Code for:
"Can Gifted Education Help
Higher-Ability Boys from Disadvantaged Backgrounds?"
David Card (UC Berkeley and NBER), Eric Chyn (UT Austin and NBER),
and Laura Giuliano (UC Santa Cruz and NBER)

Overview

This replication package includes the code to run the analysis in Card, Chyn and Giuliano (2025). The package includes three main folders:

- **Code:** This folder includes code to produce all figures and tables in the paper. All files are written in STATA, and the file "master.do" runs all necessary code.
    - Note that all the programs rely on restricted-access educational records or a derived output of these data. See below for additional information.
- **Input_data:** This folder is empty at baseline. Researchers who obtain the necessary restricted-access data can store the original data in this location.
- **Output:** This folder is empty at baseline and stores all figures and tables produced for the paper.

Data Availability and Provenance Statements

The main results in the paper rely on confidential education records from a large school district in Florida (hereafter, the "District"). Access was granted under a Data Use Agreement (DUA) that **prohibits disclosing the District's identity** and **prohibits providing access instructions or facilitating access for third parties**. The authors do not have redistribution rights. Accordingly, no third-party access is available.

Summary of Availability

- ☐ All data **are** publicly available.
- ☐ Some data **cannot be made** publicly available.
- ☒ **No data can be made** publicly available.

Details on each Data Source

Our analysis relies on the following original education record data files from the District:

1. Administrative educational records that include student-level records on demographics (e.g., gender and FRL-status); school enrollment; gifted status; courses taken (grades 3-12); state-wide standardized achievement (grades 3-8); academic grades (middle and high-school); disciplinary actions (grades 3-8); Stanford Achievement Test scores (2nd grade); PSAT and SAT scores (for those who took the exam during high-school); student surveys (available for select cohorts).

2. National Student Clearinghouse (NSC) records provided to the District (and therefore included in the files provided to us) and available for all students who completed high school in the District.

3. Gifted status IQ testing scores for students (i) tested by the District or (ii) submitted scores on their own.

We expect that the data will be preserved by the District indefinitely. We will preserve the data and code for a minimum of five years following publication of the paper in order to provide reasonable assistance to requests for clarification and replication.

Dataset List

The main data to conduct our analysis is student-level data from the District's record who were enrolled in 5th grade in the 2003-2012 academic years. The list below describes the list of individual files that are inputs to creating our main sample and our main analysis.

Data file	Notes
table1_column1.dta	All district students K-12
table1_columns2-6.dta	"Plan B" sample
planb_rdsmpl.dta	"Plan B" analytic sample with outcomes
appendix_table_a1.dta	Sample of fourth-graders used to estimate mean characteristics for compliers with GHA placement
appendix_table_a2.dta	Broad sample of "Plan B" eligible students
appendix_table_c2.dta	Sample with outcomes to conduct gifted-high-achiever (GHA) RD analysis
figure1.dta	Sample to create Figure 1 on achievement gaps by gender
appendix_figure_a5.dta	"Plan B" sample restricted to variables for heterogeneity analysis
appendix_figure_a2.dta	Sample of "Plan B" students with year first identified as gifted

Computational requirements

To replicate the results in the paper, you must first install Stata (code was last run with version 16.1). Note that the code uses the following STATA packages and schemes (for graphics):

modern (scheme)
coefplot (package)
seq (package)
outreg2 (package)

Memory and Runtime Requirements

Summary

Approximate time needed to reproduce the analyses on a standard (2025) desktop machine is one hour.

Details

The code was last run on a Dell Latitude 7430 laptop with a 12th-Gen Intel® Core™ i7-1265U CPU (10 cores / 12 logical processors, base 1.80 GHz), 16 GB RAM, running Microsoft Windows 11 Enterprise (version 10.0.22631, build 22631). Free-disk space at run time was not recorded; we recommend ≥50 GB free for smooth execution.

List of tables/figures and programs

The provided code reproduces:

- ☒ All tables and figures in the paper

Table Number	Program	Output File:
Table 1	JPE_TABLES.do	table_1_formatted.csv
Table 2	JPE_TABLES.do	table_2_formatted.csv
Table 3	JPE_TABLES.do	table_3_formatted.xls
Table 4	JPE_TABLES.do	table_4_formatted.xls
Table 5	JPE_TABLES.do	table_5_formatted.xls
Table 6	JPE_TABLES.do	table_6_formatted.xls
Table 7	JPE_TABLES.do	table_7.csv
Appendix Table A1	JPE_TABLES.do	table_A1.csv
Appendix Table A2	JPE_TABLES.do	table_A2_formatted.csv
Appendix Table A3	JPE_TABLES.do	table_A3_formatted.csv
Appendix Table A4	JPE_TABLES.do	table_A4_formatted.csv
Appendix Table A5	JPE_TABLES.do	table_A5_*.csv
Appendix Table A6	JPE_TABLES.do	table_A6_*.csv
Appendix Table A7	JPE_TABLES.do	table_A7.csv
Appendix Table A8	JPE_TABLES.do	table_A8_formatted.csv
Appendix Table A9	JPE_TABLES.do	table_A9_*.xls
Appendix Table A10	JPE_TABLES.do	table_a10_*.csv
Appendix Table C1	JPE_TABLES.do	table_C1_formatted.csv
Appendix Table C2	JPE_TABLES.do	table_C2.xlm
Appendix Table D1	JPE_TABLES.do	table_D1_formatted.csv
Appendix Table D2	JPE_TABLES.do	table_D2_*,csv

Figure 1	JPE_FIGURES.do	fig_1.png
Figure 2	JPE_FIGURES.do	fig_2*.png
Figure 3	JPE_FIGURES.do	fig_3*.png
Figure 4	JPE_FIGURES.do	fig_4*.png
Figure 5	JPE_FIGURES.do	fig_5.xlsx
Figure 6	JPE_FIGURES.do	fig_6.png
Figure 7	JPE_FIGURES.do	fig_7*.png
Figure 8	JPE_FIGURES.do	fig_8.png
Figure 9	JPE_FIGURES.do	fig_9*.png
Figure 10	JPE_FIGURES.do	fig_10*.png
Figure 11	JPE_FIGURES.do	fig_11*.png
Figure 12	JPE_FIGURES.do	fig_12*.png
Appendix Figure A1	JPE_FIGURES.do	fig_A1.png
Appendix Figure A2	JPE_FIGURES.do	fig_A2.png
Appendix Figure A3	JPE_FIGURES.do	fig_A3*.png
Appendix Figure A4	JPE_FIGURES.do	fig_A4*.png
Appendix Figure A5	JPE_FIGURES.do	fig_A5.png
Appendix Figure A6	JPE_FIGURES.do	fig_A6*.png
Appendix Figure A7	JPE_FIGURES.do	fig_A7*.png
Appendix Figure B1	JPE_FIGURES.do	fig_B1*.png
Appendix Figure B2	JPE_FIGURES.do	fig_B2*.png
Appendix Figure B3	JPE_FIGURES.do	fig_B3.png
Appendix Figure B4	JPE_FIGURES.do	fig_B4*.png
Appendix Figure C1	JPE_FIGURES.do	fig_C1*.png
Appendix Figure D1	JPE_FIGURES.do	fig_D1*.png
Appendix Figure D2	JPE_FIGURES.do	fig_D2*.png
Appendix Figure D3	JPE_FIGURES.do	fig_D3*.png
Appendix Figure D4	JPE_FIGURES.do	fig_D4*.png

Notes:

Code produces CSV and XLS/XML files. Additional steps are taken to manually format this output to create the tables, and figure 5, as they appear in the manuscript.

JPE_TABLES.do includes code that implements an adjustment for multiple hypothesis testing. This code is based on the approach described in Anderson (2008).

References

Anderson, Michael L. "Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American statistical Association* 103, no. 484 (2008): 1481-1495.

Acknowledgements

This ReadMe file is based on the Social Science Data Editors' template. See https://socialscience-data-editors.github.io/template_README/.

Appendix A: Codebook / Data Dictionary

Note: List covers all variables that appear in analysis datafiles.

Variable	Description
above_med_g3	Indicator for grade 2 standardized score above median
above_med_sch_frl	Indicator for school-level FRL fraction above median
above_med_sch_zg5_mean	Indicator for school-level average of grade 2 standardized score above median
age	Student age in grade 5
alg8	Indicator for algebra enrollment in 8th grade
any_enroll_yr2	Indicator for graduating HS on-time and college enrollment within 2 years
black	Student race: Black indicator
bw15gem	Indicator for sample bandwidth of 15 for GEM analysis
bw20gem	Indicator for sample bandwidth of 20 for GEM analysis
bw25gem	Indicator for sample bandwidth of 25 for GEM analysis
cohort	Year student was enrolled in 5th grade
coll_1yr	Indicator for graduating HS and college enrollment within 1 year
collot	Indicator for on-time college enrollment
collot_v2	Indicator for on-time college enrollment (only for cohorts 2003-2011)
d	Above-cutoff/eligibility indicator
d	Above-cutoff/eligibility indicator, RD analysis of 4th-grade GHA for high achievers
diq	Above-cutoff/eligibility indicator, RD analysis of 4th-grade GHA for gifted
donut	Indicator used for “donut” RD/sample restriction
dxr	Interaction term for variables “r” and “d”
dxriq	Interaction of diq X riq
enjoy_g3	Grade 3 survey indicator for high enjoyment of learning (score of 5)
f	Female indicator (binary 0/1).
fcat_678	Standard test score (average) in grades 6-8
fcat678_avgany	Standard test score (average of math and reading) in grades 6-8
fcat678_math	Standard math test score (average) in grades 6-8
fcat678_read	Standard reading test score (average) in grades 6-8
frl	Free/reduced-price lunch eligibility indicator
g2_zscore	Grade 2 overall standardized FCAT score (z-score)
g2math	Grade 2 math FCAT score (z-score)
g2read	Grade 2 reading FCAT score (z-score)
g3fcat	Grade 3 average FCAT score
g3math	Grade 3 math FCAT score
g3read	Grade 3 reading FCAT score
g3rm	Grade 3 overall FCAT score (zcore)
g3svy_enj	Grade 3 student survey enjoyment measure (0-5 scale)
g6svy	Grade 6 response to challenging course survey question

Variable	Description
gem6	Indicator for GEM enrollment in 6th grade
gemd	Above-cutoff/eligibility for GEM indicator
gemdxr	Interaction term for variables “gemr” and “gemd”
gemelig	Indicator for GEM eligibility
gemr	Distance from GEM cutoff (running variable)
gf	Indicator for gifted (as of 4th grade)
gha	Indicator for gifted, high-achiever classroom participation
gifted	Gifted program participation indicator (binary 0/1).
gpa_hs	All course GPA In high school
gr3zmath	Grade 3 reading FCAT score (z-score)
gr3zread	Grade 3 reading FCAT score (z-score)
gr4cavgLmath	Average grade 3 math FCAT scores of peers in 4th-grade classroom
gr4cavgLread	Average grade 3 reading FCAT scores of peers in 4th-grade classroom
gr4cpctgifted	Gifted share of peers in 4th-grade classroom
grad	Indicator for graduating HS
grad_v2	Indicator for grading HS (only for cohorts 2003-2011)
grdot	Indicator for graduating HS on-time
grdot_v2	Indicator for graduating HS on-time (only for cohorts 2003-2011)
ha	Indicator for non-gifted high achiever
has_g6svy	Indicator for having responses to grade 6 challenging course survey question
has_sat	Indicator for non-missing SAT score
hipctgft	School-level indicator for gifted share of GHA classroom between 0.3 and 1
hispanic	Student ethnicity: Hispanic indicator
lang_arts_gpa_hs	Language-arts (LA) course GPA in HS
lep	Limited English / English learner indicator
m	Male indicator (binary 0/1).
math_gpa_g68	Math GPA in grades 6-8
math_gpa_hs	Math course GPA in high school (HS)
maxpsatmathpct	Maximum percentile on PSAT math
maxpsatreadpct	Maximum percentile on PSAT reading
maxpsattotpct	Maximum percentile on PSAT (all components)
minority	Indicator for whether student race is Black or ethnicity is Hispanic
nai	Non-verbal ability index from 2nd-grade gifted screening test (NNAT)
no_susp_ms	Indicator for no suspensions in middle school
num_ap_course	Number of AP courses enrolled
peer_g35fcat_g68	Grades 6-8 peer quality in all courses, measured by their grades 3-5 FCAT avg
peer_g35fcat_g68_lang	Grades 6-8 peer quality in LA courses, measured by their grades 3-5 FCAT avg
peer_g35fcat_g68_math	Grades 6-8 peer quality in math courses, measured by their grades 3-5 FCAT avg
peer_g35fcat_hs	HS peer quality in all courses, measured by their grades 3-5 FCAT avg

Variable	Description
peer_g35fcat_hs_lang	HS peer quality in LA courses, measured by their grades 3-5 FCAT avg
peer_g35fcat_hs_math	HS peer quality in math courses, measured by their grades 3-5 FCAT avg
pr_collot	Predicted prob on-time college enrollment using 3rd-grade FCAT scores
pr_collot_g2nai	Predicted prob on-time college enrollment using 2nd-grade nonverbal ability index
pr_collot_g2sat	Predicted prob on-time college enrollment using 2nd-grade SAT scores
prntlang_eng	Parent language is English
r	Distance from IQ cutoff (running variable)
riq	Running variable in RD analysis of 4^{th}-grade GHA participation for gifted
sat	SAT score
sch_black	School-level Black composition
sch_frl	School-level FRL composition
sch_hispanic	School-level Hispanic composition
sch_medhhy	School-level median household income of students ZIP code (2000 Census)
sch_nonwhite	School-level share of students who are non-white
sch_zg5_mean	School-level mean standardized achievement in grade 5 measure (z-score mean)
sch_zsat_mean	School-level mean standardized achievement measure (z-score mean)
schid	School identifier
smpl	Indicator for inclusion in Plan B sample (defined for entire student population)
studid	Student identifier
tr	Treatment indicator for participating in 4^{th} grade GHA classroom
white	Student race: White indicator
x	Running variable in RD analysis of 4^{th}-grade GHA participation for high achievers