N of Effects	Favors PreK & Significant (% Row)	Favors PreK & Not Significant (% Row)	Favors Control or Neither & Not Significant (% Row)	Favors Control & Significant (% Row)
End of Pre-K Achievement Outcomes										
Nonequivalent comparison with no pretests	14	75%	18%	-	7%	35	74%	17%	-	9%
Nonequivalent comparison with pretests	5	25%	39%	36%	-	22	23%	36%	41%	-
Age-cutoff regression discontinuity	25	81%	17%	3%	-	106	81%	16%	3%	-
Difference-in-Difference	0	-	-	-	-	0	-	-	-	-
Random assignment	2	63%	27%	10%	-	12	67%	25%	8%	-
Total	46	72%	20%	6%	2%	175	71%	19%	7%	2%
Post Pre-K Achievement Outcomes										
Nonequivalent comparison with no pretests	33	64%	17%	16%	3%	287	52%	24%	17%	7%
Nonequivalent comparison with pretests	1	-	63%	38%	-	8	-	62%	38%	-
Age-cutoff regression discontinuity	0	-	-	-	-	0	-	-	-	-
Difference-in-Difference	6	51%	31%	13%	6%	24	54%	25%	17%	4%
Random assignment	4	1%	34%	63%	2%	79	1%	33%	61%	5%
Total	44	55%	22%	20%	3%	398	41%	27%	26%	6%

Supplemental Table 3 (continued)

End of Pre-K Other Outcomes										
Nonequivalent comparison with no pretests	13	62%	6%	21%	12%	35	57%	11%	26%	6%
Nonequivalent comparison with pretests	1	-	80%	20%	-	5	-	80%	20%	-
Age-cutoff regression discontinuity	8	52%	17%	31%	-	20	55%	20%	25%	-
Difference-in-Difference	0	-	-	-	-	0	-	-	-	-
Random assignment	2	8%	92%	-	-	10	10%	90%	-	-
Total	24	51%	20%	23%	6%	70	46%	30%	21%	3%
Post Pre-K Other Outcomes										
Nonequivalent comparison with no pretests	14	44%	24%	32%	-	68	25%	38%	37%	-
Nonequivalent comparison with pretests	0	-	-	-	-	0	-	-	-	-
Age-cutoff regression discontinuity	0	-	-	-	-	0	-	-	-	-
Difference-in-Difference	0	-	-	-	-	0	-	-	-	-
Random assignment	3	19%	20%	58%	3%	43	19%	26%	51%	5%
Total	17	40%	24%	36%	-	111	23%	33%	42%	2%

a. The different significance categories are proportioned across effects within each study, then averaged across studies.

b. The proportions of the different significance categories are computed across the full set of effects irrespective of different numbers within different studies.

Multilevel Model with Effects Nested Within Studies: Results for Direction of Effects (Proportion Favoring Pre-K)

	End of Pre-K Achievement Outcomes N = 51 studies, 208 effects	Post Pre-K Achievement Outcomes N = 50 studies, 468 effects	End of Pre-K Other Outcomes N = 29 studies, 89 effects	Post Pre-K Other Outcomes N = 21 studies, 136 effects
	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Intercept	.900*** (.162)	.347** (.118)	1.00*** (.261)	.390* (.184)
Nonequivalent comparison with no pretests	-.014 (.171)	.492*** (.123)	-.221 (.275)	.346† (.199)
Nonequivalent comparison with pretests	-.167 (.187)	.403† (.205)	-.200 (.451)	-
Age-cutoff regression discontinuity	.075 (.169)	-	-.312 (.291)	-
Difference-in-Difference	-	.470** (.153)	-	-
Random assignment (Reference group)				

Note. Analyzed at the study level using the within-study proportions of positive effects as the dependent variable.

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

Supplemental Table 4b*Pairwise Comparisons (Based on estimated marginal means from multilevel models)***End of Pre-K Achievement Outcomes**

Design 1 (Mean)	Design 2 (Mean)	Difference	SE	p-value
QED.noPre (.886)	RCT (.900)	-.014	.171	.936
QED.Pre (.733)	RCT (.900)	-.167	.187	.377
RDD.age (.975)	RCT (.900)	.075	.169	.658
DiffnDiff -	-	-	-	-

Post Pre-K Achievement Outcomes

Design 1 (Mean)	Design 2 (Mean)	Difference	SE	p-value
QED.noPre (.839)	RCT (.347)	.492	.123	<.001
QED.Pre (.750)	RCT (.347)	.403	.205	.055
RDD.age -	-	-	-	-
DiffnDiff (.817)	RCT (.347)	.470	.153	.003

End of Pre-K Other Outcomes

Design 1 (Mean)	Design 2 (Mean)	Difference	SE	p-value
QED.noPre (.779)	RCT (1.00)	-.221	.275	.428
QED.Pre (.800)	RCT (1.00)	-.200	.451	.661
RDD.age (.687)	RCT (1.00)	-.312	.291	.294
DiffnDiff -	-	-	-	-

Post Pre-K Other Outcomes

Design 1 (Mean)	Design 2 (Mean)	Difference	SE	p-value
QED.noPre (.735)	RCT (.390)	.346	.199	.099
QED.Pre -	-	-	-	-
RDD.age -	-	-	-	-
DiffnDiff -	-	-	-	-

Supplemental Table 5a

Multilevel Model with Effects Nested Within Studies: Results for Direction of Effects for Only Statistically Significant Effects (Proportion Favoring Pre-K)

	End of Pre-K Achievement Outcomes N = 40 studies, 128 effects	Post Pre-K Achievement Outcomes N = 35 studies, 188 effects	End of Pre-K Other Outcomes N = 17 studies, 34 effects	Post Pre-K Other Outcomes N = 11 studies, 27 effects
	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
Intercept	1.00*** (.113)	.200 (.240)	1.00* (.342)	.500* (.167)
Nonequivalent comparison with no pretests	-.083 (.122)	.729** (.243)	-.182 (.357)	.500* (.184)
Nonequivalent comparison with pretests	.000 (.160)	-	-	-
Age-cutoff regression discontinuity	.000 (.117)	-	.000 (.375)	-
Difference-in-Difference	-	.633* (.259)	-	-
Random assignment (Reference group)				

Note. Analyzed at the study level using the within-study proportions of positive effects as the dependent variable.

*** $p < .001$, ** $p < .01$, * $p < .05$, † $p < .10$

Supplemental Table 5b*Pairwise Comparisons (Based on estimated marginal means from multilevel models)***End of Pre-K Achievement Outcomes**

Design 1 (Mean)	Design 2 (Mean)	Difference	SE	p-value
QED.noPre (.917)	RCT (1.00)	-.083	.122	.498
QED.Pre (1.00)	-	.000	.160	1.00
RDD.age (1.00)	-	.000	.117	1.00
DiffnDiff -	-	-	-	-

Post Pre-K Achievement Outcomes

Design 1 (Mean)	Design 2 (Mean)	Difference	SE	p-value
QED.noPre (.929)	RCT (.200)	.729	.243	.005
QED.Pre -	-	-	-	-
RDD.age -	-	-	-	-
DiffnDiff (.833)	RCT (.200)	.633	.259	.020

End of Pre-K Other Outcomes

Design 1 (Mean)	Design 2 (Mean)	Difference	SE	p-value
QED.noPre (.818)	RCT (1.00)	-.182	.357	.619
QED.Pre -	-	-	-	-
RDD.age (1.00)	RCT (1.00)	.000	.375	1.00
DiffnDiff -	-	-	-	-

Post Pre-K Other Outcomes

Design 1 (Mean)	Design 2 (Mean)	Difference	SE	p-value
QED.noPre (1.00)	RCT (.500)	.500	.184	.024
QED.Pre -	-	-	-	-
RDD.age -	-	-	-	-
DiffnDiff -	-	-	-	-

Supplemental Table 6*Additional Lottery-Design Pre-K Evaluations*

Type and number of effects	% positive effects for treatment group	
	End of Treatment	Longer-Term (K Spring and Onward)
Achievement Outcomes (<i>n</i> = 65)	100% (<i>n</i> = 3)	32% (<i>n</i> = 62)
Non-Achievement Outcomes (<i>n</i> = 8)	-	25% (<i>n</i> = 8)
Total (<i>n</i> = 73)	100% (<i>n</i> = 3)	31% (<i>n</i> = 70)

Note. Effects coded from five reports:

- Humphries, J. E., Neilson, C., Ye, X., & Zimmerman, S. D. (2024). *Parents' earnings and the returns to universal pre-kindergarten* (No. w33038). National Bureau of Economic Research.
- Braga, B., Doromal, J. B., Greenberg, E., Monarrez, T., Restrepo, L. & Lamb, R. (2024). *The effects of public pre-k for 3-year-olds on early elementary school outcomes: Evidence from the DC centralized lottery*. (EdWorkingPaper: 24-1019). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/br2m-wb87>
- Woodyard, H., Sass, T. R. & Fazlul, I. (2023). *Assessing the benefits of education in early childhood: Evidence from a pre-k lottery in Georgia*. Technical Report No. 23-880, Annenberg Institute for School Reform at Brown University.
- Woodyard, H., Sass, T., & Fazlul, I. (2022). *The efficacy of school-based pre-k program sites in a metro- Atlanta school district*. Georgia Policy Labs. <https://doi.org/10.57709/30728972>
- Durkin, K., Lipsey, M. W., Farran, D. C., & Wiesen, S. E. (2022). Effects of a statewide pre-kindergarten program on children's achievement and behavior through sixth grade. *Developmental Psychology*, 58(3), 470. <https://dx.doi.org/10.1037/dev0001301>

Supplemental Table 7*Full List of Reports That Passed the Eligibility Round of Coding*

State	City	Publication Year	Report Citation
Alabama		2016	Alabama Department of Early Childhood Education. (2016). <i>Chronic absenteeism: Differences between First Class Pre-K students and non-First Class Pre-K students. (First Class Pre-K Issue Brief 1)</i> . https://children.alabama.gov/wp-content/uploads/2021/09/Absenteeism_01-10-17.pdf
Alabama		2017	Alabama Department of Early Childhood Education. (2017). <i>School readiness at kindergarten entry: Differences between students who participated in First Class Pre-K and those who did not, fall 2016. (First Class Pre-K Issue Brief 4)</i> . Alabama, Department of Early Childhood Education.
Alabama		2018	Alabama Department of Early Childhood Education. (2018). <i>Special education needs: Differences between First Class Pre-K students and non-First Class Pre-K students and potential cost savings. (First Class Pre-K Issue Brief 5a)</i> . https://children.alabama.gov/wp-content/uploads/2021/09/Issue-Brief-5a_SpecialEd_overall_2-2018.pdf
Alabama		2018	Preskitt, J., Fifolt, M., Ernest, J. & Adams, J. (2018). <i>Achievement gap closure and gains associated with Alabama First Class Pre-K. [Research Issue Brief]</i> . Alabama Department of Early Childhood Education. https://children.alabama.gov/wp-content/uploads/2021/09/Achivement-Gap-Closure-and-Gains-Proficiency-in-Reading-and-Math-at-3rd-grade.pdf

- | | | |
|---------|------|--|
| Alabama | 2019 | Adams, J., Becker, D., Ernest, J., Fifolt, M., Johnson, H., Ross, J., Sen, B., Strichik, T., & Preskitt, J. (July, 2019) <i>Disciplinary infractions in Alabama public schools: First-Class Pre-K students have lower discipline rates. [Research Issue Brief]</i> . https://children.alabama.gov/wp-content/uploads/2021/09/7_Discipline_7-2019.pdf |
| Alabama | 2020 | Preskitt, J., Johnson, H., Becker, D., Ernest, J., Fifolt, M., Adams, J., ... & Sen, B. (2020). The persistence of reading and math proficiency: The benefits of Alabama's pre-kindergarten program endure in elementary and middle school. <i>International Journal of Child Care and Education Policy</i> , 14(1), 1-12. https://doi.org/10.1186/s40723-020-00073-3 |
| Alabama | 2012 | Tucker, K. (2012). <i>The differences in reading readiness among kindergartners who attended state and federally funded pre-k in Alabama</i> . [Unpublished doctoral dissertation]. Liberty University. |
| Alabama | 2016 | Alabama Department of Early Childhood Education. (2016). <i>Retention in grade: Differences between First Class Pre-K students and non-First Class Pre-K students. (First Class Pre-K Issue Brief 2)</i> . https://children.alabama.gov/wp-content/uploads/2021/09/Retention_01-10-17.pdf |
| Alaska | 2011 | Alaska Department of Education and Early Development. (2011). <i>The Alaska Pilot Pre-Kindergarten Project (AP3): Project evaluation -- year one</i> . https://www.akleg.gov/basis/get_documents.asp?session=27&docid=9133 |
| Alaska | 2012 | Alaska Department of Education and Early Development. (2012). <i>Alaska Pilot Pre-Kindergarten Project (AP3): Year two report</i> . https://omb.alaska.gov/ombfiles/13_budget/CapBackup/proj58943.pdf |
| Alaska | 2016 | Alaska Department of Education and Early Development. (2016). <i>Alaska Pre-Kindergarten Program</i> . https://www.akleg.gov/basis/get_documents.asp?session=29&docid=51578 |

- | | | |
|----------|------|--|
| Arkansas | 2007 | Hustedt, J. T., Barnett, W. S., Jung, K., & Thomas, J. (2007). <i>The effects of the Arkansas Better Chance Program on young children's school readiness</i> . National Institute for Early Education Research. https://nieer.org/sites/default/files/2023-08/arkansasyear1.pdf |
| Arkansas | 2008 | Hustedt, J. T., Barnett, W. S., & Jung, K. (2008). <i>Longitudinal effects of the Arkansas Better Chance Program: Findings from kindergarten and first grade</i> . National Institute for Early Education Research. https://nieer.org/sites/default/files/2023-08/arkansaslongitudinal.pdf |
| Arkansas | 2013 | Jung, K., Barnett, W. S., Hustedt, J. T., & Francis, J. (2013). <i>Longitudinal effects of the Arkansas Better Chance Program: Findings from first grade through fourth grade</i> . National Institute for Early Education Research. http://nieer.org/sites/nieer/files/ArkansasLongitudinalReportMay2013n.pdf . |
| Arkansas | 2015 | Hustedt, J. T., Jung, K., Barnett, W., & Williams, T. (2015). Kindergarten readiness impacts of the Arkansas Better Chance state prekindergarten initiative. <i>Elementary School Journal</i> , 116(2), 198-216. https://doi.org/10.1086/684105 |
| Arkansas | 2019 | Argue, S. E., & Holland, G. O. (2019). <i>Arkansas Better Chance (ABC) program longitudinal study (2009-2018)</i> . Arkansas Research Center. https://arc.arkansas.gov/arcweb/research/2019-ABC-Longitudinal-Report.pdf |
| Arkansas | | Argue, S. E., & Holland, G. (n.d.). <i>Arkansas Better Chance (ABC) shrinks gap in kindergarten readiness for economically disadvantaged students</i> . Arkansas Research Center. https://files.eric.ed.gov/fulltext/ED607412.pdf |
| Arkansas | 2017 | Argue, S. E., & Holland, G. O. (2017). <i>Late elementary school outcomes for children participating in Arkansas Better Chance (ABC) program</i> . Arkansas Research Center. https://files.eric.ed.gov/fulltext/ED607417.pdf |
| Arkansas | 2020 | Argue, S. E., & Holland, G. O. (2020). <i>Arkansas Better Chance (ABC) program longitudinal study (2009-2019)</i> . Arkansas Research Center. https://arc.arkansas.gov/arcweb/research/2021-ABC-Longitudinal-Report.pdf |

California	2009	Barnett, W. S., Howes, C., & Jung, K. (2009). <i>California's state preschool program: Quality and effects on children's cognitive abilities at kindergarten entry</i> . National Institute for Early Education Research.
California	2013	Applied Survey Research. (2013). <i>Evaluating Preschool for All effectiveness</i> . First 5 San Francisco. https://glenpricegroup.com/sites/laupresearch/wp-content/uploads/sites/2/2014/05/Evaluating-PFA-Effectiveness-Research-Brief.pdf
California	2017	Kyger, M., & Barnhart, M. (2017). <i>Effects of high-quality preschool on early literacy and mathematics skills- a regression-discontinuity approach</i> . Los Angeles Universal Preschool. https://researchconnections.org/childcare/resources/33498
Colorado	2010	Colorado Department of Education. (2010). <i>Colorado Preschool Program: Legislative report 2010</i> . https://www.cde.state.co.us/sites/default/files/documents/cpp/download/cppdocs/2010_cpp_legislative_report.pdf
Colorado	2011	Colorado. Department of Education. (2011). <i>Colorado Preschool Program: Working together: 2011 legislative report</i> . Colorado, Department of Education. https://spl.cde.state.co.us/artemis/edserials/ed55069internet/ed550692011internet.pdf
Colorado	2012	Colorado Department of Education. (2012). <i>Colorado Preschool Program: 2012 Legislative Report</i> . https://spl.cde.state.co.us/artemis/edserials/ed55069internet/ed550692012internet.pdf
Colorado	2013	Colorado Department of Education. (2013). <i>Colorado Preschool Program: Legislative report 2013</i> . https://www.cde.state.co.us/sites/default/files/documents/cpp/download/cppdocs/2013_cpp_legislative_report.pdf

Colorado	2014	Colorado Department of Education. (2014). <i>Colorado Preschool Program: Legislative report 2014</i> . https://spl.cde.state.co.us/artemis/edserials/ed55069internet/ed550692014internet.pdf
Colorado	2015	Colorado Department of Education. (2015). <i>Colorado Preschool Program: Legislative report 2015</i> . https://spl.cde.state.co.us/artemis/edserials/ed55069internet/ed550692015internet.pdf
Colorado	2016	Colorado Department of Education. (2016). <i>Colorado Preschool Program: Legislative report 2016</i> . https://spl.cde.state.co.us/artemis/edserials/ed55069internet/ed550692016internet.pdf
Colorado	2017	Colorado Department of Education. (2017). <i>Colorado Preschool Program: Legislative report 2017</i> . https://spl.cde.state.co.us/artemis/edserials/ed55069internet/ed550692017internet.pdf
Colorado	2018	Colorado Department of Education. (2018). <i>Colorado Preschool Program: Legislative report 2018</i> . https://spl.cde.state.co.us/artemis/edserials/ed55069internet/ed550692018internet.pdf
Colorado	2019	Colorado Department of Education. (2019). <i>Colorado Preschool Program: Legislative report 2019</i> . https://spl.cde.state.co.us/artemis/edserials/ed55069internet/ed550692019internet.pdf
Colorado	2020	Colorado Department of Education. (2020). <i>Colorado Preschool Program: Legislative report 2020</i> . https://teachingstrategies.com/wp-content/uploads/2021/08/Colorado-Preschool-Program-2020-Report.pdf

Colorado		2021	Colorado Department of Education. (2021). <i>Colorado Preschool Program: Legislative report 2021</i> . https://leg.colorado.gov/sites/default/files/images/committees/2017/2021_cpp_report.pdf
Colorado	Denver	2009	Palaich, R. (2009). <i>DPP 2008-2009 cohort TCAP results</i> . Denver Preschool Program. https://researchconnections.org/childcare/resources/33479
Colorado	Denver	2010	Klute, M. M. (2010). <i>Denver Preschool Program: Report on child outcomes: 2009-10 school year</i> . Denver Preschool Program. https://dpp.org/wp-content/uploads/2022/05/2009-2010-Child-Outcomes.pdf
Colorado	Denver	2011	Klute, M. M., & Ponce, C. (2011). <i>Denver Preschool Program: Report on child outcomes: 2010-11 school year</i> . Denver Preschool Program. https://dpp.org/wp-content/uploads/2022/05/2010-2011-Child-Outcomes.pdf
Colorado	Denver	2012	Klute, M. M. (2012). <i>Denver Preschool Program: Report on child outcomes: 2011-12 school year</i> . Denver Preschool Program. https://dpp.org/wp-content/uploads/2022/05/2011-2012-Child-Outcomes.pdf
Colorado	Denver	2013	Klute, M. M. (2013). <i>Denver Preschool Program: Report on child outcomes: 2012-13 school year</i> . Denver Preschool Program. https://dpp.org/wp-content/uploads/2022/05/2012-2013-Child-Outcomes.pdf
Colorado	Denver	2014	Green, S., Ponce, C., & Mangels, D. (2014). <i>Denver Preschool Program: Report on child outcomes: 2013-2014 school year</i> . https://researchconnections.org/childcare/resources/33485
Colorado	Denver	2015	Green, S., Ponce, C., Mangels, D., & Landgraff, C. (2015). <i>Denver Preschool Program child outcome evaluation 2014-2015: Part A: Preschool progress and school readiness, evaluation cohort 7 preschool report</i> [unpublished report]. https://dpp.org/cohort-7-preschool-report/
Colorado	Denver	2016	Green, S. and Mangels, D. (2016). <i>Denver Preschool Program child outcome evaluation 2015-2016: Part A: Preschool progress and school readiness, evaluation cohort 8 preschool report</i> [unpublished report]. https://dpp.org/cohort-8-preschool-report/

Colorado	Denver	2018	Green, S., Mangels, D. & Reale, M. (2017). <i>Denver Preschool Program child outcomes study 2016-2017: Part A: Preschool progress and school readiness report, cohort 9</i> . https://dpp.org/cohort-9-preschool-report/
Colorado	Denver	2018	Green, S., Mangels, D. & Reale, M. (2018). <i>Denver Preschool Program child outcomes study 2017-2018: Part A: Preschool progress and school readiness report, cohort 10</i> . https://dpp.org/cohort-10-report/
Colorado		2010	Colorado Department of Education. (2010). <i>Colorado Preschool Program: Building a brighter future for Colorado's children: 2010 legislative report</i> . https://www.cde.state.co.us/sites/default/files/documents/cpp/download/cppdocs/2010_cpp_legislative_report.pdf
Connecticut		2019	Montrosse-Moorhead, B., Dougherty, S. M., La Salle, T. P., Weiner, J. M., & Dostal, H. M. (2019). The overall and differential effects of a targeted prekindergarten program: Evidence from Connecticut. <i>Early Childhood Research Quarterly</i> , 48, 134-145. https://doi.org/10.1016/j.ecresq.2019.02.006
Connecticut		2016	Connecticut Academy of Science and Engineering. (2016). <i>Early childhood regression discontinuity study</i> . https://www.ctoec.org/wp-content/uploads/2020/10/EARLY_CHILDHOOD_REGRESSION_DISCONTINUITY_STUDY_06_16_16.pdf
Delaware		2002	Gamel-McCormick, M., & Amsden, D. (2002). <i>Investing in better outcomes: The Delaware early childhood longitudinal study</i> . Delaware State Department of Education. https://files.eric.ed.gov/fulltext/ED482164.pdf
Delaware		2002	Cornwell, J. R., & Kersteter, P. G. (2002). <i>Head Start/ECAP outcomes report: State of Delaware: Fiscal year 2001-2002</i> . Delaware Department of Education. https://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/146/earclychildhood_statereport01-02.pdf

- | | | |
|----------|------|--|
| Delaware | 2005 | Delaware Interagency Resource Management Committee, Delaware Department of Education, Delaware Department of Health and Social Services, & Delaware Department of Services for Children, Youth, and Their Families. (2005). <i>Investing in better outcomes: Reaping continued dividends: The Delaware early childhood longitudinal study follow-up with fifth graders</i> . University of Delaware, Center for Disabilities Studies.
https://udspace.udel.edu/server/api/core/bitstreams/f4924daf-f06c-4054-b99b-183426499371/content |
| Delaware | 2014 | Schwartz, H. L., Karoly, L. A., Le, V. N., Tamargo, J., & Setodji, C. M. (2014). <i>Evaluation of Delaware Stars for Early Success: Year 1 report</i> [research report]. RAND Corporation.
https://www.rand.org/pubs/research_reports/RR606.html |
| Florida | 2008 | Florida Office of Program Policy Analysis and Government Accountability. (2008). <i>61% of voluntary prekindergarten children are ready for kindergarten: Accountability process needs improvement</i> . (Report No. 08-23). Office of Program Policy Analysis and Government Accountability.
https://oppaga.fl.gov/Documents/Reports/08-23.pdf |
| Florida | 2009 | Figlio, D., & Roth, J. (2009). The behavioral consequences of pre-kindergarten participation for disadvantaged youth. In <i>The Problems of Disadvantaged Youth</i> (pp. 15-42). University of Chicago Press. https://oppaga.fl.gov/Documents/Reports/08-23.pdf |
| Florida | 2010 | Florida Office of Program Policy Analysis and Government Accountability. (2010). <i>Steps taken to improve the VPK program: Additional actions needed to increase program accountability</i> . (Report No. 10-57). https://oppaga.fl.gov/Documents/Reports/10-57.pdf |

- | | | | |
|---------|-------------------|------|--|
| Florida | | 2012 | Florida Office of Program Policy Analysis and Government Accountability. (2012). <i>VPK teachers' education level does not substantially influence kindergarten readiness scores: Program and provider differences affect readiness rates</i> . (Report No. 12-06). Office of Program Policy Analysis and Government Accountability. https://oppaga.fl.gov/Documents/Reports/12-06.pdf |
| Florida | | 2019 | Goff, S., & Webster, M. (2019). <i>Early learning update</i> [presentation]. Florida Organization of Instructional Leaders, Department of Education. https://oppaga.fl.gov/Documents/Reports/12-06.pdf |
| Florida | | 2019 | Miller, L. C., & Bassok, D. (2019). The effects of universal preschool on grade retention. <i>Education Finance and Policy</i> , 14(2), 149-177. https://doi.org/10.1162/edfp_a_00248 |
| Florida | | 2019 | Melin, M. (2019). <i>The sustained impact of prekindergarten participation on third grade academic performance</i> . Electronic Theses and Dissertations, 2004-2019. https://stars.library.ucf.edu/etd/6538 |
| Florida | | 2019 | Florida Department of Education. (2019). <i>2020-2025: Data on strategic plan measures</i> . https://oppaga.fl.gov/Documents/Reports/12-06.pdf |
| Florida | | 2020 | Goff, S., & Webster, M. (2020). <i>Early learning update</i> [presentation]. Florida Organization of Instructional Leaders, Department of Education. https://oppaga.fl.gov/Documents/Reports/12-06.pdf |
| Florida | Miami | 2008 | Winsler, A., Tran, H., Hartman, S. C., Madigan, A. L., Manfra, L., & Bleiker, C. (2008). School readiness gains made by ethnically diverse children in poverty attending center-based childcare and public school pre-kindergarten programs. <i>Early Childhood Research Quarterly</i> , 23(3), 314-329. https://oppaga.fl.gov/Documents/Reports/12-06.pdf |
| Florida | Miami-Dade County | 2017 | Ansari, A., López, M., Manfra, L., Bleiker, C., Dinehart, L. H., Hartman, S. C., & Winsler, A. (2017). Differential third-grade outcomes associated with attending publicly funded preschool programs for low-income Latino children. <i>Child Development</i> , 88(5), 1743-1756. https://doi.org/10.1111/cdev.12663 |

- | | | | |
|---------|-------|------|--|
| Florida | Miami | 2019 | Conger, D., Gibbs, C. R., Uchikoshi, Y., & Winsler, A. (2019). New benefits of public school pre-kindergarten programs: Early school stability, grade promotion, and exit from ELL services. <i>Early Childhood Research Quarterly</i> , 48, 26–35. https://doi.org/10.1016/j.ecresq.2018.10.015 |
| Georgia | | 2004 | Henry, G. T., Ponder, B. D., Rickman, D. K., Mashburn, A. J., Henderson, L. W., & Gordon, C. S. (2004). <i>An evaluation of the implementation of Georgia's pre-k program: Report of the findings from the Georgia Early Childhood Study (2002-03)</i> . Georgia State University, School of Policy Studies. https://oppaga.fl.gov/Documents/Reports/12-06.pdf |
| Georgia | | 2008 | Fitzpatrick, M. D. (2008). Starting school at four: The effect of Universal pre-kindergarten on children's academic achievement. <i>The B.E. Journal of Economic Analysis and Policy</i> , 8(1), 1-38. https://oppaga.fl.gov/Documents/Reports/12-06.pdf |
| Georgia | | 2013 | Peisner-Feinberg, E. S., Schaaf, J. M., & LaForett, D. R. (2013). <i>Children's growth and classroom experiences in Georgia's Pre-K Program: Findings from the 2011–2012 evaluation study</i> . The University of North Carolina, FPG Child Development Institute. https://files.eric.ed.gov/fulltext/ED541940.pdf |
| Georgia | | 2014 | Peisner-Feinberg, E. S., Schaaf, J. M., LaForett, D. R., Hildebrandt, L.M., & Sideris, J. (2014). <i>Effects of Georgia's Pre-K Program on children's school readiness skills: Findings from the 2012–2013 evaluation study</i> . The University of North Carolina, FPG Child Development Institute. https://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/GAPreKEval_RDDReport%203-4-2014.pdf |
| Georgia | | 2015 | Peisner-Feinberg, E. S., Schaaf, J. M., Hildebrandt, L.M., & Pan, Y. (2015). <i>Children's pre-k outcomes and classroom quality in Georgia's Pre-K Program: Findings from the 2013–2014 evaluation study</i> . The University of North Carolina, FPG Child Development Institute. https://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/GAPreKEval2013-2014%20Report.pdf |

- Georgia 2016 Peisner-Feinberg, E. S., Garwood, J. D., & Mokrova, I. L. (2016). *Children's outcomes and classroom quality from pre-k through kindergarten: Findings from year 2 of Georgia's pre-k longitudinal study*. The University of North Carolina, FPG Child Development Institute. <https://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/GA%20Pre-K%20Longitudinal%20Study%20Yr%202%20Report%20Final.pdf>
- Georgia 2017 Early, D. M., Li, W., & Maxwell, K. L. (2017). *Third-grade achievement for children who participated in Georgia's pre-k*. Child Trends. <https://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/GA%20Pre-K%20Longitudinal%20Study%20Yr%202%20Report%20Final.pdf>
- Georgia 2017 Peisner-Feinberg, E. S., Mokrova, I. L., & Anderson, T. L. (2017). *Children's outcomes through first grade: Findings from year 3 of Georgia's pre-k longitudinal study*. The University of North Carolina, FPG Child Development Institute. <https://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/GA%20Pre-K%20Longitudinal%20Study%20Yr%202%20Report%20Final.pdf>
- Georgia 2019 Peisner-Feinberg, E., Van Manen, K., Mokrova, I., & Burchinal, M. (2019). *Children's outcomes through second grade: Findings from year 4 of Georgia's pre-k longitudinal study*. The University of North Carolina, FPG Child Development Institute. <https://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/GA%20Pre-K%20Long%20Study%20Yr4%20Report.pdf>

- | | | |
|----------|------|--|
| Georgia | 2021 | Soliday Hong, S., Zadrozny, S., Walker, J., Love, E.N.G., Osborne, J.D., Owen, J. L., Jenkins, G., & Peisner-Feinberg, E. (2021). <i>Longitudinal study of Georgia's pre-k program: Third grade report</i> . The University of North Carolina, FPG Child Development Institute.
https://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/GA%20Pre-K%20Long%20Study%20Yr4%20Report.pdf |
| Georgia | 2003 | Henry, G. T., Henderson, L. W., Ponder, B. D., Gordon, C. S., Mashburn, A. J., & Rickman, D. K. (2003). <i>Report of the findings from the Early Childhood Study: 2001-02</i> . Georgia State University, School of Policy Studies. https://files.eric.ed.gov/fulltext/ED481261.pdf |
| Georgia | 2005 | Henry, G. T., Rickman, D. K., Ponder, B. D., Henderson, L. W., Mashburn, A. J., & Gordon, C. S. (2005). <i>The Georgia Early Childhood Study: 2001-2004: Final report</i> . Georgia State University, School of Policy Studies. https://files.eric.ed.gov/fulltext/ED481261.pdf |
| Guam | 2017 | Ada, P. A., Fernandez, J., Sanchez, J., Sanchez, E., & Napa, Z. (2017). <i>Pre-kindergarten program evaluation report SY2015-16</i> . Guam Department of Education. https://gdoe.net/files/user/1/file/Pre-K-Eval-Report-as-of-Feb-27%2C-2017-with-cover.pdf |
| Hawaii | 2020 | Department of Education of Hawaii. (2020). <i>Relating to special education and Title I-funded prekindergarten programs report to the 2020 legislature</i> . https://gdoe.net/files/user/1/file/Pre-K-Eval-Report-as-of-Feb-27%2C-2017-with-cover.pdf |
| Hawaii | 2015 | Zellman, G., & Kilburn, M. R. (2015). <i>Final report on the Hawaii P-3 evaluation</i> . Rand Corporation. https://www.rand.org/pubs/research_reports/RR1100.html |
| Illinois | 2004 | Illinois State Board of Education Data Analysis and Progress Reporting Division. (2004). <i>Illinois prekindergarten program for children at risk for academic failure: FY 2003 evaluation report</i> . Illinois State Board of Education. https://www.rand.org/pubs/research_reports/RR1100.html |

- | | | |
|----------|------|--|
| Illinois | 2007 | Illinois State Board of Education Data Analysis and Progress Reporting Division. (2007). <i>Illinois prekindergarten program for children at risk for academic failure: 2005-2006 evaluation report</i> . Illinois State Board of Education. https://www.isbe.net/Documents/prek_evaluation06.pdf |
| Illinois | 2010 | Illinois State Board of Education Data Analysis and Progress Reporting Division. (2010). <i>Illinois prekindergarten program report: 2008-2009</i> . Illinois State Board of Education. https://www.isbe.net/Documents/pre-k_program_report08-09.pdf |
| Illinois | 2012 | Gaylor, E., Spiker, D., Fleming, J., & Korfmacher, J. (2012). <i>Illinois Preschool For All (PFA) program evaluation</i> . Erikson Institute, Herr Research Center for Children and Social Policy. https://www.erikson.edu/wp-content/uploads/PFA-Evaluation-Final-Report-June-2012.pdf |
| Illinois | | Illinois State Board of Education Division of Early Childhood Education. (n.d.). <i>Illinois Preschool for All Children program: A three year report: SY 2010 through SY 2012</i> . Illinois State Board of Education. https://www.erikson.edu/wp-content/uploads/PFA-Evaluation-Final-Report-June-2012.pdf |
| Iowa | 2013 | Iowa Department of Education Early Learning Team. (2013). <i>Statewide voluntary preschool programs for four-year-old children: 2012-2013: Legislative committee request</i> . Iowa, Department of Education. https://educate.iowa.gov/media/1606/download?inline= |
| Iowa | 2013 | American Institutes for Research. (2013). <i>State-funded preschool: Analyses of third-grade ITBS results of children enrolled in Iowa's 2007-08 statewide voluntary preschool programs</i> . Iowa, Department of Education. https://www.researchconnections.org/sites/default/files/pdf/rc25685.pdf |
| Iowa | 2018 | Iowa Department of Education. (2018). <i>Legislative report early childhood assessment</i> . https://www.researchconnections.org/sites/default/files/pdf/rc25685.pdf |
| Iowa | 2019 | Iowa Department of Education. (2019). <i>Legislative report early childhood assessment</i> . https://educate.iowa.gov/media/7258/download?inline= |

Iowa		Iowa Department of Education. (2013). <i>Statewide voluntary preschool programs for four-year-old children: 2010-2011 preliminary report</i> . Iowa, Department of Education.
Kentucky	2017	Office of Accountability (2017). <i>Preschool program review and full-day kindergarten</i> (Research Report No. 450). Legislative Research Commission. https://educate.iowa.gov/media/7258/download?inline=
Louisiana	2011	Picard Center for Child Development and Lifelong Learning. (2011). <i>LA 4 longitudinal study: The impact of LA 4 participation on academic achievement in 3rd & 4th grade: iLEAP & LEAP performance for cohorts 1-4</i> . University of Louisiana at Lafayette, Cecil J. Picard Center for Child Development and Lifelong Learning. https://picardcenter.louisiana.edu/sites/picardcenter/files/LA%204%20iLEAP%20and%20LEAP%20Report.pdf
Louisiana	2013	Picard Center for Child Development and Lifelong Learning. (2013). <i>Eighth-grade outcomes for LA 4 cohort 1 students</i> (Technical Brief). University of Louisiana at Lafayette, Cecil J. Picard Center for Child Development and Lifelong Learning. https://www.louisianabelieves.com/docs/default-source/early-childhood/picard-early-childhood-longitudinal-study.pdf?sfvrsn=2ab1713d_4
Louisiana	2021	Amus, G., Stokes, B., Ramey, S., & Ramey, C. (2021). <i>The LA 4 longitudinal study: Findings from the high school years</i> . A Project of the Council for A Better Louisiana. https://cabl.org/wp-content/uploads/2021/08/The-LA4-Longitudinal-Study.pdf
Louisiana		Louisiana Department of Education, Loyd J. Rockhold Center for Child Development, LSU HSC, Department of Psychiatry. (n.d.). <i>LA 4 prekindergarten program evaluation 2002–2003</i> . https://web.archive.org/web/20050312054836/http://www.doe.state.la.us/lde/uploads/3843.pdf

- Louisiana Cecil J. Picard Center for Child Development and Lifelong Learning. (n.d.). *LA4 longitudinal report*. University of Louisiana at Lafayette, Cecil J. Picard Center for Child Development and Lifelong Learning. <https://researchconnections.org/childcare/resources/20561>
- Louisiana University of Louisiana at Lafayette Cecil J. Picard Center for Child Development and Lifelong Learning, University of Alabama at Birmingham Center for Educational Accountability, & Georgetown University Center on Health & Education. (2007). *LA 4 & Starting Points prekindergarten program evaluation 2006-2007: Full report*. Louisiana, Department of Education. <https://picardcenter.louisiana.edu/sites/picardcenter/files/LA%204%20Annual%20Report%202007-08.pdf>
- Louisiana Cecil J. Picard Center for Child Development and Lifelong Learning. (n.d.). *2010-2011 LA 4 prekindergarten annual report: Final*. University of Louisiana at Lafayette, Cecil J. Picard Center for Child Development and Lifelong Learning. <https://picardcenter.louisiana.edu/sites/picardcenter/files/LA%204%20Annual%20Report%202010-2011.pdf>
- Louisiana University of Louisiana at Lafayette Cecil J. Picard Center for Child Development and Lifelong Learning, University of Alabama at Birmingham Center for Educational Accountability, & Georgetown University Center on Health & Education. (n.d.). *LA 4 & Starting Points prekindergarten program evaluation: 2007-2008 full report*. Louisiana, Department of Education. <https://picardcenter.louisiana.edu/sites/picardcenter/files/LA%204%20Annual%20Report%202006-07.pdf>

Louisiana	University of Louisiana at Lafayette Cecil J. Picard Center for Child Development and Lifelong Learning, University of Alabama at Birmingham Center for Educational Accountability, & Georgetown University Center on Health & Education. (n.d.). <i>LA 4 & Starting Points preschool program evaluation: 2008-2009 full report</i> . Louisiana, Department of Education. https://picardcenter.louisiana.edu/sites/picardcenter/files/LA%204%20Annual%20Report%202008-09.pdf		
Maryland	2012	Maryland State Department of Education. (2012). <i>Maryland Model for School Readiness 2011- 2012: Children entering school ready to learn</i> . http://archives.marylandpublicschools.org/NR/rdonlyres/BCFF0F0E-33E5-48DA-8F11-28CF333816C2/31940/2011_12_statereport_web_.pdf	
Maryland	2016	Maryland State Department of Education. (2017). <i>Readiness matters! The 2015-2016 kindergarten readiness assessment report</i> . https://files.eric.ed.gov/fulltext/ED572305.pdf	
Maryland	2020	Maryland State Department of Education. (2020). <i>Readiness matters! The 2019-2020 kindergarten readiness assessment report</i> . https://earlychildhood.marylandpublicschools.org/system/files/filedepot/4/200178_ready5_book_web.pdf	
Maryland	Baltimore	2020	Williams, S. C., Barajas, C. B, Milam, A. J., Olson, L., Leaf, P., & Connolly, F. (2020). Preparing students for success: Differential outcomes by preschool experience in Baltimore City, Maryland. <i>Prevention Science</i> , 21(4), 467-476. https://doi.org/10.1007/s11121-019-01069-3
Maryland	Montgomery County	2010	Zhao, H., & Modarresi, S. (2010). <i>Evaluating lasting effects of full-day prekindergarten program on school readiness, academic performance, and special education services</i> . Office of Shared Accountability, Montgomery County Public Schools. https://ww2.montgomeryschoolsmd.org/departments/sharedaccountability/reports/2010/10.04.30%20Pre%20K%20report.pdf

- | | | | |
|---------------|--------|------|--|
| Maryland | | 2002 | Maryland State Department of Education. (2002). <i>Children entering school ready to learn: School readiness baseline information, school year 2001-02 by state and county</i> . http://archives.marylandpublicschools.org/NR/rdonlyres/BCFF0F0E-33E5-48DA-8F11-28CF333816C2/2490/School_Readiness_Report_all1.pdf |
| Maryland | | 2012 | Maryland State Department of Education. (2012). <i>Children entering school ready to learn: Maryland model for school readiness 2012-2013 kindergarten assessment</i> . http://archives.marylandpublicschools.org/NR/rdonlyres/BCFF0F0E-33E5-48DA-8F11-28CF333816C2/35516/MMSR_SY20122013_StateReport_.pdf |
| Maryland | | 2018 | Maryland State Department of Education. (2018). Readiness matters! The 2017-2018 kindergarten readiness assessment report. https://files.eric.ed.gov/fulltext/ED589988.pdf |
| Maryland | | 2019 | Maryland State Department of Education. (2019). <i>Readiness matters! The 2018-2019 kindergarten readiness assessment report</i> . https://files.eric.ed.gov/fulltext/ED604311.pdf |
| Massachusetts | | 2018 | Hofer, K., Checkoway, A., Goodson, B., & Nichols, A. (2018). <i>Massachusetts preschool expansion grant (PEG) on children's early academic and executive function skills: A RDD evaluation in a mixed delivery system</i> [presentation]. APPAM. https://files.eric.ed.gov/fulltext/ED606103.pdf |
| Massachusetts | Boston | 2013 | Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. <i>Child Development</i> , 84(6), 2112-2130. https://doi.org/10.1111/cdev.12099 |
| Massachusetts | Boston | 2016 | Weiland, C. (2016). Impacts of the Boston prekindergarten program on the school readiness of young children with special needs. <i>Developmental Psychology</i> , 52(11), 1763. https://doi.org/10.1037/dev0000168 |

Massachusetts	Boston	2016	Yudron, M., & Weiland, C. (2016). <i>BPS KIDS: Piloting the Boston Public Schools' prekindergarten model in community-based organizations: Final report</i> . Boston Public Schools. http://www.bpsearlychildhood.weebly.com/uploads/1/0/1/3/10131776/bpskids_final_report_feb2016_11.pdf
Massachusetts	Boston	2020	Weiland, C., Unterman, R., Shapiro, A., Staszak, S., Rochester, S., & Martin, E. (2020). The effects of enrolling in oversubscribed prekindergarten programs through third grade. <i>Child Development</i> , 91, 1401-1422. https://doi.org/10.1111/cdev.13308
Massachusetts	Boston	2021	Gray-Lobe, G., Pathak, P. A., & Walters, C. R. (2021). The long-term effects of universal preschool in Boston (No. w28756). <i>National Bureau of Economic Research</i> . https://doi.org/10.1093/qje/qjac036
Michigan		2000	Xiang, Z., Schweinhart, L. J., Hohmann, C., Smith, C., & Smith, E. (2000). <i>Points of light: Third year report of the Michigan School Readiness Program evaluation</i> . High/Scope Educational Research Foundation. https://files.eric.ed.gov/fulltext/ED440736.pdf
Michigan		2002	Xiang, Z., & Schweinhart, L. J. (2002). <i>Effects five years later: The Michigan School Readiness Program evaluation through age 10</i> . High/Scope Educational Research Foundation. https://highscope.org/wp-content/uploads/2018/03/MSRP-through-age-10_2002-five-years-later.pdf
Michigan		2005	Lamy, C., Barnett, W., & Jung, K. (2005). <i>The effects of the Michigan School Readiness Program on young children's abilities at kindergarten entry</i> . National Institute for Early Education Research. https://highscope.org/wp-content/uploads/2018/03/nieer-report-results-effects-of-MSRP_2005.pdf
Michigan		2007	Malofeeva, E. V., Daniel-Echols, M., & Xiang, Z. (2007). <i>Findings from the Michigan School Readiness Program 6 to 8 follow up study</i> . High/Scope Educational Research Foundation. https://highscope.org/wp-content/uploads/2018/03/MSRP_six-eight-follow-up-2007.pdf

- | | | | |
|-----------|------------------|------|---|
| Michigan | | 2012 | Schweinhart, L. J., Xiang, Z., Daniel-Echols, M., Browning, K., & Wakabayashi, T. (2012). <i>Michigan Great Start Readiness Program evaluation 2012: High school graduation and grade retention findings</i> . HighScope Educational Research Foundation. https://www.michigan.gov/-/media/Project/Websites/mde/2012/05/30/Item_E_GSRP_Evaluation.pdf?rev=b6c43efabfe44dd1bb7e1361c6c974ef#page=3 |
| Michigan | | | Michigan Department of Education. (n.d.). Michigan's Great Start Readiness Program (GSRP) supports dual language learners. https://www.michigan.gov/mileap/-/media/Project/Websites/mileap/Documents/Early-Childhood-Education/gsrp/eval/kentdll_summaryoct2016finalada.pdf?rev=bf58e0b8889b4fa48f90ebe45bcd8ca5 |
| Michigan | Kalamazoo County | 2013 | Bartik, T. J. (2013). <i>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</i> (No. 13-198). Upjohn Institute Working Paper. https://research.upjohn.org/cgi/viewcontent.cgi?article=1215&context=up_workingpapers |
| Michigan | | | Michigan Department of Education. (n.d.). Michigan's Great Start Readiness Program (GSRP): Replicating evidence of impact in rural Michigan. https://www.michigan.gov/mileap/-/media/Project/Websites/mileap/Documents/Early-Childhood-Education/gsrp/eval/Thumb_summary08252016-ADA_533598_7.pdf?rev=9e3ad649e54a4b29b74e75b973c6816c |
| Minnesota | St. Paul | 2011 | Gozali-Lee, E., Mohr, C., & Mueller, D. (2011). <i>Project early kindergarten evaluation: Results through 2010-11 of a Saint Paul Public Schools initiative</i> . Wilder Research. https://www.wilder.org/sites/default/files/imports/PEK_Evaluation_2010-11results_10-11.pdf |

- | | | |
|-------------|------|---|
| Mississippi | 2015 | Mississippi Department of Education. (2015). <i>Kindergarten readiness assessment results for early learning collaboratives</i> . Mississippi, Department of Education. https://www.mdek12.org/sites/default/files/Offices/MDE/OEA/OPR/2015/k-readiness-results-collaboratives-2014-2015.pdf |
| Mississippi | 2015 | PEER Mississippi. (2015). <i>The Early Learning Collaborative Act of 2013: Evaluation of the operations and effectiveness of the program</i> . https://www.peer.ms.gov/sites/default/files/peer_publications/rpt600.pdf |
| Mississippi | 2016 | Mississippi Department of Education. (2016). <i>Pre-kindergarten assessment results for Early Learning Collaboratives and other four-year-old classrooms</i> . https://www.mdek12.org/sites/default/files/Offices/MDE/OEA/OPR/2016/Pre-K-SBE-REPORT_FINAL.pdf |
| Mississippi | 2017 | Mississippi Department of Education. (2017). <i>2016-2017 pre-kindergarten assessment results for Early Learning Collaboratives and other four-year-old classrooms</i> . https://www.mdek12.org/sites/default/files/Offices/MDE/OEA/OPR/2017/ELC_PK%20Results_Spring%202017.pdf |
| Mississippi | 2019 | PEER Mississippi. (2019). <i>Report on the Early Learning Collaborative Act of 2013: An evaluation of the operation and effectiveness of the program</i> . |
| Missouri | 2016 | Hall, C. R. (2016). <i>An examination of the long-term academic impacts of students who participated in the Missouri Preschool Program in rural southwest Missouri</i> . [Doctoral dissertation]. Lindenwood University. https://digitalcommons.lindenwood.edu/dissertations/262/ |
| Missouri | 2020 | Allin, M. J. (2020). <i>The effect of the Missouri Preschool Program on school readiness</i> . [Doctoral dissertation]. Lindenwood University. https://digitalcommons.lindenwood.edu/dissertations/40 |

- | | | |
|---|------|--|
| Multistate | 2018 | Bartik, T. J., & Hershbein, B. (2018). <i>Pre-K in the public schools: Evidence from within U.S. states</i> . (Upjohn Institute Working Paper 18-285). W. E. Upjohn Institute for Employment Research. https://doi.org/10.17848/wp18-285 |
| Multistate
(AK, CA,
MI, NM, NJ,
OK, SC,
WV) | 2018 | Barnett, W., Jung, K., Friedman-Krauss, A., Frede, E., Nores, M., Hustedt, J. T., Howes, C., & et al. (2018). State prekindergarten effects on early learning at kindergarten entry: An analysis of eight state programs. <i>AERA Open</i> , 4(2), 1-16. https://doi.org/10.1177/2332858418766291 |
| Multistate
(GA/OK) | 2013 | Cascio, E., & Whitmore, D. (2013). <i>The impacts of expanding access to high-quality preschool education</i> . Brookings Papers on Economic Activity. http://www.nber.org/papers/w19735 |
| Multistate
(MI, NJ, OK,
SC, WV) | 2008 | Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. <i>Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management</i> , 27(1), 122-154. https://doi.org/10.1002/pam.20310 |
| Nebraska | 2012 | Nebraska State Department of Education. (2012). <i>Early childhood education in Nebraska public school districts and educational service units: 2010-2011 state report</i> . https://www.education.ne.gov/wp-content/uploads/2017/07/2010-2011.pdf |
| Nebraska | 2014 | Nebraska State Department of Education. (2015). <i>Early childhood education in Nebraska public school districts and educational service units: 2013-2014 state report</i> . https://www.education.ne.gov/wp-content/uploads/2017/07/2013-2014.pdf |
| Nebraska | 2014 | Nebraska State Department of Education. (2014). <i>Early childhood education in Nebraska public school districts and educational service units: 2012-2013 state report</i> . https://www.education.ne.gov/wp-content/uploads/2017/07/2012-2013.pdf |

Nebraska	Omaha	2017	Buffett Early Childhood Institute. (2017). <i>Superintendents' early childhood plan evaluation: 2016-2017</i> . https://buffettinstitute.nebraska.edu/-/media/beci/docs/superindendents-plan-evaluation-report-final-12-1-17.pdf
Nebraska	Omaha	2016	Jackson, B., Johnson, J., Maxwell, C., & Moon, D. (2016). <i>The learning community of Douglas and Sarpay counties: 2015-2016 evaluation report</i> . https://www.cde.state.co.us/21stcclc/21stcclc2017report
Nevada		2004	Leitner, D. (2004). <i>Senate Bill 8: Nevada early childhood education (ECE) program: 2003-04 evaluation report</i> . Nevada, Office of Special Education, Elementary and Secondary Education, and School Improvement Programs. https://epubs.nsla.nv.gov/statepubs/epubs/31428003054844-2003-04.pdf
Nevada		2005	Nevada Department of Education. (2005). <i>Senate Bill 8: Nevada early childhood education (ECE) program: 2004-05 evaluation report</i> . Nevada, Department of Education. https://epubs.nsla.nv.gov/statepubs/epubs/31428003054844-2004-05.pdf
Nevada		2006	Leitner, D. (2006). <i>Senate Bill 525: Nevada early childhood education (ECE) program: 2005-06 evaluation report</i> . Nevada, Office of Special Education, Elementary and Secondary Education, and School Improvement Programs. https://epubs.nsla.nv.gov/statepubs/epubs/31428003054844-2005-06.pdf
Nevada		2007	Leitner, D. (2007). <i>Senate Bill 525: Nevada early childhood education (ECE) program: 2006-07 evaluation report</i> . Nevada, Office of Special Education, Elementary and Secondary Education, and School Improvement Programs. https://epubs.nsla.nv.gov/statepubs/epubs/31428003054844-2006-07.pdf
Nevada		2009	Leitner, D. (2009). <i>Assembly Bill 627: Nevada early childhood education (ECE) program: FY 2007-08 final evaluation report</i> . Nevada, Office of Special Education, Elementary and Secondary Education, and School Improvement Programs. https://epubs.nsla.nv.gov/statepubs/epubs/31428003054844-2007-08.pdf

- | | | |
|------------|------|---|
| Nevada | 2011 | Leitner, D. (2011). <i>Assembly Bill 563: Nevada early childhood education (ECE) program: Building a foundation for school readiness and success in K-12 and beyond: FY 2009-10 final evaluation report</i> . Nevada, Office of Special Education, Elementary and Secondary Education, and School Improvement Programs. https://epubs.nsla.nv.gov/statepubs/epubs/31428003054844-2009-10.pdf |
| Nevada | 2013 | Leitner, D. (2013). <i>Assembly Bill 579: Nevada early childhood education (ECE) program: Building a foundation for school readiness and success in prek-12 and beyond: FY 2012-13 evaluation report</i> . Nevada, Office of Educational Opportunity. https://www.leg.state.nv.us/Division/Research/Library/Documents/ReportsToLeg/2013-2015/193-14.pdf#page=12 |
| Nevada | 2015 | Leitner, D. (2015). <i>Senate Bill (SB) 522: Nevada early childhood education (ECE) program: Building a foundation for school readiness and success in preK-12 and beyond: FY 2014-15: Evaluation report</i> . Nevada, Department of Education. https://epubs.nsla.nv.gov/statepubs/epubs/31428003054844-2014-15.pdf |
| Nevada | 2009 | Leitner, D. (n.d.). <i>Assembly Bill 627: Nevada early childhood education (ECE) program: Building a foundation for school readiness and success in K-12 and beyond: FY 2008-09 final evaluation report</i> . Nevada, Office of Special Education, Elementary and Secondary Education, and School Improvement Programs. https://epubs.nsla.nv.gov/statepubs/epubs/31428003054844-2008-09.pdf |
| New Jersey | 2004 | Frede, E., Lamy, C., Seplocha, H., Strasser, J., Jambunathan, S., Juncker, J., & Wolock, E. (2004). <i>A rising tide: Classroom quality and language skills in the Abbott preschool program: Year two preliminary update, 2003-2004</i> . New Jersey Office of Early Childhood Education. https://nieer.org/sites/default/files/2023-09/A-Rising-Tide.pdf |

- New Jersey 2004 Lamy, C., Frede, E., Seplocha, H., Strasser, J., Jambunathan, S., Juncker, J., Ferrar, H. M., & et al. (2004). *Inch by inch, row by row, gonna make this garden grow: Classroom quality and language skills in the Abbott preschool program: Year one report, 2002-2003*. New Jersey Office of Early Childhood Education.
<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3c5ba1aba84a0df82562f42d10b19f375751da78>
- New Jersey 2005 Lamy, C., Frede, E., Seplocha, H., Strasser, J., Jambunathan, S., Juncker, J., & Wolock, E. (2005). *Giant steps for the littlest children: Progress in the sixth year of the Abbott preschool program: Year three initial update, 2004-2005*. New Jersey Office of Early Childhood Education.
- New Jersey 2005 Lamy, C., Barnett, W. S., & Jung, K. (2005). *The effects of New Jersey's Abbott preschool program on young children's school readiness*. National Institute for Early Education Research.
<https://web.archive.org/web/20220609005047/https://nieer.org/wp-content/uploads/2005/12/nj.pdf>
- New Jersey 2005 Frede, E. (2005). *Assessment in a continuous improvement cycle: New Jersey's Abbott preschool program*. National Early Childhood Accountability Task Force.
- New Jersey 2007 Frede, E., Jung, K., Barnett, W., Lamy, C., & Figueras, A. (2007). *The Abbott preschool program longitudinal effects study (APPLES): Interim report*. National Institute for Early Education Research.
<https://nieer.org/sites/default/files/2023-08/apples.pdf>
- New Jersey 2009 Frede, E., Jung, K., Barnett, W. S., & Figueras, A. (2009). *The APPLES blossom: Abbott preschool program longitudinal effects study (APPLES), Preliminary results through 2nd grade interim report*. National Institute for Early Education Research.
https://web.archive.org/web/20230617060231/http://nieer.org/wp-content/uploads/2012/04/apples_second_grade_results.pdf

- | | | |
|------------|------|--|
| New Jersey | 2013 | Barnett, S. W., Jung, K., Youn, M., & Frede, E. C. (2013). <i>Abbott preschool program longitudinal effects study: Fifth grade follow-up</i> . National Institute for Early Education Research. Rutgers. https://nieer.org/sites/default/files/2023-07/apples205th20grade.pdf |
| New Jersey | 2021 | Barnett, W. S., & Jung, K. (2021). Effects of New Jersey's Abbott preschool program on children's achievement, grade retention, and special education through tenth grade. <i>Early Childhood Research Quarterly</i> , 56, 248-259. https://doi.org/10.1016/j.ecresq.2021.04.001 |
| New Mexico | 2007 | Hustedt, J. T., Barnett, W., Jung, K., & Pew Charitable Trusts. (2007). <i>The effects of the New Mexico prek initiative on young children's school readiness</i> . National Institute for Early Education Research. https://www.newmexicoprek.org/wpfd_file/the-effects-of-the-new-mexico-prek-initiative-on-young-childrens-school-readiness-august-2007s-school-readiness/ |
| New Mexico | 2008 | Hustedt, J. T., Barnett, W., Jung, K., & Figueras, A. (2008). <i>Impacts of New Mexico prek on children's school readiness at kindergarten entry: Results from the second year of a growing initiative</i> . National Institute for Early Education Research. https://nieer.org/sites/default/files/2023-08/newmexicordd0608.pdf |
| New Mexico | 2009 | Hustedt, J. T., Barnett, W., Jung, K., & Figueras, A. (2009). <i>Continued impacts of New Mexico prek on children's readiness for kindergarten: Results from the third year of implementation</i> . National Institute for Early Education Research. https://www.newmexicoprek.org/wpfd_file/continued-impacts-of-new-mexico-prek-on-childrens-readiness-for-kindergarten-results-from-the-third-year-of-implementation-september-2009-2/ |

- | | | | |
|------------|------|--|---|
| New Mexico | 2009 | Hustedt, J. T., Barnett, W., Jung, K., & Goetze, L. D. (2009). <i>The New Mexico prek evaluation: Results from the initial four years of a new state preschool initiative: Final report</i> . National Institute for Early Education Research. https://www.newmexicoprek.org/download/206/english/2506/the-new-mexico-prek-evaluation-results-from-the-initial-four-years-of-a-new-state-preschool-initiative-final-report-2.pdf | |
| New Mexico | 2010 | Hustedt, J. T., Barnett, W. S., Jung, K., & Friedman, A. H. (2010). <i>The New Mexico prek evaluation: Impacts from the fourth year (2008-2009) of New Mexico's state-funded prek program</i> . National Institute for Early Education Research. https://nieer.org/sites/default/files/2023-08/newmexicordd1110-1.pdf | |
| New Mexico | 2016 | New Mexico Legislative Finance Committee. (2016). <i>Accountability report: Early childhood</i> . https://nmlegis.gov/Entity/LFC/Documents/Program_Evaluation_Reports/2016%20Early%20Childhood%20Accountability%20Report.pdf | |
| New Mexico | 2020 | New Mexico Legislative Finance Committee. (2020). Program evaluation: Prekindergarten quality and educational outcomes. https://www.nmlegis.gov/Entity/LFC/Documents/Program_Evaluation_Reports/Prekindergarten%20Quality%20and%20Educational%20Outcomes.pdf | |
| New Mexico | 2021 | Hustedt, J. T., Jung, K., Friedman-Krauss, A. H., Barnett, W. S., & Slicker, G. (2021). Impacts of the New Mexico prek initiative by children's race/ethnicity. <i>Early Childhood Research Quarterly</i> , 54, 194-203. https://doi.org/10.1016/j.ecresq.2020.09.006 | |
| New York | NYC | 2016 | Westat, Inc., Metis Associates, Inc., & Branch Associates, Inc. (2016). <i>Pre-K for all: Snapshot of student learning</i> . Center for Economic Opportunity. https://www.nyc.gov/assets/opportunity/pdf/Westat_Metis_Branch_PreK_Study_Snapshot_of_Student_Learning_Final.pdf |

- | | | |
|-------------------|------|--|
| North
Carolina | 2005 | Peisner-Feinberg, E. S., Maris, C. L., & More at Four Evaluation Team. (2005). <i>Evaluation of the North Carolina More at Four pre-kindergarten program: Year 3 report: July 1, 2003-June 30, 2004</i> . University of North Carolina, FPG Child Development Institute.
https://fpg.unc.edu/sites/fpg.unc.edu/files/resource-files/MAF_Yr3_full_report.pdf |
| North
Carolina | 2006 | Peisner-Feinberg, E. S., Maris, C. L., & More at Four Evaluation Team. (2006). <i>Evaluation of the North Carolina More at Four pre-kindergarten program: Children's longitudinal outcomes and classroom quality in kindergarten</i> . University of North Carolina, FPG Child Development Institute.
https://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/MAF_Yr4_pt2_full_report.pdf |
| North
Carolina | 2005 | Peisner-Feinberg, E. S., Maris, C. L., & More at Four Evaluation Team. (2005). <i>Evaluation of the North Carolina More at Four pre-kindergarten program: Year 2 report: July 1, 2002-June 30, 2003</i> . University of North Carolina, FPG Child Development Institute.
https://fpg.unc.edu/sites/fpg.unc.edu/files/resource-files/MAF_Yr2_full_report.pdf |
| North
Carolina | 2007 | Peisner-Feinberg, E. S., Schaaf, J. M., & More at Four Evaluation Team. (2007). <i>Evaluation of the North Carolina More at Four pre-kindergarten program: Year 5 report: July 1, 2005-June 30, 2006: Children's outcomes & program quality in the fifth year</i> . University of North Carolina, FPG Child Development Institute. https://files.eric.ed.gov/fulltext/ED499809.pdf |
| North
Carolina | 2010 | Peisner-Feinberg, E. S., & Schaaf, J. M. (2010). <i>Long-term effects of the North Carolina More at Four Pre-Kindergarten Program: Children's reading and math skills at third grade</i> . University of North Carolina, Frank Porter Graham Child Development Institute.
https://fpg.unc.edu/sites/fpg.unc.edu/files/resource-files/MAF_EOG_full_report.pdf |

- | | | |
|-------------------|------|--|
| North
Carolina | 2011 | Peisner-Feinberg, E. S., & Schaaf, J. M. (2011). <i>Summary of key findings: Effects of the North Carolina More at Four Prekindergarten Program on children's school readiness skills</i> . University of North Carolina, FPG Child Development Institute. https://fpg.unc.edu/sites/fpg.unc.edu/files/resource-files/MAF_Yr9_key_findings.pdf |
| North
Carolina | 2014 | Ladd, H. F., Muschkin, C. G., & Dodge, K. A. (2014). From birth to school: Early childhood initiatives and third-grade outcomes in North Carolina. <i>Journal of Policy Analysis and Management</i> , 33(1), 162-187. https://doi.org/10.1002/pam.21734 |
| North
Carolina | 2014 | Peisner-Feinberg, E. S., LaForett, D. R., Schaaf, J. M., Hildebrandt, L. M., Sideris, J., & Pan, Y. (2014). <i>Children's outcomes and program quality in the North Carolina Pre-Kindergarten Program: 2012–2013 Statewide evaluation</i> . University of North Carolina, FPG Child Development Institute. https://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/NC%20Pre-K%20Eval%202012-2013%20Report.pdf |
| North
Carolina | 2015 | Muschkin, C. G., Ladd, H. F., & Dodge, K. A. (2015). Impact of North Carolina's early childhood initiatives on special education placements in third grade. <i>Educational Evaluation and Policy Analysis</i> , 37(4), 478-500. https://doi.org/10.3102/0162373714559096 |
| North
Carolina | 2015 | Peisner-Feinberg, E. S., Schaaf, J. M., Hildebrandt, L. M., Pan, Y. & Warnaar, B. L. (2015). <i>Children's kindergarten outcomes and program quality in the North Carolina Pre-Kindergarten Program: 2013–2014 Statewide evaluation</i> . University of North Carolina, FPG Child Development Institute. https://fpg.unc.edu/sites/fpg.unc.edu/files/resource-files/NC%20Pre-K%20Eval%202013-2014%20Report.pdf |

- | | | |
|-------------------|------|--|
| North
Carolina | 2016 | Peisner-Feinberg, E. S., Garwood, J. D., & Mokrova, I. L. (2016). <i>Children's pre-k experiences and outcomes in the North Carolina Pre-Kindergarten Program: 2014–2015 statewide evaluation</i> . University of North Carolina, FPG Child Development Institute.
https://fpg.unc.edu/sites/fpg.unc.edu/files/resource-files/2014-15%20NC%20Pre-K%20Eval%20Report%20Final.pdf |
| North
Carolina | 2017 | Dodge, K. A., Bai, Y., Ladd, H. F., & Muschkin, C. G. (2017). Impact of North Carolina's early childhood programs and policies on educational outcomes in elementary school. <i>Child Development</i> , 88(3), 996-1014. https://doi.org/10.1111/cdev.12645 |
| North
Carolina | 2017 | Peisner-Feinberg, E. S., Mokrova, I. L., & Anderson, T. L. (2017). <i>Effects of participation in the North Carolina Pre-Kindergarten Program at the end of kindergarten: 2015-2016 statewide evaluation</i> . University of North Carolina, FPG Child Development Institute.
https://fpg.unc.edu/sites/fpg.unc.edu/files/resource-files/2014-15%20NC%20Pre-K%20Eval%20Report%20Final.pdf |
| North
Carolina | 2019 | Peisner-Feinberg, E., Zadrozny, S., Kuhn, L., & Van Manen, K. (2019). Effects of the North Carolina Pre-Kindergarten program: Findings through pre-k of a small-scale RCT study. University of North Carolina, FPG Child Development Institute. https://fpg.unc.edu/sites/fpg.unc.edu/files/resource-files/NC%20Pre-K%20Eval%202017-18%20Exec%20sum.pdf |
| North
Carolina | 2020 | Muschkin, C. G., Ladd, H. F., Dodge, K. A., & Bai, Y. (2020). Gender differences in the impact of North Carolina's early care and education initiatives on student outcomes in elementary school. <i>Educational Policy</i> , 34(2), 377-407. https://doi.org/10.1177/0895904818773901 |
| North
Carolina | | Peisner-Feinberg, E. S., Kuhn, L., Zadrozny, S., Foster, T., & Burchinal, M. (2020). <i>Kindergarten follow-up findings from a small-scale RCT study of the North Carolina Pre-Kindergarten Program</i> . University of North Carolina, School of Education. |

- | | | | |
|-------------------|------------|------|---|
| North
Carolina | | | Carr, R. C., Peisner-Feinberg, E. S., Kaplan, R., & Mokrova, I. L. (2021). Effects of North Carolina's pre-kindergarten program at the end of kindergarten: Contributions of school-wide quality. <i>Journal of Applied Developmental Psychology</i> , 76, 101317. https://doi.org/10.1016/j.appdev.2021.101317 |
| North
Carolina | 2008 | | Peisner-Feinberg, E. S., & Schaaf, J. M. (2008). <i>Evaluation of the North Carolina More at Four pre-kindergarten program: Performance and progress in the seventh year (2007-2008): Year 7 report (July 1, 2007-June 30, 2008)</i> . University of North Carolina, FPG Child Development Institute. https://fpg.unc.edu/sites/fpg.unc.edu/files/resources/reports-and-policy-briefs/MAF_Yr7_full_report.pdf |
| North
Carolina | 2009 | | Peisner-Feinberg, E. S., & Schaaf, J. M. (2009). <i>Evaluation of the North Carolina More at Four Pre-kindergarten Program: Year 8 report (July 1, 2008-June 30, 2009): A look across time at children's outcomes and classroom quality from pre-k through kindergarten</i> . University of North Carolina, FPG Child Development Institute. https://fpg.unc.edu/sites/fpg.unc.edu/files/resource-files/MAF_Yr8_full_ |
| Ohio | 2020 | | Anthony, E., Rohrman, S., Cho, Y., Atwell, M. S., & Fischer, R. (2020). <i>Cuyahoga county's universal pre-kindergarten program: Evaluating its effects on kindergarten readiness and third grade reading proficiency</i> . Cuyahoga County Office of Early Childhood / Invest In Children. https://hhs.cuyahogacounty.gov/docs/default-source/invest-in-children-documents/upk-outcomes-2020-final.pdf?sfvrsn=5df0f50a_2 |
| Ohio | Cincinnati | 2020 | Innovations in Community Research and Program Evaluation, & Cincinnati Public Schools. (2020). <i>Cincinnati kindergarten readiness longitudinal evaluation: Preschool attendance: Kindergarten readiness and academic outcomes</i> . United Way of Greater Cincinnati. https://www.uwgc.org/files/UWGC_kindergarten-longitudinal-evaluation-final-1-29-20.pdf |

Ohio	Cincinnati	2020	Cincinnati Preschool Promise, Innovations in Community Research and Program Evaluation, & Cincinnati Public Schools. (2020). <i>Cincinnati preschool promise program evaluation report: Year 3 2019-2020</i> . https://cincy-promise.org/wp-content/uploads/2021/06/_CPP-Evaluation-Report-9.18.2020.pdf
Oklahoma		2005	Lamy, C., Barnett, W. S., & Jung, K. (2005). <i>The effects of Oklahoma's early childhood four-year-old program on young children's school readiness</i> . National Institute for Early Education Research. https://sde.ok.gov/sites/default/files/documents/files/Effects%20of%20OK%20EC4year-old%20Program-School%20Readiness.pdf
Oklahoma	Tulsa	2015	Smith, A. (2015). <i>The long-run effects of universal pre-k on criminal activity</i> . Social Science Research Network. http://dx.doi.org/10.2139/ssrn.2685507
Oklahoma	Tulsa	2005	Gormley, W. T., Jr., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. <i>Developmental Psychology</i> , 41(6), 872–884. https://doi.org/10.1037/0012-1649.41.6.872
Oklahoma	Tulsa	2005	Gormley Jr, W. T., & Phillips, D. (2005). The effects of universal pre-k in Oklahoma: Research highlights and policy implications. <i>Policy Studies Journal</i> , 33(1), 65-82. https://doi.org/10.1111/j.1541-0072.2005.00092.x
Oklahoma	Tulsa	2005	Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-k program. <i>The Journal of Human Resources</i> , 40(3), 533–558. http://www.jstor.org/stable/4129551
Oklahoma	Tulsa	2008	Gormley, W. T., Jr. (2008). The effects of Oklahoma's pre-K program on Hispanic children. <i>Social Science Quarterly</i> , 89(4), 916–936. https://doi.org/10.1111/j.1540-6237.2008.00591.x
Oklahoma	Tulsa	2008	Gormley, W. T., Phillips, D., & Gayer, T. (2008). Preschool programs can boost school readiness. <i>Science</i> , 320(27), 1723. https://doi.org/10.1126/science.1156019

Oklahoma	Tulsa	2008	Corrington, M. (2008). <i>Understanding gender differences in the effects of the Tulsa pre-k program on children's school readiness</i> [Masters Thesis, Georgetown University].
Oklahoma	Tulsa	2010	Gormley, Jr., W. T., Phillips, D. A., Adelstein, S., & Shaw, C. (2010). Head Start's comparative advantage: Myth or reality?. <i>Policy Studies Journal</i> , 38(3), 397-418. https://doi.org/10.1111/j.1541-0072.2010.00367.x
Oklahoma	Tulsa	2011	Gormley, W. T., Jr., Phillips, D. A., Newmark, K., Welti, K., & Adelstein, S. (2011). Social-emotional effects of early childhood education programs in Tulsa. <i>Child Development</i> , 82(6), 2095–2109. https://doi.org/10.1111/j.1467-8624.2011.01648.x
Oklahoma	Tulsa	2012	Bartik, T. J., Gormley, W., & Adelstein, S. (2012). Earnings benefits of Tulsa's pre-K program for different income groups. <i>Economics of Education Review</i> , 31(6), 1143-1161. 10.1016/j.econedurev.2012.07.016
Oklahoma	Tulsa	2012	Phillips, D. A., & Meloy, M. E. (2012). High-quality school-based pre-k can boost early learning for children with special needs. <i>Exceptional Children</i> , 78(4), 471-490. https://doi.org/10.1177/001440291207800405
Oklahoma	Tulsa	2013	Gormley, W., Kitchens, K., & Adelstein, S. (2013). <i>Do middle-class families benefit from high-quality pre-k</i> . CROCUS Policy Brief. Center for Research on Children in the US, Georgetown University. https://www.fcd-us.org/wp-content/uploads/2016/04/Policy-Brief_draft_071613_0.pdf
Oklahoma	Tulsa	2015	Hill, C. J., Gormley, W. T., Jr., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? <i>Early Childhood Research Quarterly</i> , 32, 60–79. https://doi.org/10.1016/j.ecresq.2014.12.005
Oklahoma	Tulsa	2016	Johnson, A. D., Markowitz, A. J., Hill, C. J., & Phillips, D. A. (2016). Variation in impacts of Tulsa pre-K on cognitive development in kindergarten: The role of instructional support. <i>Developmental Psychology</i> , 52(12), 2145–2158. https://doi.org/10.1037/dev0000226

Oklahoma	Tulsa	2016	Bartik, T., Belford, J., Gormley, W. T., & Anderson, S. (2016). <i>A benefit-cost analysis of the Tulsa universal pre-k program</i> . Upjohn Institute For Employment Research. https://doi.org/10.17848/wp16-261
Oklahoma	Tulsa	2017	Lawrimore, T., & Gormley, Jr., W. T. (2017). <i>Tulsa pre-K alumni are more likely to complete Algebra I early</i> . Georgetown University, Center for Research on Children in the United States. https://georgetown.app.box.com/s/05ceobhuxnii6c81kpcp41oibt9hq07x
Oklahoma	Tulsa	2018	Phillips, D. A., Anderson, S., & Gormley, W. T., Jr. (2018). The dog that didn't bark: Preschool education and middle-school attitudes in Tulsa. <i>Applied Developmental Science</i> , 22(3), 200–216. https://doi.org/10.1080/10888691.2016.1263154
Oklahoma	Tulsa	2018	Gormley Jr, W. T., Phillips, D., & Anderson, S. (2018). The effects of Tulsa's Pre-K program on middle school student performance. <i>Journal of Policy Analysis and Management</i> , 37(1), 63-87. https://doi.org/10.1002/pam.22023
Oregon		2015	Wilson, A. (2015). <i>An examination of the Oregon kindergarten assessment</i> [Doctoral Dissertation, University of Oregon].
Pennsylvania		2009	Pennsylvania Office of Child Development and Early Learning. (2009). <i>Pennsylvania Pre-K Counts: End of year report: 2008-2009</i> . http://www.pde.state.pa.us/early_childhood/lib/early_childhood/PKCendofyrrpt08-09small.pdf
Pennsylvania		2010	Pennsylvania Office of Child Development and Early Learning. (2010). <i>Kindergarten readiness with PA Pre-K Counts: A statewide summary of local evaluations</i> . (Research Brief Issue 1). https://www.pakeys.org/wp-content/uploads/2023/06/OCDEL-Research-Brief-1-Kindergarten-Readiness-with-PA-Pre-K-Counts.pdf
Pennsylvania		2010	Pennsylvania Office of Child Development and Early Learning. (2010). <i>Pennsylvania Pre-K Counts: End of year report: 2009-2010</i> . https://files.eric.ed.gov/fulltext/ED513403.pdf

- | | | |
|----------------|------|--|
| Pennsylvania | 2013 | Pennsylvania Office of Child Development and Early Learning (2014). <i>Children's progress update: Pennsylvania Pre-K Counts, 2012-13</i> . https://www.pakeys.org/wp-content/uploads/2023/06/Child-Outcomes-for-children-participating-in-PA-Pre-K-Counts-2012-13-1.pdf |
| Pennsylvania | 2015 | Pennsylvania Department of Education (2015). <i>Quality early education helps close the achievement gap</i> . https://www.pakeys.org/wp-content/uploads/2023/06/Quality-early-education-helps-close-the-achievement-gap-Dec.-2015.pdf |
| Pennsylvania | 2020 | Peisner-Feinberg, E., Soliday Hong, S., Yazejian, N., Zadrozny, S., & Burchinal, M.(2020). <i>Kindergarten impacts of the Pennsylvania Pre-K Counts program: A statewide evaluation</i> . The University of North Carolina, School of Education and Frank Porter Graham Child Development Institute. https://fpg.unc.edu/sites/fpg.unc.edu/files/resource-files/PA-PKC-Impact-Study-Report_Final.pdf |
| Pennsylvania | 2009 | Bagnato, S. J., Salaway, J., & Suen, H. K. (2009). <i>Pre-K Counts in Pennsylvania for youngsters' early school success: Authentic outcomes for an innovative prevention and promotion initiative: Research results of SPECS for Pre-K Counts: An independent authentic program evaluation research initiative (2005-2009): 2009 final research report</i> . University of Pittsburgh, Early Childhood Partnerships. https://hceconomics.uchicago.edu/sites/default/files/file_uploads/PennsylvaniaHeinzFoundationPre-KCountsFinalReport.pdf |
| South Carolina | 2004 | South Carolina State Department of Education Office of Research. (2004). <i>What is the penny buying for South Carolina?: Child development programs for four-year-olds: Longitudinal studies of later academic achievement, 1995-96 through 1999-2000 and 2000-01 through 2001-02</i> . https://www.scstatehouse.gov/Archives/DeptofEducation/PennyBuy2004.pdf |

- | | | |
|-------------------|------|---|
| South
Carolina | 2005 | Lamy, C., Barnett, W. S., & Jung, K. (2005). <i>The effects of South Carolina's early childhood programs on young children's school readiness</i> . National Institute for Early Education Research. https://web.archive.org/web/20230804001217/https://nieer.org/wp-content/uploads/2005/12/sc.pdf |
| South
Carolina | 2010 | South Carolina Education Oversight Committee. (2010). <i>Child development education pilot program (CDEPP): 2009-10 student and classroom assessment report</i> . https://web.archive.org/web/20161222051415/http://www.eoc.sc.gov/In%20the%20News/At-A-Glance/2010/CDEEPEvaluationRevisedOnePagerrek.pdf |
| South
Carolina | 2015 | South Carolina First Steps. (2015). <i>Annual Report 2015: Further steps to school readiness evaluation of the South Carolina First Steps to School Readiness initiatives</i> . https://www.scstatehouse.gov/reports/FirstSteps/2015-SC%20First%20Steps%20to%20School%20Readiness%20Annual%20Report.pdf |
| South
Carolina | | Greer, F., Googe, H., & Brown, W. H. (n.d.). <i>Analysis of individual child assessment for a sample of children who are receiving Child Development Education Pilot Program (CDEPP) services: Spring 2007 and fall 2007</i> . South Carolina Education Oversight Committee. https://dc.statelibrary.sc.gov/items/df50afac-dffb-43a2-a953-2ec4fe3f15d1 |
| South
Carolina | 2006 | High/Scope Educational Research Foundation & South Carolina First Steps to School Readiness. (2006). <i>From implementation to impact: An evaluation of the South Carolina First Steps to School Readiness program</i> . South Carolina State Library. https://dc.statelibrary.sc.gov/items/208120aa-f928-4af9-b027-605f240aaef2 |
| South
Carolina | 2009 | South Carolina Education Oversight Committee. (2009). <i>2008-2009 implementation & expansion of the Child Development Education Pilot Program (CDEPP): Evaluation report</i> . https://www.scstatehouse.gov/Archives/EducationOversightComm/CDEPPReportJanuary12009.pdf |

- | | | |
|-------------------|------|---|
| South
Carolina | 2010 | South Carolina Education Oversight Committee. (2010). <i>2009-2010 implementation & expansion of the Child Development Education Pilot Program (CDEPP): Evaluation report</i> .
https://www.scstatehouse.gov/Archives/EducationOversightComm/CDEPPReport1-12-10.pdf |
| South
Carolina | 2010 | Browning, K. G., & Xiang, Z. (2010). <i>Further steps to school readiness: 2009 evaluation of the South Carolina First Steps to School Readiness initiatives</i> . South Carolina First Steps to School Readiness.
https://www.scfirststeps.org/wp-content/uploads/2014/12/South-Carolina-First-Steps-to-School-Readiness-2009-Evaluation.pdf |
| South
Carolina | 2014 | South Carolina Education Oversight Committee. (2014). <i>2013-14: Expansion of the SC Child Development Education Pilot Program (CDEPP): Report</i> .
https://www.scstatehouse.gov/reports/EducationOversightComm/ExpansionOfSCCDEPPReportFromSCEOC.pdf |
| South
Carolina | 2015 | South Carolina Education Oversight Committee. (2015). <i>SC child early reading development & education program: Report for FY 2014 & 2015</i> .
https://eoc.sc.gov/sites/eoc/files/Documents/2015%20CDEP%20Report.pdf |
| South
Carolina | 2016 | South Carolina Education Oversight Committee. (2016). <i>FY 2014-15 & 2015-16: Evaluation of state-funded full-day 4K: Part I</i> . https://eoc.sc.gov/sites/eoc/files/Documents/CDEP%20Report%20-%20Final%201-19-16.pdf |
| South
Carolina | 2017 | South Carolina Education Oversight Committee. (2017). <i>FY 2015-16 & 2016-17: State-funded full-day 4K: Evaluation pursuant to Provisos 1.62. and 1A.30. of the 2016-17 General Appropriation Act</i> .
https://eoc.sc.gov/sites/eoc/files/Documents/State-Funded%20Full-Day%204K%20Evaluation%20-%20Final%201.17.17.pdf |

- | | | |
|-------------------|------|--|
| South
Carolina | 2019 | Institute for Families in Society. (2019). <i>Evaluation of South Carolina First Steps to School Readiness: Local partnerships from 2013-2014 to 2017-2018</i> . University of South Carolina. https://www.scstatehouse.gov/reports/FirstSteps/Evaluation-of-South-Carolina-First-Steps-to-School-Readiness-Local-Partnerships-from-2013-2014-to-2017-2018-FINAL.pdf |
| Tennessee | 2007 | Strategic Research Group (2007). <i>Assessing the effectiveness of Tennessee's pre-kindergarten program: First interim report</i> . Offices of Research and Education Accountability. https://comptroller.tn.gov/content/dam/cot/orea/advanced-search/orea-reports-2007/2007_OREA_SRGPreKPrelimSept.pdf |
| Tennessee | 2008 | Strategic Research Group (2008). <i>Assessing the effectiveness of Tennessee's pre-kindergarten program: Second interim report</i> . Tennessee Office of Research and Educational Accountability. https://comptroller.tn.gov/content/dam/cot/orea/advanced-search/orea-reports-2008/2008_OREA_SRGPreKSecondInterimAug.pdf |
| Tennessee | 2009 | Coburn, J. L. (2009). <i>The effect of Tennessee's prekindergarten programs on young children's school readiness skills: A regression discontinuity design</i> [Doctoral Dissertation, Tennessee Technical University]. ProQuest. |
| Tennessee | 2010 | Strategic Research Group. (2010). <i>Assessing the effectiveness of Tennessee's pre-kindergarten program: Third interim report</i> . Comptroller of the Treasury. https://digitalcommons.memphis.edu/cgi/viewcontent.cgi?article=1146&context=govpubs-tn-comptroller-office-research-education-accountability |
| Tennessee | 2011 | Lipsey, M. W., Farran, D., Bilbrey, C., Hofer, K. G., & Dong, N. (2011). <i>Initial results of the evaluation of the Tennessee Voluntary Pre-K Program</i> . Vanderbilt University, Peabody Research Institute. https://my.vanderbilt.edu/tnprekevaluation/files/2013/10/April2011_PRI_Initial_TN-VPK_ProjectResults.pdf |

- | | | |
|-----------|------|--|
| Tennessee | 2011 | Strategic Research Group. (2011). <i>Assessing the impact of Tennessee's pre-kindergarten program: Final report</i> . Tennessee Offices of Research and Educational Accountability. https://comptroller.tn.gov/content/dam/cot/orea/advanced-search/2011/2011_OREA_SRGPreKFinal.pdf |
| Tennessee | 2013 | Lipsey, M. W., Hofer, K. G., Dong, N., Farran, D. C., & Bilbrey, C. (2013). <i>Evaluation of the Tennessee Voluntary Prekindergarten Program: End of pre-K results from the randomized control design</i> . Vanderbilt University, Peabody Research Institute. https://files.eric.ed.gov/fulltext/ED566667.pdf |
| Tennessee | 2014 | Lipsey, M. W. (2014). <i>Immediate and longitudinal effects of the Tennessee Voluntary Pre-Kindergarten program on student cognitive and non-cognitive outcomes</i> . SREE Spring 2014 Conference Abstract. |
| Tennessee | 2015 | Lipsey, M. W., Farran, D.C., & Hofer, K. G., (2015). <i>A randomized control trial of the effects of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade</i> . Vanderbilt University, Peabody Research Institute. https://cdn.vanderbilt.edu/vu-my/wp-content/uploads/sites/1147/2013/10/14112801/VPKthrough3rd_final_withcover.pdf |
| Tennessee | 2018 | Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. <i>Early Childhood Research Quarterly</i> , 45, 155-176. https://doi.org/10.1016/j.ecresq.2018.03.005 |
| Tennessee | 2019 | Watts, T. W., Duncan, G. J., & Rivas, M. (2019). <i>A reanalysis of impacts of the Tennessee Voluntary Prekindergarten Program</i> (EdWorkingPaper No.19-28). Annenberg Institute at Brown University. http://edworkingpapers.com/ai19-28 |
| Tennessee | 2020 | Pearman, F. A. (2020). The moderating effect of neighborhood poverty on preschool effectiveness: Evidence from the Tennessee Voluntary Prekindergarten Experiment. <i>American Educational Research Journal</i> , 57(3), 1323-1357. https://doi.org/10.3102/0002831219872977 |

- | | | | |
|-----------|---------|--|---|
| Tennessee | 2020 | Pion, G. M., & Lipsey, M. W. (2021). Impact of the Tennessee Voluntary Prekindergarten Program on children's literacy, language, and mathematics skills: Results from a regression-discontinuity design. <i>AERA Open</i> , 7. https://doi.org/10.1177/23328584211041353 | |
| Tennessee | | Strategic Research Group (n.d.). <i>Assessing the effectiveness of Tennessee's pre-kindergarten program: Annual report 2008-2009</i> . Offices of Research and Education Accountability. https://comptroller.tn.gov/content/dam/cot/orea/advanced-search/orea-reports-2009/SRGAnnualReport08-09.pdf | |
| Tennessee | Memphis | 2020 | First Eight Memphis. (2020). <i>Investing in Pre-K as a key strategy to improve long-term economic viability quarter 2 pre-k report: School year 2020-2021</i> . |
| Texas | | 2012 | Andrews, R. J., Jargowsky, P., & Kuhne, K. (2012). <i>The effects of Texas's targeted pre-kindergarten program on academic performance</i> . Working Paper 84. National Center for Analysis of Longitudinal Data in Education Research (CALDER). https://caldercenter.org/sites/default/files/wp-84.pdf |
| Texas | | 2012 | Huston, A., Gupta, A., & Schexnayder, D. (2012). <i>Study of early education in Texas: The relationship of pre-k attendance to 3rd grade test results</i> . The University of Texas at Austin. https://repositories.lib.utexas.edu/items/1e48b498-bd12-49af-b2d6-0613dae125e9 |
| Texas | | 2016 | Sanborn, R., McConnell, K., Kimball, M., Canales, A., Davila, J., Everitt, S., O'Quinn, K., & et al. (2016). Pre-k in Texas: A critical component for academic success. <i>Journal of Applied Research on Children: Informing Policy for Children at Risk</i> , 7(2), 1-42. https://digitalcommons.library.tmc.edu/cgi/viewcontent.cgi?article=1317&context=childrenatrisk |
| Texas | | 2017 | Texas Education Agency. (2017). <i>Prekindergarten outcomes for Texas public school students</i> . https://researchconnections.org/childcare/resources/36118 |

Texas	Houston	2017	Spikes, S. (2017). <i>Effects of HISD prekindergarten programs on third-grade students' academic achievement, 2015-2016: Evaluation report</i> . Houston Independent School District. https://files.eric.ed.gov/fulltext/ED603136.pdf
Texas	Houston	2017	Baumgartner, E. M. (2017). <i>The benefits of HISD pre-kindergarten: The relationship between years of exposure and school readiness</i> . Houston Education Research Consortium, Rice University Kinder Institute for Urban Research. https://rice.app.box.com/s/82rb3aqicxl4jwrq82q2ok43t1uu3t2f
Texas	San Antonio	2014	Edvance Research. (2014). <i>Pre-K 4 SA evaluation report: Year 1</i> . https://prek4sa.com/wp-content/uploads/2019/01/Pre-K-4-SA_Year-1-Evaluation-Report_Edvance-Research-Inc_Web.pdf
Texas	San Antonio	2015	Edvance Research. (2015). <i>Pre-K 4 SA evaluation report: Year 2</i> . https://prek4sa.com/wp-content/uploads/2019/01/Pre-K-4-SA_Year-2-Evaluation-Report_FINAL.pdf
Texas	San Antonio	2016	Decker-Woodrow, L., & Price, E. (2016). <i>Pre-K 4 SA evaluation report: Year 3</i> . Westat/Edvance Research. https://prek4sa.com/wp-content/uploads/2019/01/Pre-K-4-SA_Year-3-Evaluation-Report.pdf
Texas	San Antonio	2017	Decker-Woodrow, L., Diaz, E., Barfield, D., & Lamey, G. (2017). <i>Pre-K 4 SA evaluation report: Year 4</i> . Westat. https://prek4sa.com/wp-content/uploads/2019/01/Pre-K-4-SA_Year-4-Evaluation-Report.pdf
Texas	San Antonio	2018	Decker-Woodrow, L., Diaz, E., Adachi, E., Barfield, D., & Lamey, G. (2018). <i>Pre-K 4 SA evaluation report: Year 5</i> . Westat. https://prek4sa.com/wp-content/uploads/2019/01/PreK4SA_Year5_EvaluationReport.pdf
Texas	San Antonio	2019	Villarreal, M. (2019). <i>Impact study of prekindergarten for San Antonio</i> . Urban Education Institute at UTSA. https://prek4sa.com/wp-content/uploads/2019/08/UTSA_PK4SA_web_v2.pdf
Texas	San Antonio	2019	Decker-Woodrow, L., Diaz, E., Lamey, G., Hartman, N., Adachi, E., & Barfield, D. (2019). <i>Pre-K 4 SA evaluation report: Year 6</i> . Westat. https://prek4sa.com/wp-content/uploads/2020/10/PreK4SA_Year6_EvaluationReport.pdf

Utah	Salt Lake City	2018	Evaluation and Training Institute. (2018). <i>Utah high-quality school readiness expansion (HQSR-E) program evaluation 2017-2018 findings</i> . https://www.schools.utah.gov/policy/_policy/_utahlegislative-reports/_2018_/2018NovemberHQSR-E.pdf
Utah	Salt Lake City	2019	Evaluation and Training Institute. (2019). <i>Utah high-quality school readiness expansion (HQSR-E) program evaluation 2018-2019 findings</i> . https://www.waterford.org/wp-content/uploads/2020/10/HQSRE-2018-2019-Impact-Final-Report-1-2.pdf
Utah	Salt Lake City	2011	Dubno, J. (2011). <i>A sustainable financing model: High quality preschool for at-risk children</i> . Voices for Utah Children. https://uw.org/wp-content/uploads/2018/02/a-sustainable-financing-model-for-high-quality-preschool-evidence-from-the-granite-school-district-in-utah-updated-10-26-11.pdf
Vermont		2017	Vermont Agency of Education, & Vermont Agency of Human Services. (2017). <i>Preliminary evaluation of pre-kindergarten education programs to promote optimum outcomes for children and to collect data that will inform future decisions: Response to 16 VSA Sec 829(e)(10) [Amended 2015 Act 11 Sec 13. Restart 5-yr clock at 2015]: Report to the General Assembly</i> . https://legislature.vermont.gov/assets/Legislative-Reports/edu-legislative-report-preliminary-evaluation-of-prek-programs-2017.pdf
Vermont		2018	Vermont Agency of Education. (2018). <i>Ready for Kindergarten! Survey (R4K!S) report: 2017 - 2018: Report to supervisory unions/school districts</i> . https://files.eric.ed.gov/fulltext/ED628413.pdf
Vermont		2019	Vermont. Agency of Education. (2019). <i>Ready for Kindergarten! Survey (R4K!S) report: 2018 - 2019: Report to supervisory unions/school districts</i> . https://education.vermont.gov/sites/aoe/files/documents/edu-early-education-ready-for-kindergarten-report-2018-2019_0.pdf

- | | | |
|----------|------|--|
| Vermont | 2019 | Vermont Agency of Education, & Vermont Agency of Human Services. (2019). <i>Annual evaluation of pre-kindergarten education programs to promote optimum outcomes for children and to collect data that will inform future decisions: Response to 16 VSA Sec 829(e)(10) [Amended 2015 Act 11 Sec 13. Restart 5-yr clock at 2015]</i> . Report to the General Assembly.
https://legislature.vermont.gov/assets/Legislative-Reports/edu-final-prek-annual-evaluation-legislative-report-2019.pdf |
| Vermont | 2020 | Vermont Agency of Education. (2020). <i>Ready for Kindergarten! Survey (R4K!S) report: 2019 - 2020: Report to supervisory unions/school districts</i> .
https://education.vermont.gov/sites/aoe/files/documents/edu-early-education-ready-for-kindergarten-report-2019-2020.pdf |
| Virginia | 2012 | Huang, F. L., Invernizzi, M. A., & Drake, E. A. (2012). The differential effects of preschool: Evidence from Virginia. <i>Early Childhood Research Quarterly</i> , 27(1), 33–45.
https://doi.org/10.1016/j.ecresq.2011.03.006 |
| Virginia | 2015 | Almarode, J., Edwards, E. N., Bradburn, I., Downer, J. T., Ruzek, E. A., DeCoster, J., Jonas, D. L., & Gupta, S. (2015). <i>Predicting on-time promotion to and literacy achievement in eighth grade in relation to public prekindergarten in Virginia</i> . Virginia Early Childhood Foundation.
https://hac.virginia.gov/subcommittee/commissions_and_joint_subcommittees/Jt_Preschool_Initiative_Sub/06-11-15/IVb%20-%20VECF%20Smart%20Beginnings%20-%20Predicting%20OnTime%20Promotion%20Study%20Report%20-%20May%202015.pdf |
| Virginia | 2017 | Huang, F. L. (2017). Does attending a state-funded preschool program improve letter name knowledge? <i>Early Childhood Research Quarterly</i> , 38, 116–126.
https://doi.org/10.1016/j.ecresq.2016.08.002 |
| Virginia | 2017 | University of Virginia Center for Advanced Study of Teaching and Learning. (2017). <i>Virginia Preschool Initiative - Plus: CASTL year 2 implementation report</i> . Virginia Department of Education.
https://researchconnections.org/childcare/resources/35783 |

- | | | | |
|------------|---------|------|---|
| Virginia | | 2018 | Haslip, M. (2018). The effects of public pre-kindergarten attendance on first grade literacy achievement: A district study. <i>International Journal of Child Care and Education Policy</i> , 12(1), 1-19. https://doi.org/10.1186/s40723-017-0040-z |
| Virginia | | 2019 | Gaylor, E., Golan, S., Chow, K., Grindal, T., Mercier, B., Williamson, C., & Tiruke, T. (2019). <i>Virginia Preschool Initiative Plus (VPI+): Final evaluation report</i> . SRI International. https://www.sri.com/wp-content/uploads/2021/12/VPI_FinalEvalReport_NoCostStudy_2020_Acc.pdf |
| Washington | | 2014 | Bania, N., Kay, N., Aos, S., & Pennucci, A. (2014). <i>Outcome evaluation of Washington State's early childhood education and assistance program</i> , (Document No. 14-12-2201). Washington State Institute for Public Policy. https://www.wsipp.wa.gov/ReportFile/1576/Wsipp_Outcome-Evaluation-of-Washington-States-Early-Childhood-Education-and-Assistance-Program_Report.pdf |
| Washington | | | Washington Department of Early Learning. (n.d.). <i>ECEAP outcomes: 2011-12 school year</i> . https://researchconnections.org/childcare/resources/27244 |
| Washington | | | Washington Department of Early Learning. (n.d.). <i>2018-2019 ECEAP outcomes report</i> . https://teachingstrategies.com/wp-content/uploads/2021/08/Washington-ECEAP-Outcomes-Report.pdf |
| Washington | Seattle | 2016 | Nores, M., Barnett, W. S., Joseph, G., Stull, S. Figueras-Daniel, A. & Soderberg, J. S. (2016). <i>Year 1 report: Seattle pre-k program evaluation</i> . National Institute for Early Education Research & Childcare Quality & Early Learning Center for Research & Professional Development. https://nieer.org/sites/default/files/2023-08/SPP-Evaluation-Year-1-Report-v10.26.16.pdf |
| Washington | Seattle | 2018 | Nores, M., Barnett, Jung, K., W. S., Joseph, G., Bachman, L., & Soderberg, J. S. (2018). <i>Year 3 report: Seattle pre-k program evaluation</i> . National Institute for Early Education Research & Cultivate Learning. https://nieer.org/sites/default/files/2023-08/Seattle_Preschool_Evaluation-Year-3_SPP10.8.18.pdf |

Washington	Seattle	2019	Nores, M., Barnett, W.S., Jung, K., Joseph, G. & Bachman, L. (2019). <i>Year 4 report: Seattle preschool program evaluation</i> . National Institute for Early Education Research & Cultivate Learning. https://nieer.org/sites/default/files/2023-08/SPP-Evaluation-Year-4-Report-FINAL_v9.30.19-.pdf
Washington		2016	District of Columbia Office of the State Superintendent of Education. (2016). <i>The state of pre-k in the District of Columbia: 2016 pre-k report</i> . https://osse.dc.gov/sites/default/files/dc/sites/osse/publication/attachments/OSSE%20Annual%20Pre-K%20Report%202016%20Feb%202%201017.pdf
Washington	Seattle	2017	Nores, M., Barnett, W.S., Joseph, G., Stull, & Soderberg, J.S. (2017). <i>Year 2 report: Seattle pre-k program evaluation</i> . National Institute for Early Education Research & Cultivate Learning. https://nieer.org/sites/default/files/2023-08/SPP-Evaluation-Year-2-Final-Report.pdf
West Virginia		2005	Lamy, C., Barnett, W., & Jung, K. (2005). <i>The effects of West Virginia's early education program on young children's school readiness</i> . National Institute for Early Education Research. https://researchconnections.org/childcare/resources/7629
West Virginia		2017	Nores, M., Riley-Ayers, S., Kent, A., Jung, K., Allenger, M., & Contreras, C. (2017). <i>Evaluation of West Virginia universal pre-k: First year outcomes</i> . National Institute for Early Education Research.
West Virginia		2019	Nores, M., Jung, K., Contreras, C., Valle, E., & Allenger, M. (2019). <i>Evaluation of West Virginia universal pre-k: Third year longitudinal outcomes</i> . National Institute for Early Education Research.
West Virginia		2019	Nores, M., Jung, K., Riley-Ayers, S., Kent, A., Contreras, C., & Allenger, M. (2019). <i>Evaluation of West Virginia universal pre-k: Second year longitudinal outcomes</i> . National Institute for Early Education Research.
West Virginia		2020	Nores, M., Jung, K., Valle, E. Contreras, C., & Allenger, M. (2020). <i>Evaluation of West Virginia universal pre-k: Fourth year longitudinal outcomes</i> . National Institute for Early Education Research.

- Wisconsin 2016 Artz, B., & Welsch, D. M. (2016). The impact of publicly provided early childhood education programs on district-level test scores. *Contemporary Economic Policy*, 34(1), 89-106. <https://doi-org.ezproxy.cul.columbia.edu/10.1111/coep.12128>
-