Readme for replicating results in “Genetic Endowments, Educational Outcomes, and the Moderating Influence of School Quality”, by Benjamin W. Arold, Paul Hufe, and Marc Stoeckli.

If experiencing trouble running the code or replicating results, contact Paul Hufe (paul.hufe@bristol.ac.uk).

Overview

This replication package generates all tables, and figures of the paper “Genetic Endowments, Educational Outcomes, and the Moderating Influence of School Quality.”

- All raw data shall be stored in folder “rawdata.”

- All script files shall be stored in the folder “scripts.”

- All (user-written) packages for statistical programs shall be installed in the folders “scripts/libraries/python,” “scripts/ libraries/R,” “scripts/ libraries/stata.”

Raw data is not shared as part of the replication package. However, download and storage instructions are provided below.

Some analyses will require the installation of additional user-written packages that may experience version updates over time (see above). Contact Paul Hufe (paul.hufe@bristol.ac.uk) for more information which versions of user-written packages were used for the analyses. We are happy to share the folder “scripts/libraries” used for our analyses with replicators upon request.

Data Availability and Provenance Statements

The National Longitudinal Study of Adolescent to Adult Health (Add Health)

Citation	Harris, K.M., C.T. Halpern, E.A. Whitsel, J.M. Hussey, L. Killeya-Jones, J. Tabor, and S.C. Dean. 2019. Cohort Profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health). *International Journal of Epidemiology* 48(5):1415-1425
Location in archive	“rawdata/add health”
Last download	01.12.2023
Data dictionary	https://addhealth.cpc.unc.edu/documentation/codebook-explorer/#/

US Census 1970 Extract

Citation	Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Grace Cooper, Julia A. Rivera Drew, Stephanie Richards, Renae Rodgers, Jonathan Schroeder, and Kari C.W. Williams. IPUMS USA: Version 16.0 [dataset]. Minneapolis, MN: IPUMS, 2025. https://doi.org/10.18128/D010.V16.0
Location in archive	“rawdata/census1970/census.dta”
Last download	04.03.2020
Data dictionary	https://usa.ipums.org/usa-action/samples

American Community Survey (ACS)

Citation	Steven Ruggles, Sarah Flood, Matthew Sobek, Daniel Backman, Grace Cooper, Julia A. Rivera Drew, Stephanie Richards, Renae Rodgers, Jonathan Schroeder, and Kari C.W. Williams. IPUMS USA: Version 16.0 [dataset]. Minneapolis, MN: IPUMS, 2025. https://doi.org/10.18128/D010.V16.0
Location in archive	"rawdata/acs/ usa_00005.dta"
Last download	09.08.2021
Data dictionary	https://usa.ipums.org/usa-action/samples

Current Population Survey (CPS)

Citation	Sarah Flood, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Megan Schouweiler, and Michael Westberry. IPUMS CPS: Version 12.0 [dataset]. Minneapolis, MN: IPUMS, 2024. https://doi.org/10.18128/D030.V12.0
Location in archive	"rawdata/cps/cps.dta" "rawdata/cps/cps_1994.dta"
Last download	09.08.2021
Data dictionary	https://cps.ipums.org/cps-action/variables/group

Description of programs/code

Software requirements.

All analyses were run on Windows 11 using Stata version 18, R version 4.4.3, Python 3.8.19 called through Windows PowerShell ISE (x86).

Runtime.

Approximate time needed to reproduce the analyses on a standard desktop machine (PC (64-bit x86-64), Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz, 16GB RAM): 11 hours.

Most of the calculation time (>10h) is due to the simulations for Appendix C.

Instructions for Replicators.

1. Open replication directory and store folders "scripts" and "rawdata" as described above.
2. Open "scripts/_master.do" and set relevant paths.
3. Install (user-written) packages for statistical programs into the folders "scripts/libraries/python," "scripts/ libraries/R," "scripts/ libraries/stata."
4. Execute script-files following the sequence indicated in "scripts/_master.do."