

A Report on “The Long-Term Effects of
Universal Preschool in Boston” by
Gray-Lobe et al. (2023)

Reviewer 2

February 25, 2026

v1



isitcredible.com

Disclaimer

This report was generated by large language models, overseen by a human editor. It represents the honest opinion of The Catalogue of Errors Ltd, but its accuracy should be verified by a qualified expert. Comments can be made [here](#). Any errors in the report will be corrected in future revisions.

I am wiser than this person; for it is likely that neither of us knows anything fine and good, but he thinks he knows something when he does not know it, whereas I, just as I do not know, do not think I know, either. I seem, then, to be wiser than him in this small way, at least: that what I do not know, I do not think I know, either.

Plato, *The Apology of Socrates*, 21d

To err is human. All human knowledge is fallible and therefore uncertain. It follows that we must distinguish sharply between truth and certainty. That to err is human means not only that we must constantly struggle against error, but also that, even when we have taken the greatest care, we cannot be completely certain that we have not made a mistake.

Karl Popper, 'Knowledge and the Shaping of Reality'

Overview

Citation: Gray-Lobe, G., Pathak, P. A., and Walters, C. R. (2023). The Long-Term Effects of Universal Preschool in Boston. *Quarterly Journal of Economics*, Vol. 138, No. 1, pp. 363–411.

URL: <https://academic.oup.com/qje/article/138/1/363/6701924>

Abstract Summary: This study uses admissions lotteries to estimate the effects of large-scale public preschool in Boston on college-going, college preparation, standardized test scores, and behavioral outcomes. The findings illustrate possibilities for large-scale modern, public preschool and highlight the importance of measuring long-term and non-test score outcomes in evaluating the effectiveness of education programs.

Key Methodology: The study uses a lottery-based research design with admissions lotteries as instruments for preschool enrollment to estimate causal effects, employing two-stage least squares (2SLS) in the sample of randomized applicants.

Research Question: What are the long-term effects of large-scale public preschool in Boston on college-going, college preparation, standardized test scores, and behavioral outcomes?

Summary

Is It Credible?

Gray-Lobe et al. evaluate the long-term effects of Boston’s public preschool program using a lottery-based design. They claim that preschool enrollment “boosts college attendance as well as SAT test taking and high school graduation” and decreases “high school disciplinary measures including juvenile incarceration,” while having “no detectable effect on state achievement test scores” (p. 363). The article is notable for providing randomized, long-term evidence on a large-scale public program, tracking students for roughly two decades. The central narrative is that early childhood interventions can yield significant long-term educational and behavioral benefits even if medium-term test score gains fade out.

The headline claim regarding college enrollment is the most robust finding in the article. The authors utilize National Student Clearinghouse data, which suffers from very low attrition, to show an 8.3 percentage point increase in on-time college enrollment (p. 382). However, the claim that the program boosts college graduation is not statistically significant due to a lack of precision, as the youngest cohorts had not yet reached graduation age (p. 385). Furthermore, the estimated treatment effect represents a local average treatment effect relative to a mixed, unobserved counterfactual. The authors estimate that lottery losers often attended Head Start or private preschools (p. 379). Consequently, the study measures the value of winning a Boston public preschool lottery relative to a mix of alternative early childhood environments, rather than a strict comparison against no preschool at all.

The claims regarding secondary outcomes—such as high school graduation, disciplinary infractions, and the lack of effects on state achievement tests—require more cautious interpretation. These outcomes rely on the Massachusetts state administrative database, which exhibits significant differential attrition. Applicants who re-

ceived a preschool offer were 2.8 percentage points more likely to be observed in the state data and 3.8 percentage points more likely to have *any* state test score across grades 3–10 (p. 381). This differential follow-up compromises the initial random assignment, potentially introducing selection bias into the analysis of test scores, discipline, and high school graduation. The authors acknowledge this limitation, noting these results should be interpreted with caution, but they do not employ bounding exercises to quantify the potential magnitude of this bias.

Other methodological choices also complicate the secondary findings. The positive effect on SAT performance is largely driven by an 8.5 percentage point increase in the likelihood of taking the test at all (p. 388). While the authors find significant effects on achieving top-quartile scores, suggesting some intensive margin gains, the unconditional score effects heavily reflect this extensive margin (pp. 388–391). The reduction in juvenile incarceration relies on a very narrow measure—attendance at a Department of Youth Services facility—which has an extremely low base rate of 0.7% in the control group (p. 395). While the relative effect size is large (effectively negating the baseline risk), this measure may not capture broader criminal justice involvement. Additionally, the study tests a large number of outcomes without formal adjustments for multiple comparisons, raising the risk that some marginally significant findings are spurious.

Finally, the external validity of the findings is constrained by the historical context and the specific nature of the program. The cohorts studied applied between 1997 and 2003, a period predating the large-scale expansion of charter schools that now alters the educational landscape (p. 372). While described as a “universal” program because eligibility was not income-restricted—a definition consistent with academic literature (p. 368)—the program was heavily rationed and primarily served a disadvantaged, self-selected applicant pool (pp. 368–374). Extrapolating these historical results to contemporary proposals for fully funded, non-rationed universal entitlements should be done carefully. While the authors compare the program’s costs

to Head Start and other interventions, the absence of a formal cost-benefit analysis limits the study's direct policy utility, as the program was relatively expensive at roughly \$13,000 per student in 2020 dollars (p. 369). Despite these limitations, the study provides credible evidence that winning a lottery for this specific public preschool program meaningfully increased the likelihood of subsequent college enrollment.

The Bottom Line

The study provides credible evidence that winning a lottery to attend Boston's public preschool program in the late 1990s and early 2000s increased long-term college enrollment. However, claims regarding high school graduation, disciplinary improvements, and a lack of medium-term test score effects are less certain due to differential attrition in the state administrative data. Furthermore, the results represent the effect of the program relative to a mixed counterfactual of other preschool options and no preschool, complicating direct assessments of the program's absolute value.

Potential Issues

Differential attrition may bias findings for non-college outcomes: The study's causal claims for outcomes derived from the Massachusetts state administrative database (SIMS) are potentially compromised by significant differential attrition. The authors report that applicants who received a preschool offer were 2.8 percentage points more likely to be observed in the SIMS file and 3.8 percentage points more likely to have *any* state achievement test (MCAS) score across grades 3–10 (p. 381, Table II). This differential follow-up means the treatment and control groups are no longer as-good-as-randomly-assigned for the analysis of high school graduation, SAT-taking, disciplinary outcomes, and MCAS test scores. The authors acknowledge this limitation, stating that these results “should be interpreted with more caution than results for our primary postsecondary outcomes” (p. 381). They investigate the issue by showing that the post-attrition samples have observable imbalances; for instance, the offered group is significantly more likely to be female and have limited English proficiency (Appendix Table A5, p. 59). While the authors argue that this attrition may itself be a causal effect of preschool increasing attachment to the public school system (p. 381), the article does not present formal bounding exercises, such as Lee bounds, to quantify the potential magnitude of the selection bias. This omission is important, as the central narrative of test score “fade-out” and the positive findings on high school graduation and behavioral improvements rest on this potentially biased data.

The estimated treatment effect is relative to a mixed and unobserved counterfactual: The study estimates the Local Average Treatment Effect (LATE) for students who attend Boston Public Schools (BPS) preschool only when offered a spot. The interpretation of this effect is complicated because the alternative for these students is not necessarily the absence of preschool. The specific counterfactual for each student is unobserved. To approximate it, the authors use aggregate enrollment data

from a period of program expansion to estimate what lottery losers do instead, concluding that on the margin, new BPS students are drawn from Head Start (33%), private preschool (29%), and no preschool (38%) (p. 379). This analysis relies on the “strong assumption” that changes in BPS enrollment are unrelated to other factors affecting these shares. Because the counterfactual is an estimated mix of different alternatives rather than a single, known condition, it is difficult to isolate the value of public preschool compared to no preschool. While the authors argue this is the “policy-relevant parameter for evaluating expansions of BPS preschool” (p. 379), this mixed counterfactual complicates cost-benefit interpretations and understanding the specific sources of the program’s effects.

The study’s external validity is limited by its historical context and program characteristics: The findings are based on cohorts who applied to preschool between 1997 and 2003. The authors note that this period predates significant changes in the urban education landscape, such as the “large-scale expansion of Boston’s charter sector” (p. 372). They argue that the absence of a large charter sector is a strength for isolating the preschool effect, as charters could be a “confounding fallback option” in later years (p. 372), though they do find suggestive evidence that preschool reduced charter attendance in early grades (p. 400). Nonetheless, this historical context differs from the current policy environment. Furthermore, the article’s framing of the program as “universal” may not align with the common lay understanding of the term. The authors define “universal” based on eligibility being open to all residents regardless of income, following academic convention (p. 368), but they note that in practice, the program was heavily rationed. The applicant pool was therefore a self-selected, relatively disadvantaged group, not a representative sample of all Boston children (pp. 368, 374). While the authors suggest the program’s “unevenness or even mediocre implementation” may make the results more generalizable to the challenges of scaling up programs (p. 371), extrapolating findings from a rationed program serving a specific population two decades ago to contemporary

proposals for a fully-funded, non-rationed universal entitlement requires caution.

The analysis does not explore changes in parental investment as a potential mechanism: The article probes mechanisms for its long-term findings by examining student-level outcomes on the pathway to college, such as test scores, discipline, and peer effects (pp. 365–366). However, it does not investigate a plausible alternative mechanism: changes in parental behavior and investment. Access to free, public preschool provides a significant transfer of resources (both time and money) to families. These resources could be reinvested in children in other ways during their subsequent K–12 years, potentially driving the observed long-term effects on college-going. The study is limited by the absence of data on household expenditures or parental time use, and the authors appropriately focus on school-based mechanisms that their administrative data can illuminate (pp. 399–400). Still, the omission of the parental investment channel leaves a significant gap in understanding *why* the preschool program was effective.

The interpretation of SAT results is complicated by endogenous test-taking: The study finds that preschool attendance increases the rate of SAT test-taking by a statistically significant 8.5 percentage points (p. 388). This effect on the extensive margin complicates the interpretation of effects on SAT scores. The authors analyze unconditional scores (where non-takers are coded as zero) and find that preschool increases the likelihood of scoring above the bottom quartile and in the top quartile for math. They are transparent that this unconditional effect “combines the extensive-margin impact on SAT-taking and any intensive-margin effects on scores” (p. 391). They also correctly note that estimates of average scores conditional on taking the test “may be contaminated by composition effects” and are statistically indistinguishable from zero (p. 391). While the authors argue that the top-quartile effects are large enough to suggest real gains beyond just marginal entrants (p. 391), readers should be cautious not to interpret the general SAT results as unambiguous evidence of improved underlying academic ability, given the strong influence of test-taking rates.

The measure of juvenile incarceration is narrow: The study reports a statistically significant 1 percentage point reduction in juvenile incarceration, which is a key component of the positive finding for the high school disciplinary index (p. 395). The measure of incarceration is defined as “whether a student is ever observed attending a Massachusetts Department of Youth Services (DYS) school” (p. 394). This captures a severe outcome—residential commitment—but does not capture other significant interactions with the justice system, such as arrests or probation. The base rate for this outcome is extremely low in the control group (0.7%, p. 395), suggesting the measure captures only the most severe cases of delinquency. While the relative effect size is substantial—effectively eliminating the baseline risk—the specific finding on incarceration may not reflect the program’s impact on broader criminal justice involvement.

The absence of a cost-benefit analysis limits the study’s direct policy utility: The article is situated within a literature where cost-benefit analysis is a central feature of evaluating early childhood interventions. The authors report that the BPS program was relatively expensive, costing “roughly \$13,000 (2020 dollars) for full-day preschool,” which is more than many other state and federal programs (p. 369). While the article provides the first randomized long-term evidence on educational attainment for such a program and compares these effects to Head Start and Perry Preschool (p. 403), it does not attempt to monetize these benefits (e.g., through projected lifetime earnings) to compare them against the program’s costs. This omission makes it difficult for policymakers to assess whether the program represents an efficient use of public funds and limits the ability to judge the practical, economic significance of the findings.

The analysis of numerous outcomes lacks consistent adjustment for multiple comparisons: The study tests a large number of outcomes across several domains, including postsecondary enrollment, high school progress, SAT scores, and disciplinary incidents. When conducting many statistical tests, there is an elevated

risk of finding statistically significant results purely by chance. The article does not report using standard methods to adjust for multiple comparisons, such as Bonferroni corrections, in its main results tables. This concern is mitigated for some findings; the primary college enrollment effects are highly significant ($p < 0.01$), and the authors use a summary index for disciplinary outcomes to aggregate signals (p. 394). They also employ joint testing strategies in their heterogeneity analysis to address this concern (pp. 396–398). However, for main effect estimates that are only marginally significant, such as “Ever enrolling in any college” ($p < 0.1$) (p. 383), the lack of formal adjustment in the primary tables reduces confidence that the findings are not spurious.

The evidence for an effect on college graduation is statistically imprecise: The article’s abstract highlights that preschool boosts “college attendance,” and the analysis of this outcome is robust. The evidence for the more definitive outcome of college *graduation*, however, is not statistically significant. The authors report a point estimate suggesting a 5.2 percentage point increase in the likelihood of ever graduating from any college, but they are transparent about the lack of statistical power, stating, “we cannot reject that these graduation effects equal zero due to a lack of statistical precision” (p. 385, Table III). This imprecision is partly due to the fact that the two most recent cohorts in the study were too young to have their graduation outcomes observed. While the authors are clear about this limitation, the narrative emphasis on the “quantitatively large” point estimate (p. 385) should be weighed against the statistical uncertainty.

Minor transparency and presentation issues are present: Several minor issues related to sample definitions and baseline balance could be clarified. First, the main descriptive table shows a statistically significant pre-treatment imbalance on gender ($p < 0.01$), with the offered group being more likely to be female (p. 374, Table I). The authors address this by controlling for gender in all models and conducting a sensitivity analysis that shows the results are not driven by this control, but the

imbalance itself is notable (p. 385). Second, there is a small discrepancy between the total randomized sample of 4,215 applicants mentioned in the text (p. 375) and the sample of 4,175 used in the main regression tables (p. 383); this drop of 40 students is explained in the online appendix as a failure to match students to the NSC database (Appendix Table B1, p. 72). Finally, the subgroup analysis in Table VIII uses a smaller sample for the analysis by free/reduced-price lunch status because this variable is only available for students who appear in the state administrative data, a necessary data constraint that is explained in the table notes (p. 397).

Future Research

Bounding exercises for differential attrition: Future work should apply bounding techniques, such as Lee bounds, to the state administrative data to quantify the potential selection bias introduced by differential follow-up rates. This would clarify the robustness of the findings on high school graduation, discipline, and medium-term test scores.

Formal cost-benefit analysis: Given the high per-pupil cost of the Boston preschool program, future research could integrate the estimated effects on college enrollment and other outcomes into a comprehensive cost-benefit framework. Projecting lifetime earnings and monetizing behavioral impacts would help policymakers evaluate the program's return on investment beyond the comparative effectiveness analysis provided.

Exploring parental investment mechanisms: Future studies could investigate how access to public preschool alters household resource allocation. Collecting data on parental time use and expenditures would help determine whether the long-term benefits of preschool are mediated by parents reinvesting the time and money saved on childcare into other forms of child development.

© 2026 The Catalogue of Errors Ltd

This work is licensed under a

Creative Commons Attribution 4.0 International License

(CC BY 4.0)

You are free to share and adapt this material for any purpose,
provided you give appropriate attribution.

isitcredible.com