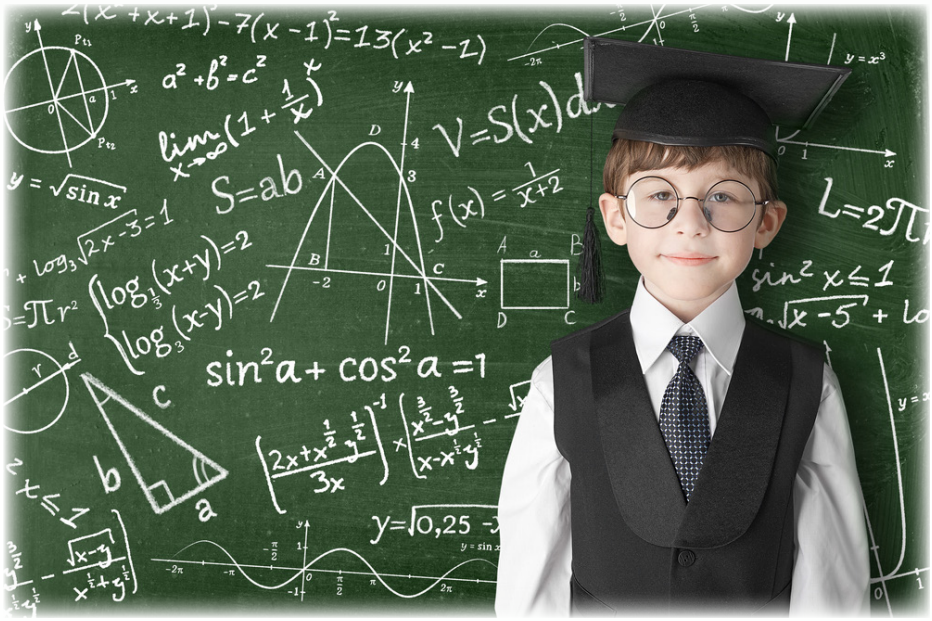
NEPC Review: Ohio's Lost Einsteins: The Inequitable Outcomes of Early High Achievers (Thomas B. Fordham Institute, September 2021)



Reviewed by:

Eunice S. Han and Yongmei Ni
University of Utah

December 2021

National Education Policy Center

School of Education, University of Colorado Boulder
Boulder, CO 80309-0249
(802) 383-0058
nepc.colorado.edu

Acknowledgements

NEPC Staff

Faith Boninger
Publications Manager

Francesca López
Academic Editor

Elaine Duggan
Production Design

Alex Molnar
Publications Director

Kevin Welner
NEPC Director

Suggested Citation: Han, E.S & Ni, Y. (2021). *NEPC review: Ohio's lost Einsteins: The inequitable outcomes of early high achievers*. Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/thinktank/einsteins

Funding: This review was made possible in part by funding from the Great Lakes Center for Educational Research and Practice.

NEPC Review: Ohio's Lost Einsteins: The Inequitable Outcomes of Early High Achievers (Thomas B. Fordham Institute, September 2021)

Reviewed by:

Eunice S. Han and Yongmei Ni
University of Utah

December 2021

Summary

A recent report from the Thomas B. Fordham Institute, titled *Ohio's Lost Einsteins: The Inequitable Outcomes of Early High Achievers,* examines the achievement growth of Ohio's "early high achievers," some of whom are identified as gifted and talented (GT). After tracking their academic performance from fourth grade to enrollment in college, the report finds that Black, Hispanic, and low-income, early high-achieving students tend to perform worse than White, non-disadvantaged peers in various educational outcomes. In addition, Black, Hispanic, and economically disadvantaged early high-achieving students are less likely to be identified as GT than their White, Asian/Pacific Islander, and non-disadvantaged peers. Finally, the report finds that GT identification improves the academic performance of early high achievers, especially for Black students and students in high-minority but low-poverty schools. Based on these findings, the report concludes that closing the "gifted identification gap" can help to close the "excellence gap." However, given the data limitations and key assumptions in its research design, this study does not support the causal inference on the effectiveness of GT identification. Thus, it seems premature to offer any clear policy suggestions for "the use of universal screening for GT services" to help early higher achievers "maintain their altitude" and close the "excellence gaps" for economically disadvantaged students and students of color.

NEPC
NATIONAL EDUCATION POLICY CENTER

NEPC Review: Ohio's Lost Einsteins: The Inequitable Outcomes of Early High Achievers (Thomas B. Fordham Institute, September 2021)

Reviewed by:

Eunice S. Han and Yongmei Ni
University of Utah

December 2021

I. Introduction

In September 2021, the Thomas B. Fordham Institute published *Ohio's Lost Einsteins: The Inequitable Outcomes of Early High Achievers,* authored by Scott Imberman, a professor of economics and education policy at Michigan State University.[1] Although many different analyses are conducted, the main purposes of the study are to examine whether identification as gifted makes a difference in the performance of early high-achieving students on state exams and long-term college-going outcomes, and whether closing the "gifted identification gap" is an effective way to close the "excellence gap."

Gifted/talented (GT) programs are provided in nearly all states, although they vary widely from state to state and even district to district within a state in terms of identification, access, funding, professional training, and services.[2] According to the Civil Rights Data Collection, in 2017-18, around 3.3 million, or 6.5% of public school students nationwide, enrolled in GT programs.[3] Although enrichment and acceleration programs have long been advocated to facilitate GT students' development of their full potential and continuous progress in school, existing studies examining the effects of GT programs on students' educational outcomes have showed mixed results. In addition, the effects vary considerably by programs and among students from different backgrounds.

This report examines whether GT identification is an effective way to help early high-achieving students maintain their academic status, especially for Black, Hispanic, economically disadvantaged, and other historically underrepresented students. Due to the lack of data on who actually received GT services in Ohio, however, this report examines whether GT identification improves student performance on state exams, and whether the effects vary among students from different backgrounds.

II. Findings and Conclusions of the Report

The report has two main parts. The first part identifies early high achievers, defined as students who scored in the top quintile (20%) on their third-grade state exams, and compares their achievement as they progress through elementary and middle school (until eighth grade) with those who scored in the bottom quintile on the same exams. It also looks at how early high achievers from different demographic groups compare to each other in long-term outcomes (e.g., high school performance and college attendance). The second part of the report turns to examine the impact of GT identification of early high achievers on various outcomes. The analysis is further disaggregated by student demographic information, including gender, race, and socioeconomic status.

The report yields the following main findings:

- Among all early high achievers, Black, Hispanic, and low-income students tend to perform worse than non-disadvantaged peers along multiple dimensions when simply comparing group means, but the differences become less consistent in the regression analyses where multiple variables are controlled.
- Black, Hispanic, and economically disadvantaged early high achievers are less likely to be identified for gifted services than their White, Asian/Pacific Islander, and non-disadvantaged peers.
- On average, GT-identified students perform better on all the long-run outcomes than non-GT students who are also early high achievers.
- GT identification has a small but positive impact on math achievement, especially for Black students and students in high-minority but low-poverty schools.

Based on these findings, the report asserts that obtaining GT status increases the chance of early high achievers sustaining their status, and that closing the "GT identification gap" could help close the "excellence gap" for economically disadvantaged students and students of color.

III. The Report's Rationale for Its Findings and Conclusions

The findings and conclusions are based on several ungrounded assumptions. First is the assumption that GT identification is a good proxy for GT program participation. The report uses findings from the evaluation of GT identification on student performance to infer the effectiveness of GT programs. Second, the report assumes that biases against economically disadvantaged students and students of color in the identification practices and the participation in GT services can be corrected by the statistical approaches of this study. The third assumption is that the student fixed-effects models are sufficient to control for the possibility of reverse causality (i.e., students are more likely to be identified as gifted if they perform well on state exams). As we elaborate below, these strong assumptions, combined with the narrow focus of the analysis and data limitations, weaken the conclusions of the report that

closing the "GT identification gap" could be an effective way to close the "excellence gap."

IV. The Report's Use of Research Literature

There are two main issues with the use of research literature. First, the report states that the empirical evidence of the impact of GT programs is mixed and lists several well-designed studies: four studies with no effects and two studies with significantly positive effects. However, the report does not provide any explanation—whether different designs of GT programs or the complexity of implementation, for instance, might have contributed to the mixed results—for the lack of consensus in the cited studies.

Second, and more importantly, the literature reviewed in the report does not serve its purpose, which is not about evaluating the effectiveness of GT programs but about "identifying" GT students as a way of promoting outcomes for high-achieving students. Accordingly, it would have been more appropriate to cite literature concerned with GT identification, such as the various issues related to the GT identification process in different states or the link between different identification processes and student outcomes.[4] In particular, it is critical to explain what "giftedness" is, what measures are used in identifying GT students, and which groups of students are more likely to be GT identified than others.[5] Whether GT-identified students among early high achievers tend to score higher on state exams than other non-identified students even before identification, and what school and district characteristics serve as the determinants for identification, are also important topics that the report overlooks. With the current presentation, it is difficult to situate this study in literature.

Additionally, because this study focuses on "GT identification," explaining potential channels through which GT identification, with or without the services, influences student outcomes is essential. The literature suggests a number of such channels, such as the right match between student ability and the level of instruction, teacher quality, and the composition of classroom peers in terms of gender and ethnicity,[6] but the report does not offer any discussion of these potential mechanisms. Thus, this study's contribution is minimal in advancing the existing knowledge on GT identification and services.

V. Review of the Report's Methods

The report utilizes several statistical techniques, including descriptive statistics by group, regression analysis, and student fixed-effects models. The descriptive analysis often compares the simple means between subgroups of students and offers a quick interpretation indicating that one group performs better than the other(s). However, without statistical testing of these differences, the interpretations could be misleading.[7]

With the linear regression analysis, the report overclaims that the method is able to find "virtually identical" non-GT identified students for GT-identified students by controlling some factors, including a set of cohort-by-school fixed effects based on students' third-grade

schools. However, there are many important factors that the regressions fail to control, including parents' attributes, teacher quality, classroom features, and school characteristics, many of which change from year to year. These regression analyses, therefore, by no means provide "apples-to-apples" comparisons.

The report argues that the results based on the fixed effects models in examining the effect of GT identification on student test scores are causal, "if one is willing to assume that the unobserved aspects of students that affect both GT identification and achievement are only made up of these 'time-invariant' characteristics."[8] However, this assumption is too strong to hold, because as described in the report, "assessment for gifted identification may occur either through a referral process that is largely subjective, as it is typically initiated by teachers and/or parents or via whole-grade screenings that occur at least once during grades K-2 and once during grades 3-6."[9] This implies that the assessment for GT identification can occur at different times through various ways and is based on student, parent, teacher, classroom, and school characteristics that change over time. Notably, the same set of variables may also affect student achievement on exams and who actually participates in GT programs. These multiple selection biases cannot be solved with the student fixed effects model, resulting in unreliable estimates. In addition, due to data limitations, the student fixed effects models are insufficient to correct the potential reverse causality: When students become high achievers, they are more likely to be identified as gifted, not the other way around. For these reasons, the causal inference of the effects of GT identification is not fully justified.

VI. Review of the Validity of the Findings and Conclusions

The dataset used in this study is appropriate to track early high achievers,[10] but not to tackle the issue of GT identification and GT programs, mainly due to the absence of individual students' data on the assessment results informing the GT identification, the actual participation in GT programs, and the longitudinal factors that vary across different classrooms within the same school. Moreover, GT identification faces critical measurement issues. GT assessment varies and is often subjective, compromising the reliability of this identification. The data also do not control for differences among districts that are more likely to conduct a referral vs. screening. Nor is information provided on what whole-grade screenings are based on. Based on the report's description of the Ohio GT programs (i.e., GT identification is often based on teacher referrals and the standards for enrollment in GT programs are restrictive), one would expect that the identification practices and the participation in GT services in Ohio may be biased against economically disadvantaged students and students of color. These biases, however, cannot be fully corrected by statistical models due to the data limitations described above.

Only 20-30% of identified GT students received services during the study. This makes it impossible to know if the performance gain of the GT students is the result of the learning experience under the GT programs or of high-ability students who are labeled as GT simply doing better without such programs. The finding that GT-identified students performed bet-

ter than non-GT students even though most of them did not receive any GT services may be indicating that GT programs are not effective. Or it may merely reflect the fact that the metrics used in GT identification are highly correlated with performance indices in state exams.

The report uses a combined race category Asian/Pacific Islander (API), but the US census has used a separate category for Asian American and for Native Hawaiian and Other Pacific Islander since 2000.[11] The academic performance of the two groups substantially differs, so it is vital to separate these two categories.[12]

In addition, the findings do not fully support the conclusion that "if early-high-achieving, low-income students and students of color were identified for GT programs at the same rate as their White and Asian high-achieving peers, this could reduce achievement gaps down the road"[13] because the report also finds that students in more heavily minority and lower-poverty schools benefit more from GT identification. However, heavily minority schools tend to be high-poverty schools, so these findings seem contradictive. One suggestion is to consider the interaction between the racial and SES subgroups (for example, Black and economically disadvantaged vs. Black and non-disadvantaged, etc.) to better understand students from diverse backgrounds.

VII. Usefulness of the Report for Guidance of Policy and Practice

The report provides interesting findings in its comparison of the paths of early high achievers by different subgroups. However, with the current datasets and study design, no causal inferences on the effectiveness of GT identification can be drawn. Also, the empirical findings of this study only show small magnitudes of the gains in test scores of the identified GT students, of which few actually received GT services. The results of this study, therefore, offer no solid base for clear policy suggestion as to whether more investment should be made for GT identification for early high achievers. Even if we take the findings at face value, this study fails to provide clear evidence for policy intervention for students of color and students of economically disadvantaged backgrounds, as it shows seemingly conflicting results for students from different subgroups. Based on the findings of this study, therefore, it seems premature to promote the increased "use of universal screening for GT services" as an effective way of helping early high achievers maintain their academic success and closing the achievement gaps.[14]

Notes and References

1 Imberman, S. (2021, September 26,). *Ohio's lost Einsteins: The inequitable outcomes of early high achievers.* Washington, DC: Thomas B. Fordham Institute. Retrieved October 20, 2021, from https://fordhaminstitute.org/ohio/research/ohios-lost-einsteins-inequitable-outcomes-early-high-achievers

2 Rinn, A.N., Mun, R.U., & Hodges, J. (2020). *2018-2019 State of the states in gifted education*. National Association for Gifted Children and the Council of State Directors of Programs for the Gifted. Retrieved October 20, 2021, from https://www.nagc.org/2018-2019-state-states-gifted-education

3 Calculated based on data retrieved October 20, 2021, from https://ocrdata.ed.gov/resources/downloaddatafile

4 For a review of literature on GT identification and its effectiveness, see Hodges, J., Tay, J., Maeda, Y., & Gentry, M. (2018). A meta-analysis of gifted and talented identification practices. *Gifted Child Quarterly, 62*(2),147-174.

Olszewski-Kubilius, P., & Corwith, S. (2018). Poverty, academic achievement, and giftedness: A literature review. *Gifted Child Quarterly*, *62*(1), 37-55.

Reis, S.M., & McCoach, D.B. (2000). The underachievement of gifted students: What do we know and where do we go? *Gifted Child Quarterly, 44*(3),152-170.

5 For example, high-ability learners are diverse and many underachieve on state exams for various reasons. GT identification tools, including teacher referrals, often are biased against diverse learners. For a review, see: Hodges J., Tay J., Maeda Y., & Gentry, M. (2018). A meta-analysis of gifted and talented identification practices. *Gifted Child Quarterly, 62*(2),147-174.

Olszewski-Kubilius, P., & Corwith, S. (2018). Poverty, academic achievement, and giftedness: A literature review. *Gifted Child Quarterly*, *62*(1), 37-55.

Reis, S.M., & McCoach, D.B. (2000). The underachievement of gifted students: What do we know and where do we go? *Gifted Child Quarterly*, *44*(3), 152-170.

6 Betts, J. R. (2011). The economics of tracking in education. In E.A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education*, *3,* 41-81. Amsterdam: North Holland.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentive, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review, 101*(5), 1739-1774.

7 For example, in Figures 2-4, are the differences statistically significant in a t-test? What are the sample sizes for each group category? What fraction of high-achieving students are boys or girls in math and English?

8 Imberman, S. (2021, September 26). *Ohio's lost Einsteins: The inequitable outcomes of early high achievers* (p. 10). Washington, DC: Thomas B. Fordham Institute. Retrieved October 20, 2021, from https://fordhaminstitute.org/ohio/research/ohios-lost-einsteins-inequitable-outcomes-early-high-achievers

9 Imberman, S. (2021, September 26). *Ohio's lost Einsteins: The inequitable outcomes of early high achievers* (p. 25). Washington, DC: Thomas B. Fordham Institute. Retrieved October 20, 2021, from https://fordhaminstitute.org/ohio/research/ohios-lost-einsteins-inequitable-outcomes-early-high-achievers

10 In the first part of the analysis, we have several comments: i) It is not clear whether the test scores are normally distributed and how the author standardizes the test scores. ii) In Figure 1, the author should explain the reason for sudden dips in the 5th grade in both math and English test scores. iii) In Figure 5, the solid horizontal line in the center of the box plot indicates "median" not "mean."

11 In the 2000 US Census, the Federal Government defines "Asian American" to include persons having origins

in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent. "Native Hawaiian and Other Pacific Islander" includes Native Hawaiian, Samoan, Guamanian or Chamorro, Fijian, Tongan, or Marshallese peoples and encompasses the people within the United States jurisdictions of Melanesia, Micronesia and Polynesia. The previous "Asian and Pacific Islander" (API) category was separated into "Asian American" and "Native Hawaiians and Other Pacific Islander" (NHOPI).

12 For instance, in 2016, the total college enrollment rate was substantially higher for Asian young adults (58 percent) than for Pacific Islander young adults (21 percent). Data source: Retrieved October 20, 2021, from https://nces.ed.gov/programs/raceindicators/indicator_rea.asp

13 Imberman, S. (2021, September 26). *Ohio's lost Einsteins: The inequitable outcomes of early high achievers* (p. 43). Washington, DC: Thomas B. Fordham Institute. Retrieved October 20, 2021, from https://fordhaminstitute.org/ohio/research/ohios-lost-einsteins-inequitable-outcomes-early-high-achievers

14 Imberman, S. (2021, September 26). *Ohio's lost Einsteins: The inequitable outcomes of early high achievers* (p. 44). Washington, DC: Thomas B. Fordham Institute. Retrieved October 20, 2021, from https://fordhaminstitute.org/ohio/research/ohios-lost-einsteins-inequitable-outcomes-early-high-achievers