ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 23-885

Why are Preschool Programs Becoming Less Effective?

Anamarie A. Whitaker
University of Delaware

Margaret Burchinal
University of Virginia

Jade M. Jenkins
University of California, Irvine

Drew H. Bailey
University of California, Irvine

Tyler W. Watts
Teachers College, Columbia University

Greg J. Duncan
University of California, Irvine

Emma R. Hart
Teachers College, Columbia University

Ellen Peisner-Feinberg
University of North Carolina

High-quality preschool programs are heralded as effective policy solutions to promote low-income children's development and life-long wellbeing. Yet evaluations of recent preschool programs produce puzzling findings, including negative impacts, and divergent, weaker results than demonstration programs implemented in the 1960s and 70s. We provide potential explanations for why modern preschool programs have become less effective, focusing on changes in instructional practices and counterfactual conditions. We also address popular theories that likely do not explain weakening program effectiveness, such as lower preschool quality and low-quality subsequent environments. The field must take seriously the smaller positive, null, and negative impacts from modern programs and strive to understand why effects differ and how to improve program effectiveness through rigorous, longitudinal research.

VERSION: December 2023

Suggested citation: Whitaker, Anamarie A., Margaret Burchinal, Jade M. Jenkins, Drew H. Bailey, Tyler W. Watts, Greg J. Duncan, Emma R. Hart, and Ellen S. Peisner-Feinberg. (2023). Why are Preschool Programs Becoming Less Effective?. (EdWorkingPaper: 23-885). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/smqa-n695

Why are Preschool Programs Becoming Less Effective?

Anamarie A. Whitaker
University of Delaware

Margaret Burchinal
University of Virginia

Jade M. Jenkins
Drew H. Bailey
University of California, Irvine

Tyler W. Watts
Teachers College, Columbia University

Greg J. Duncan
University of California, Irvine

Emma R. Hart
Teachers College, Columbia University

Ellen Peisner-Feinberg
University of North Carolina

Author Note

Anamarie A. Whitaker https://orcid.org/0000-0002-6865-5850

Margaret Burchinal https://orcid.org/0000-0002-3606-7843

Jade M. Jenkins https://orcid.org/0000-0002-2000-3087

Tyler W. Watts https://orcid.org/0000-0002-2741-0873

Greg J. Duncan https://orcid.org/0000-0002-9869-6311

Emma R. Hart https://orcid.org/0000-0003-3808-0838

Ellen Peisner-Feinberg https://orcid.org/0000-0001-9716-9977

We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Anamarie A. Whitaker, 212 Alison Hall, Newark, DE 19716. Email: awhit@udel.edu

Abstract

High-quality preschool programs are heralded as effective policy solutions to promote low-income children's development and life-long wellbeing. Yet evaluations of recent preschool programs produce puzzling findings, including negative impacts, and divergent, weaker results than demonstration programs implemented in the 1960s and 70s. We provide potential explanations for why modern preschool programs have become less effective, focusing on changes in instructional practices and counterfactual conditions. We also address popular theories that likely do not explain weakening program effectiveness, such as lower preschool quality and low-quality subsequent environments. The field must take seriously the smaller positive, null, and negative impacts from modern programs and strive to understand why effects differ and how to improve program effectiveness through rigorous, longitudinal research.

Why are preschool programs becoming less effective?

Evidence from small-scale randomized control trials (RCTs) conducted decades ago demonstrated that preschool programs can support young children from low-income families in developing the skills needed to succeed in school and adult life (Elango et al., 2016). These studies, plus quasi-experimental evaluations showing long-run benefits to early cohorts of the Head Start program (e.g., Bailey et al., 2021), have convinced many people that preschool could be one of the most effective policy levers for equalizing opportunity in the United States and have often been used to justify large investments by federal, state, and local governments in preschool programs for low-income children (White House, 2023). At the same time, rigorous evaluations of more recent preschool programs have not replicated the results of the early evaluations. Although some have found positive impacts, others have shown no impacts, and one has even shown significantly lower school achievement and worse behavior among children attending state-funded preschool programs when compared with children who did not. As public spending on preschool increases, it is important to understand why some programs seem to work much better than others.

High-Intensity Preschool Programs

RCT evaluations of the Perry Preschool and Abecedarian programs demonstrate that preschool can provide lifetime benefits for disadvantaged children (Elango et al., 2016). However, both programs were much more expensive than today's preschool programs, more comprehensive in terms of services offered, and were run by researchers in a small number of classrooms. Perry (n = 123) provided half-day center care and home visiting to 3- and 4-year-olds from low-income Black families in the 1960s in Ypsilanti, Michigan at a cost of approximately $43,500[1] per child. Abecedarian (n = 111) provided full-time preschool to low-income, predominantly Black children from infancy through entry into kindergarten during the

1970s in Chapel Hill, North Carolina, costing approximately $120,000[1] per child. However, current high-quality, full-day pre-k program costs typically are in the range of approximately $15,000-16,500 (in 2023$)—substantially less than the demonstration programs (Karoly et al., 2021).

Both of these early programs focused on encouraging positive, affirming adult-child interactions, hands-on learning activities, and frequent conversations between children and teachers (Ramey et al., 2014; Schweinhart & Weikart, 1980). The teachers, their supervisors, and researchers collaborated in developing the interventions and classroom activities that were later incorporated into early childhood curricula. However, the evaluations were not designed to disentangle the separate contributions of program components for improving children's development.

Long-run evaluations of both programs showed higher levels of adult education, employment, earnings, and health, and lower levels of crime and incarceration in the treatment group, with estimated program-generated benefits totaling six to ten times the program costs (García et al., 2021). In short, results from rigorous evaluations of these two programs provide convincing evidence that preschool programs can produce lasting benefits for children by improving educational attainment, income, and health in adulthood.

Publicly Funded Preschool Programs

Head Start. Head Start, a federally funded and regulated preschool program, began in 1965 as a part of President Johnson's War on Poverty. Currently, the program serves nearly 900,000 children and their low-income families at a cost to the federal government of approximately $10 billion annually (U.S. Department of Health and Human Services [DHHS],

[1] Cost figures for Perry Preschool and Abecedarian are taken from Elango et al. (2016), specifically Table 4.8. Figures were converted in 2023 dollars.

2022). The majority of Head Start programs are based in centers, serve 3- and 4-year old children and use a "whole-child" approach that initially focused on enhancing nutrition, health, and social skills, and added a focus on teaching reading and math skills starting in the mid 2000s (U.S. DHHS, 2007).

Quasi-experimental research on Head Start has reported positive long-run impacts in cohorts who attended the program prior to 1990 (Bailey et al., 2021; Deming, 2009). In contrast, an extension of the Deming (2009) study using more recent cohorts yielded null or negative impacts in early adulthood (Pages et al., 2020). A large RCT evaluation of the program (the Head Start Impact Study [HSIS]), conducted in a nationally representative set of Head Start sites in 2002, yielded some positive impacts at the end of the Head Start year, but these differences disappeared rapidly after children entered elementary school (Puma et al., 2012). By third grade, few statistically significant results were detected, and some of those indicated slightly more emotional problems and problems with peer relations among the 4-year-old Head Start than control children (Puma et al., 2012). No long-term follow-up of the HSIS has yet been conducted.

State and Local Pre-k. Pre-k programs, with a primary focus on promoting academic skills, run by states and some cities, began to expand in the 1990s. As of 2023, state-funded pre-k programs serve approximately 1.5 million preschoolers, most of whom are from low-income families, at a cost of almost $10 billion (Friedman-Krauss et al., 2023). Although the effectiveness of these programs has been examined in various ways, only four program evaluations with publicly reported impacts on child outcomes have been based on admission lotteries or explicit random assignment (Boston, Tennessee, North Carolina, and Georgia). As we note below, the results of these evaluations have generated vexing questions for the ECE field.

Boston Pre-k. The most positive results come a lottery study of children enrolled in Boston's pre-k program between 1997 and 2003 (Gray-Lobe et al., 2023), prior to quality enhancements that characterize the current program (Boston Public Schools, n.d.). Although no information was available to gauge program impacts at the end of their pre-k year, school administrative records comparing children who won or lost the admissions lottery showed no differences in achievement test scores during elementary school, but statistically significant differences in high school graduation, SAT scores, disciplinary problems, and college attendance. A second, shorter-run lottery-based evaluation of Boston pre-k using more recent cohorts of attendees also demonstrates null results on standardized achievement tests in third grade but has not followed participants beyond that point (Weiland et al., 2020). This pattern of results is perplexing, as the best indicators of proximate program impacts (i.e., elementary school achievement scores) produced null effects in both evaluations. Thus, the positive long-term impacts reported by Gray-Lobe et al. on adult outcomes lack a clear mechanism linking pre-k attendance to adult functioning.

Tennessee's Voluntary Pre-k Program (TNVPK). The TNVPK program was the first statewide pre-k program to be evaluated with an RCT and longitudinal follow-up. At the end of the pre-k year, TNVPK children had significantly higher academic skills than children who lost the admissions lottery (Lipsey et al., 2018). However, in third to sixth grade, TNVPK children had lower academic skills, increased rates of disability identification, greater absences, and more disciplinary experiences than students who applied but were not offered a pre-k slot (Durkin et al., 2022). Specifically, TNVPK children were more likely to have a disciplinary infraction than those who were not offered a slot (effect size of .17). Their math and reading scores were lower as well, with effect sizes ranging from -0.24 to -0.33 on standardized assessments of sixth-grade achievement using treatment-on-the-treated adjustments (i.e., adjusting for actual pre-k

attendance). The negative long-term results reported in TN also present a perplexing pattern of results for the field, as the program had positive impacts on child achievement at the end of pre-k. The quality of the TNVPK program has been debated – an argument we turn to below.

North Carolina Pre-k (NCPK). In 2017, NCPK programs with waiting lists in two large North Carolina counties randomly assigned applicants to attend NCPK (Peisner-Feinberg et al., 2020). As with the TNVPK program, initial evaluation results (not yet peer reviewed) indicated the NCPK children ended pre-k with higher levels of language and early reading skills than the control children. By the end of kindergarten, however, the NCPK children showed lower math skills (d = -.18) and more behavior problems (d = .23) albeit, not significantly so after multiple hypothesis testing corrections (Peisner-Feinberg et al., 2020).

Georgia Pre-K Program. A final lottery study, also not yet peer reviewed, is of children seeking to enter Georgia's pre-k program between 2012 and 2018 (Woodyard et al., 2023). It found big gains in math and reading scores at the end of the program that quickly faded out and, by fourth grade, were negative but not always statistically significant. Adverse impacts appeared to be concentrated among the more economically-advantaged students.

Non-RCT Evidence. A plethora of quasi-experimental evidence from state pre-k evaluations, relying on correlational methods, report moderate to large initial positive associations that decline and sometimes, but not always, disappear during the first years of elementary school (Phillips et al., 2017). Follow-up studies of the children who attended the Tulsa pre-k programs and NCPK programs showed initial impacts and some longer-term impacts (Dodge et al., 2017; Gormley et al., 2023). Quasi-experimental research in North Carolina has found that increased funding for the program has positive impacts on later measures of achievement, especially for severely disadvantaged children (Watts et al., 2023).

Comparing the Impacts of Early and More Recent Programs. These mixed findings from evaluations of scaled-up programs raise important questions about what medium- to long-term impacts should be expected from today's preschool programs. A systematic look at preschool RCT impacts across decades suggests that, even if the negative results from Tennessee, North Carolina, and Georgia are set aside, most programs are producing smaller impacts than what was found in the foundational literature for preschool (Duncan & Magnuson, 2013).

The most recent evidence on the temporal impacts of preschool programs comes from the Meta-Analysis of Educational RCTs with Follow-up (MERF), a dataset comprised of post-test and follow-up impacts on cognitive and social-emotional outcomes from a broad range of educational RCTs systematically sampled from eight existing meta-analyses (for more details see Hart et al., 2023). As depicted in Figure 1, a meta-analytic evaluation of 17 preschool treatment-control group contrasts suggests that end-of-treatment impacts for programs that started in 1960-1999 (.47 *SD*) are more than twice as large as those from ECE programs that started in 2000 to 2011 (.19 *SD*). Moreover, even conditional on posttest effect sizes, the impacts of the more recent programs appear to fade out more quickly than the impacts for the older evaluations. These patterns were apparent across a broad range of ECE treatment contrasts that were both exposure-based (i.e., control group receives no preschool offer) and curricular (i.e., control group receives business-as-usual preschool curriculum).

Possible Explanations for Declining Preschool Impacts

Societal changes in the last 50-60 years now provide all children with some of the advantages that the early preschool projects provided in the 1960s and 70s. We argue that these contrasting advantages have become narrower over time, both because the environmental

conditions experienced by children have improved sharply over the past several decades and because the foci of preschool programs have become narrower and more academic.

Improved Counterfactual Conditions. The early-childhood enrichment opportunities for children have improved in two ways—increasing support for and quality of children's home environments and the expansion of community-based preschool. We discuss each domain below.

Secular Trends in Children's Environments. In the 1960s and 1970s, children from low-income families, especially Black children, faced deplorable conditions. In the mid-1960s, the Food Stamp program and programs such as the Earned Income Tax Credit had not yet been introduced. Racial discrimination in parts of the country denied Black children access to quality schools and Black families access to hospital care, including physician-assisted childbirth (Reynolds, 2004). Many safety-net programs for children in low-income families were introduced, and others were markedly expanded between 1960 and 2000. As shown in Table 1, spending on welfare programs and ECE services has risen dramatically, with a large increase seen in Medicaid and SNAP funding from 2000 to 2019, as well as substantial increases in Head Start and Child Care and Development Fund dollars.

Home environments have improved as well. Mothers with low income completed 2 more years of schooling in 2000 compared with 1960, and family size shrunk by over 30%. Additionally, parents spend more time with their children now: in the 1990s, parents increased the amount of time they spent with their children dramatically, by nearly a third for non-college-educated mothers and nearly 100% for non-college-educated fathers (Ramey & Ramey, 2009). Such broad environmental improvements have likely contributed to improved child outcomes. Child health has improved in the U.S., with infant and child mortality falling by over 70% between 1960 and 2000 (See Table 1). Potentially offsetting some of the changes to family size, maternal education, public expenditures on children's health, nutrition, and education, child

health, and ECE participation are increases in crime and single motherhood over the same time period.

Even assuming equal or increasing program quality across time, these factors plausibly made it much easier for initial Head Start programs or a model program like Perry Preschool, to demonstrate effectiveness for enrolled children when compared with children experiencing business-as-usual conditions. Consistent with this explanation, the positive long-run effects of Head Start, even in earlier cohorts, were smaller when children had more access to Food Stamps and when they lived in counties with lower poverty levels (Bailey et al., 2021). This explanation likely accounts, at least in part, for the fact that end-of-treatment impacts are substantially smaller in more recent evaluations (Duncan & Magnuson, 2013).

In addition, changes in public education could account for why impacts of current programs may diminish more rapidly than for earlier programs. Kindergarten instruction changed from focusing on socializing young children to school to teaching early reading and math skills at the same time that preschool programs were also increasing their instruction in early literacy and math skills (Bassok et al., 2016). This resulted in redundancy in instruction (Cohen-Vogel et al., 2021) and almost certainly contributes to fadeout because kindergarten teachers promote the same early academic skills that have increasingly become the focus of preschool classrooms.

Greater Availability of Preschool Services. Evidence shows that Head Start impacts on language were larger and more likely to be sustained in the short run when the children who attended Head Start would otherwise have stayed home with their parents or been cared for in a home-based preschool setting (Kline & Walters, 2016). This suggests that smaller impacts observed in more recent evaluations might be in part due to much higher rates of center-based preschool attendance among the population of children ages five and under (Cascio, 2021; see Table 1). However, a large portion (49%) of children assigned to the control condition in the

TNVPK evaluation also stayed home (Lipsey et al., 2013). Further, over half of the Abecedarian control group attended center care (Garcia et al., 2020). Thus, the increased use of center-based care may help explain smaller initial impacts but does not straightforwardly account for the faster fadeout of impacts and certainly not the significant negative impacts found in elementary school in the TNVPK and Georgia programs.

Changes in Instructional Modes and Focus Over Time. Abecedarian and Perry focused on strong caregiver-child relationships with frequent multi-turn conversations, and hands-on learning activities in which teachers scaffolded learning (Ramey et al., 2014; Weikart, 2008). Head Start initially focused on promoting health development (Zigler & Styfco, 2010). Around 2000, preschool programs began to emphasize teaching early literacy and math skills to address kindergarten-entry gaps between low- and middle-income children (U.S. DHHS, 2007). At the same time, instruction shifted from teaching primarily through hands-on learning activities to teacher-managed instruction. For example, a comparison of nationally representative Head Start program data from 2001 to 2015 shows children spending less time in hands-on learning opportunities and more time in teacher-led large group instruction focused on basic academic skills (Markowitz & Ansari, 2020). Similarly, data from different pre-k evaluations offer some evidence that time spent in literacy and math instruction, along with more developmentally inappropriate whole-group instruction, has increased. Most of this academic instruction occurred in large groups and was didactic teacher-led instruction (Burchinal et al., 2021).

This change from using hands-on learning and focusing on promoting health, language, and social skills to using didactic large group instruction to teach early reading and math skills could account for changes in the medium-term findings across time. Some have argued that preschool instruction in language, executive functioning, and social skills are foundational skills

and thus more likely to be maintained over time than instruction in rote skills like basic reading and math skills (McCormick et al., 2021), while others argue that the success of early intervention programs depends on teaching skills that are not easily acquired later (Bailey et al., 2017). In early programs like Abecedarian that did not have an explicit focus on teaching early academic skills, children entered kindergarten with large treatment impacts on cognitive skills and smaller impacts on reading and math skills. Impacts on both reading and math skills persisted or grew in second grade and were maintained through 14-21 years of age (Campbell et al., 2012; García et al., 2020). In contrast, the strongest causal evaluations of contemporary programs that focus on teaching academic skills reported K-12 academic skills that are either negative or null (Durkin et al., 2022; Gray-Lobe et al., 2023; Peisner-Feinberg et al., 2020; Puma et al., 2012).

This emerging focus on teaching early academic skills in preschool and how they are often taught is likely difficult for preschoolers. These skills are often taught in whole-group, teacher-led instruction, with preschoolers expected to sit still for relatively long periods (Burchinal et al., 2021). Sitting still in large group instruction can result in behavioral dysregulation for young children, which can trigger impatience and harshness in teachers and, in turn, further exacerbate children's problem behaviors (Christopher & Farran, 2020). Current programs' focus on basic skills may optimize for impacts in the pre-k year (Phillips et al., 2017), but because these same skills are often taught again in kindergarten (Cohen-Vogel et al., 2020), children in the control group are likely to catch up substantially during the subsequent year. The combination of redundant instruction and harsh interactions with teachers may cause children to disengage from learning, perhaps setting them on less positive academic trajectories during the early school years. The pattern of findings observed in Tennessee and Georgia with test score impacts flipping from mostly positive to mostly negative, is consistent with the possibility that

positive impacts on learning basic skills can be offset by negative impacts on other skills, such as behavioral or emotional outcomes.

Other Widely Discussed Fadeout Explanations

Scaling Up Research Projects. In hindsight, it is naïve to believe that findings from the early RCT studies would generalize to public programs operating at scale that serve millions of children each year at a fraction of the cost. Operating an at-scale program almost always involves lower costs for serving each child to make the program financially feasible. Launching at-scale programs also entails getting buy-in from the many stakeholders and almost always requires changing the program to meet their goals and needs (Tseng et al., 2017).

Changes in Program Quality Over Time. Many critics cite lower overall program quality to explain disappointing longer-term impacts, but the evidence for this is difficult to parse. As described above, long-term positive impacts are reported for the presumably high-quality Perry and Abecedarian projects and for the first cohorts of children who attended Head Start. Yet the quality of Head Start's programming today is substantially higher than in the early years of the program because the performance quality standards became more rigorous over time (U.S. DHHS, 2023). Furthermore, the quality assessments of the two RCT evaluations with negative impacts ranged from average in Tennessee (Lipsey et al., 2018) to above average in North Carolina (Peisner-Feinberg et al., 2020) compared with other preschool programs. Thus, positive impacts for the first cohorts of Head Start and not more recent cohorts of Head Start (Pages et al., 2020) or today's preschool programs (Phillips et al., 2017) and average to high quality in the two pre-k programs with RCT evaluations does not suggest that program quality accounts for fadeout. Further, TNVPK, NCPK and Georgia pre-k demonstrated positive end of pre-k year results that then turned negative (nonsignificant in NCPK) in early elementary school, suggesting

quality was sufficiently high enough to boost initial skills on average compared with children in other care settings (Lipsey et al., 2018; Peisner-Feinberg et al., 2020; Woodyard et al., 2023).

Quality of Subsequent Schooling. A widely accepted explanation is that the quality of schools that children from programs serving low-income children attend are too low to build on the skills children acquire in preschool (Abenavoli, 2019). However, a comprehensive meta-analysis of studies testing the sustaining environment hypothesis found little evidence to support this explanation (Bailey et al., 2020), although one compelling analysis of the effects of Head Start attendance has found larger long-run impacts for children who attended elementary schools that received additional funding (Johnson & Jackson, 2019). Instead, recent evidence suggests environments that promote more learning in the post-intervention period may only further exacerbate fadeout by accelerating the eventual catch-up of children without preschool experience (Watts et al., 2023).

Conclusion

In summary, many factors likely explain why today's preschool programs are not having the same long-term impacts as earlier classic programs. Changes in scale, who administers the programs, increased opportunities and safety net coverage for low-income minoritized families, and increased access to preschool center care likely account for why today's programs have smaller impacts at entry to kindergarten. Those arguments, however, cannot account for impacts that fade dramatically in most longer-term evaluations and, especially, why initially positive impacts turned negative in direction in three randomized evaluations of programs meeting quality benchmarks relied upon by preschool researchers (Friedman-Krauss et al., 2023). Further worrisome is that these high-quality programs are often the ones supported by public funds serving children from low-income families with the goal of promoting equity.

Perhaps a central lesson learned is that it is unreasonable to expect similar results from demonstration programs funded at high per child levels given current preschool funding levels are substantially lower. This alone may require researchers, policymakers, and practitioners to adjust expectations on what today's programs can produce in terms of promoting children's development.

We make two recommendations regarding future research to further understand these fadeout issues. First, additional long-term follow-up studies are needed from RCTs that can connect short-term and long-term effects. The long-term follow-up studies of children who attended pre-k over 20 years provided surprising evidence of young adult impacts, even in the absence of short or intermediate impacts (Gray-Lobe et al., 2023). Careful follow-up studies of rigorous RCTs like the HSIS or the pre-k lottery studies are needed to determine whether they too will show these young adult impacts despite the fading of the initial short-term impacts. Such evaluations could allow for systematic investigation of the extent to which long-term impacts are forecasted by short-term impacts, with major implications for developmental theory and intervention design.

Second, we argue that research should examine whether changes in environmental conditions and in program focus and style of instruction explain, at least in part, why today's programs' impacts are smaller, null, or negative, and fade faster when children enter elementary school. The almost complete fadeout of today's preschool programs in RCT studies suggests that focusing on early literacy and, perhaps, numeracy skills in preschool programs, at best, is ineffective and may also result in negative outcomes in the medium-term for children (Durkin et al., 2022). Additionally, future research can examine whether different patterns of medium- and longer-term impacts could be obtained with the instructional practices of the earlier programs– which are similar to those recommended in a recent What Works Clearinghouse guide

(Burchinal et al., 2022) on preparing young children for school (i.e., language-rich, hands-on learning activities, and frequent conversations between teachers and students). We believe the field should take seriously the null and negative results arising from recent preschool program evaluations, and, rather than dismiss the findings using explanations that do not hold when thoroughly examined, strive to understand how best to make preschool programs effective in the current conditions.

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.

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.

Bailey, D. H., Jenkins, J. M., & Alvarez-Vargas, D. (2020). Complementarities between early educational intervention and later educational quality? A systematic review of the sustaining environments hypothesis. *Developmental Review, 56*, 100910.

Bailey, M. J., Sun, S., & Timpe, B. (2021). Prep School for poor kids: The long-run impacts of Head Start on Human capital and economic self-sufficiency. *American Economic Review, 111*(12), 3963-4001.

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? *AERA Open, 2*(1), 2332858415616358.

Boston Public Schools (n.d). What Does Research Say? Retrieved from: https://www.bpsearlylearning.org/results.

Burchinal, M. R., Krowka, S., Newman-Gonchar, R., Jayanthi, M., Gersten, R., Wavell, S., ... & Rosen, E. (2022). Preparing Young Children for School. Educator's Practice Guide. WWC 2022009. *What Works Clearinghouse*.

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

Cascio, E. U. (2021). *Early Childhood Education in the United States: What, When, Where, Who, How, and Why*. (No. w28722). National Bureau of Economic Research. Retrieved from: https://www.nber.org/papers/w28722

Christopher, C., & Farran, D. (2020). Academic gains in kindergarten related to eight classroom practices. *Early Childhood Research Quarterly, 53*, 638-649.

Cohen-Vogel, L., Little, M., Jang, W., Burchinal, M., & Bratsch-Hines, M. (2021). A missed opportunity? Instructional content redundancy in Pre-K and kindergarten. *AERA Open, 7*, 23328584211006163.

Cohen-Vogel, L., Sadler, J. R., Little, M., & Merrill, B. (2020). (Mis) Alignment of instructional policy supports in Pre-K and kindergarten: Evidence from rural districts in North Carolina. *Early Childhood Research Quarterly, 52*, 30-43.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics, 1*(3), 111-134.

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

Duncan, G. J., & Magnuson, K. (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.

Elango, S., García, J. L., Heckman, J. J., & Hojman, A. (2016). Early childhood education. In *Economics of means-tested transfer programs in the United States, volume 2* (pp. 235-297). University of Chicago Press.

Friedman-Krauss, A., Barnett, W. S., Hodges, K. S., Garver, K., Weisenfeld, G. G., Gardiner, B. A., & Jost, T. M. (2023). *The State of Preschool 2022*. https://nieer.org/wp-content/uploads/2023/05/YB2022_FullReport.pdf

García, J. L., Bennhoff, F. H., Leaf, D. E., & Heckman, J. J. (2021). *The dynastic benefits of early childhood education*. (No. w29004). National Bureau of Economic Research. Retrieved from: https://www.nber.org/papers/w29004

García, J. L., Heckman, J. J., Leaf, D. E., & Prados, M. J. (2020). Quantifying the life-cycle benefits of an influential early-childhood program. *Journal of Political Economy, 128*(7), 2502-2541.

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.

Hart, E. R., Bailey, D. H., Luo, S., & Watts, T. W. (2023). *Do intervention impacts on social-emotional skills persist at higher rates than impacts on cognitive skills? A meta-analysis of educational RCTs with follow-up* (EdWorkingPaper: 23-782). https://doi.org/10.26300/7j8s-dy98

Johnson, R. C., & Jackson, C. K. (2019). Reducing inequality through dynamic complementarity: Evidence from Head Start and public school spending. *American Economic Journal: Economic Policy, 11*(4), 310-349.

Karoly, L. A., Cannon, J. S., Gomez, C. J., & Whitaker, A. A. (2021). Understanding the Cost to Deliver High-Quality Publicly Funded Pre-Kindergarten Programs. *RAND Corporation* https://www.rand.org/pubs/research_reports/RRA252-1.html

Kline, P., & Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of Head Start. *The Quarterly Journal of Economics, 131*(4), 1795-1848.

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.

Lipsey, M. W., Hofer, K. G., Dong, N., Farran, D. C., & Bilbrey, C. (2013). Evaluation of the Tennessee Voluntary Prekindergarten Program: End of pre-k results from the randomized control design. *Nashville, TN: Vanderbilt University, Peabody Research Institute*.

Markowitz, A. J., & Ansari, A. (2020). Changes in academic instructional experiences in Head Start classrooms from 2001–2015. *Early Childhood Research Quarterly, 53*, 534-550.

McCormick, M., Weiland, C., Hsueh, J., Pralica, M., Weissman, A. K., Moffett, L., Snow, C., & 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.

Pages, R., Lukes, D. J., Bailey, D. H., & Duncan, G. J. (2020). Elusive longer-run impacts of head start: replications within and across cohorts. *Educational Evaluation and Policy Analysis, 42*(4), 471-492.

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, School of Education.*

Phillips, D., Lipsey, M., Dodge, K., Haskins, R., Bassok, D., Burchinal, M., & Weiland, C. (2017). Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects*. The current state of scientific knowledge on pre-kindergarten effects*, 19-30. Brookings Institution.

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Mashburn, A., & Downer, J. (2012). Third Grade Follow-Up to the Head Start Impact Study: Final Report. OPRE Report 2012-45. *Administration for Children & Families*.

Ramey, C. T., Sparling, J. J., & Ramey, S. L. (2014). Interventions for students from low resource environments: The Abecedarian approach. *Essentials of planning, selecting, and tailoring interventions for unique learners*, 415-448.

Ramey, G., & Ramey, V. A. (2009). *The rug rat race*. National Bureau of Economic Research. Retrieved from https://www.nber.org/papers/w15284.

Reynolds, P. P. (2004). Professional and hospital discrimination and the US Court of Appeals Fourth Circuit 1956–1967. *American Journal of Public Health, 94*(5), 710-720.

Schweinhart, L. J., & Weikart, D. P. (1980). *Young children grow up: The effects of the Perry Preschool Program on youths through age 15*. ERIC.

Tseng, V., Easton, J. Q., & Supplee, L. H. (2017). Practice Partnerships: Building Two-Way Streets of Engagement. Social Policy Report. Volume 30, Number 4. *Society for Research in Child Development*.

U. S. Department of Health and Human Services (2007). Head Start Act. Retrieved from: https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/HS_Act_2007.pdf

U. S. Department of Health and Human Services (2022). Head Start Program Facts: Fiscal Yeare 2022. Retrieved from: https://eclkc.ohs.acf.hhs.gov/data-ongoing-monitoring/article/head-start-program-facts-fiscal-year-2022

U. S. Department of Health and Human Services (2023). History of Head Start. Retrieved from: https://www.acf.hhs.gov/ohs/about/history-head-start.

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.

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.

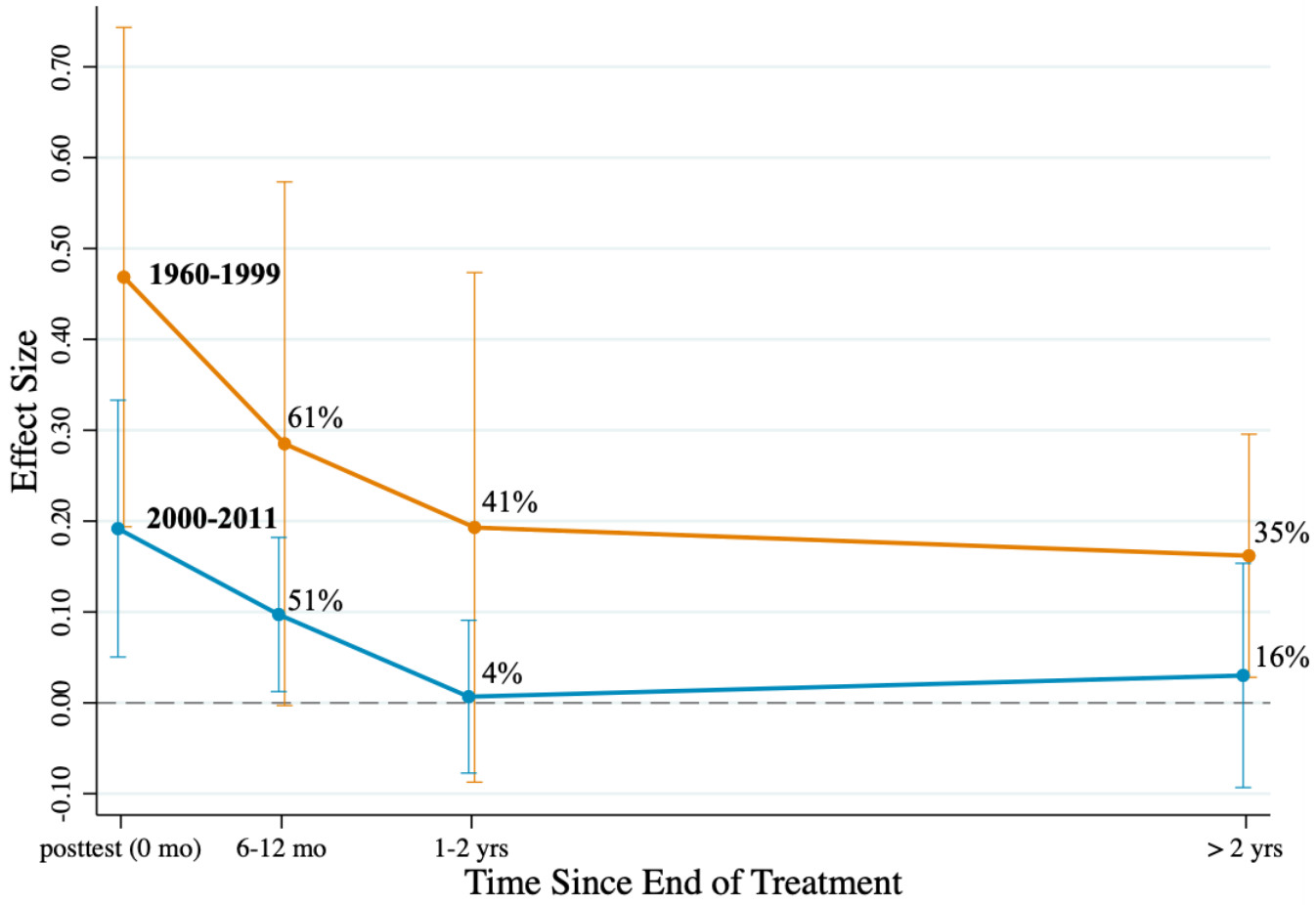
White House. (2023). *Economic Report of the President 2023.* https://www.whitehouse.gov/wp-content/uploads/2023/03/ERP-2023.pdf

Woodyard, H. T., Sass, T. R., & Fazlul, I. (2023). *Assessing the Benefits of Education in Early Childhood: Evidence from a Pre-K Lottery in Georgia.* (EdWorkingPaper: 23-880). Annenberg Institute at Brown University.

Zigler, E., & Styfco, S. J. (2010). *The hidden history of Head Start*. Oxford University Press.

Figure 1.

Average Treatment Impacts for Early Childhood Educational Interventions



Note: Figure displays the average treatment impacts across end of treatment and follow-up assessments for language/literacy, math, general cognitive, and social-emotional outcomes for early childhood educational interventions. Averages were estimated using meta-regression with weights (1/se^2) and a random effect for study. Orange lines and coordinates depict trajectories for interventions that began between 1960-1999. Blue lines and coordinates depict trajectories for interventions that began between 2000-2011. The x-axis ticks are scaled by the average time elapsed since treatment end for each bin (e.g., the average time since post-test for a greater than 2-year follow-up assessment was 76 months. The x-axis is scaled accordingly). Percentages reflect the percentage of the average post-test effect observed at each follow-up wave. Meta-analytic averages come from MERF, the Meta-Analysis of Educational RCTs with Follow-up (see Hart et al., 2023 for more details).

Table 1.

Conditions in the United States, Expenditures, Child Health, Crime, and ECE Participation Across Time

	1960		2000		2019	
	Bottom Income Quintile	**National Average**	**Bottom Income Quintile**	**National Average**	**Bottom Income Quintile**	**National Average**
Demographics for families with children <5						
Number of Children in Home	3.8	3.8	2.8	3.2	2.5	2.8
Maternal Labor Force Participation	21.9%	20.6%	57.4%	66.8%	55.2%	64.6%
Maternal Education Level (in years)	9.1	10.7	11.3	12.6	11.3	12.8
Single Mothers	3.4%	0.8%	34.4%	14.4%	37.4%	14.4%
Federal Expenditures (in billions, 2023$)	**1960**		**2000**		**2019**	
Medicaid	0		40.6		117	
AFDC/TANF	6		19.8		15.1	
SNAP	0		16.7		32.8	
WIC	0		6		5.6	
Head Start	0		7.6		11.2	
Child Care Development Fund	0		5.7		8.6	
Per Capita Expenditures (2023$)						
Total Federal Child Expenditures (per capita)	362		3,407		6,230	
Per Pupil Public K-12 Spending	4,535		14,211		17,910	
Child Health	**1960**		**2000**		**2019**	
Infant (< 1 year) Mortality (per 100k)	2,700		728.7		558.3	
Child (1- 4 years) Mortality (per 100k)	110		32.9		23.3	
Crime	**1960**		**2000**		**2019**	
Violent Crime (per 100k)	160.9		506.5		366.7	
Property Crime (per 100k)	1,726.3		3,618.3		2,109.9	
ECE Participation (% Enrolled)	**1970**		**2000**		**2019**	
Age 3 Child Care or Pre-K	12.9%		39.2%		53.75%[a]	
Age 4 Child Care or Pre-K	27.8%		64.9%			
Age 5 Child Care, Pre-K, or Kindergarten	69.3%		87.6%		90.7%	

Note. Demographic information comes from authors' calculations of the harmonized Decennial Census and American Community Surveys made available by the Integrated Multiuse Public Microdata Series (IPUMS) at the University of Minnesota (Ruggles et al., 2023).

Expenditure figures are in billions of dollars and are in 2023 dollars. Total child expenditure is per child capita. Federal expenditure and federal child per capita expenditure information comes from https://www.urban.org/sites/default/files/publication/102614/kids-share-2020-chartbook_0.pdf. Per pupil K-12 spending information comes from https://nces.ed.gov/programs/digest/d22/tables/dt22_236.55.asp. Specifically, we use data from expenditure per pupil in fall enrollment, total expenditures. 1959-60; 1999-00; 2018-19 (National Center for Education Statistics [NCES], 2023a).

Infant mortality rate is number of infant deaths before age 1 per 100,000. Child mortality rate is number of child deaths (ages 1-4) per 100,000. Data from 1960 comes from: https://www.cdc.gov/nchs/data/vsus/VSUS_1960_2A.pdf; Data from 2000 comes from: https://www.cdc.gov/nchs/data/nvsr/nvsr50/nvsr50_15.pdf; Data from 2019 comes from: https://www.cdc.gov/nchs/data/databriefs/db395-H.pdf.

Crime data comes from the U.S. Census Bureau for 1960 and 2000 (https://www2.census.gov/library/publications/2004/compendia/statab/123ed/hist/hs-23.pdf). The 2019 data comes from Federal Bureau of Investigation (https://www.fbi.gov/news/press-releases/fbi-releases-2019-crime-statistics#:~:text=The%202019%20statistics%20show%20the,2%2C109.9%20offenses%20per%20100%2C000%20inhabitants.). Violent crime includes murder, manslaughter, rape, robbery, and aggravated assault. Property crimes include burglary, larceny-theft, and motor vehicle theft.

Early childhood education participation information comes from NCES (NCES, 2023b): https://nces.ed.gov/programs/digest/d19/tables/dt19_202.10.asp, and https://nces.ed.gov/programs/digest/d22/tables/dt22_202.20.asp. Specifically, we used percent of enrolled 3, 4, and 5 year olds in school.
a = In 2019, the number of 3 and 4 year olds enrolled in school was not available individually for each age, thus the combined percentage of 3 and 4 year olds enrolled in school is displayed.