"Evolution vs. Creationism in the Classroom: The Lasting Effects of Science Education"

Benjamin W. Arold

Replication README

Overview

The code in this replication package creates all empirical figures and tables presented in the main paper and online appendix. The replication process is organized by dataset. It is documented in "master_file.do". The replication package consists of the master file "master_file.do", and the three folders "data", "code", and "output". The folder "data" contains eight subfolders, one for each dataset. The folder "code" contains eight subfolders, one for each analysis. The folder "output" contains two subfolders, one for "tables" and one for "figures".

As visible in "master_file.do", the setup and installation of Stata packages is followed by the analysis. The analysis is conducted in eight steps corresponding to eight datasets. For a given step using dataset X, the file "X_prepare.do" converts the raw datafiles to the analysis datafile, and the file "X_output.do" generates the tables and figures that appear in the paper. All of these files are referenced to in "master_file.do".

This Replication README consists of six sections. The "Data Availability and Provenance Statement" is followed by the sections "Computational Requirements", "Instructions to Replicators", "List of Tables, Figures and Programs", "Statement about Rights", and "References".

Data Availability and Provenance Statement

This paper uses 4 individual-level datasets and 4 state-level datasets. The following table provides the name, description, source, and provision status of all 8 datasets. I do not have permission to post the 4 individual-level datasets on a public repository. Hence, below the table, I provide detailed information on how researchers interested in obtaining restricted-use data licenses can proceed.

Dataset	Description	Source	Provided

naep.dta	Contains individual-level data of the National Assessment of Educational Progress (NAEP), 1996-2009 Science Assessments for Grade 12	U.S. Department of Education, National Center for Education Statistics	No
gss.dta	Contains individual-level data of the General Social Survey (GSS), waves 2006-2016	National Opinion Research Center (NORC) at the University of Chicago (Davern et al., 2024)	No
acs.dta	Contains individual-level data of the American Community Survey (ACS), waves 2000-2017	Integrated Public Use Microdata Series (IPUMS USA) (Ruggles et al., 2020)	No
hsbt.dta	Contains individual-level data of the US National Survey of High School Biology Teachers	The Association of Religion Data Archives (ARDA), National Survey of High School Biology Teachers, 2007 (Plutzer, Eric and Michael B. Berkman). (www.theARDA.com)	No
vaccinations.dta	Contains state-level data about COVID-19 vaccination	Barro (2022), Opportunity Insights, Economic Tracker (Chetty, Friedman, Stepner, et al. (2023), Centers for Disease Control and Prevention, New York	Yes

	rates and related political and societal variables	Times, Johns Hopkins Coronavirus Resource Center, U.S. Department of Health & Human Services, Federal Election Commission, General Social Survey, Mead and Mates (2009)	
word_count.xlsx	Contains state-level word count data from US State Science Education Standards	Author's collection of US State Science Education Standards	Yes (Note: File Contains only word counts. Original text not provided due to copyright laws)
statevars.dta governors19912009.dta	Contains state-level data about economic, political, and educational conditions	Bureau of Economic Analysis, Websites of United States House of Representatives/Senate, State Elections Offices, wikipedia.org, National Center for Education Statistics, International Center for Home Education Research	Yes
googletrends.dta	Contains state-level data about search frequencies of given keywords in google	Google Trends (https://www.google.com/trends)	Yes

Information on how to get access to the 4 individual-level datasets:

1) National Assessment of Educational Progress (NAEP):

For all analyses using student achievement and particularly "knowledge about evolution", I make use of the NAEP. The NAEP is a congressionally mandated

project also known as the Nation's Report Card. It is administered by the National Center for Education Statistics (NCES), a body within the Institute of Education Sciences (IES) and the US Department of Education.

I use restricted individual-level data from the Main-NAEP, Science Test, Grade 12. To access the data, I obtained a restricted-use data license for NAEP from NCES. The license is intended to protect the confidentiality of individual students and schools.

I am able to post all of the code used for the data analysis, however, I do not have the permission to post the data on a repository. Researchers interested in obtaining the restricted-use data license for NAEP need to apply at the IES through Electronic Application System (https://nces.ed.gov/statprog/instruct.asp). I will assist with any reasonable replication attempts for two years following publication. The IES Data Security Office can also be contacted for further information (as of May 22, 2024):

IES Data Security Office

Department of Education/IES/NCES

550 12th Street, SW, Room 4055

Washington, DC 20202

Mishéla King

(202) 245-6146

IESData.Security@ed.gov

2) General Social Survey (GSS)

For all analyses using attitudes and beliefs and particularly "belief in evolution", I make use of the GSS. The GSS is a biennial cross-sectional survey which monitors societal change in the US. It is administered by the National Opinion Research Center (NORC) at the University of Chicago.

I use restricted individual-level data of the GSS. To access the data, I obtained a restricted-use data license for GSS from NORC. The license is intended to protect the confidentiality of survey respondents.

I am able to post all of the code used for the data analysis, however, I do not have the permission to post the data on a repository. Researchers interested in obtaining the restricted-use data license for GSS need to submit a GSS Sensitive Data Application at NORC by email to GSS@NORC.org. The application process is described here: https://gss.norc.org/documents/other/ObtainingGSSSensitiveDataFiles.pdf. I will assist with any reasonable replication attempts for two years following publication. Further information can also be obtained here:

NORC at the University of Chicago

1155 E 60th Street

Chicago, IL 60637

GSS@norc.org

3) American Community Survey (ACS)

For all analyses using occupational choice and particularly "the probability to work in life sciences", I make use of the ACS. The ACS is a large-scale demographic survey which draws from a national random sample of the US population. The ACS is an annual cross-sectional survey to which participants are required to answer by law. It is administered by the U.S. Census Bureau.

I use restricted individual-level data of the ACS from IPUMS USA. IPUMS USA collects, preserves and harmonizes U.S. census microdata and provides easy access to this data with enhanced documentation. To access the data, I obtained a user license from IPUMS USA. The license is intended to protect the confidentiality of survey respondents.

I am able to post all of the code used for the data analysis, however, I do not have the permission to post the data on a repository. Researchers interested in obtaining the user license for ACS need to submit an application form at https://uma.pop.umn.edu/usa/user/new?return_url=https%3A%2F%2Fusa.ipums.org%2Fusa-action%2Fmenu. I will assist with any reasonable replication attempts for two years following publication. Further information can also be obtained here:

ipums@umn.edu

4) National Survey of High School Biology Teachers (HSBT)

For all analyses using high school biology teachers, I make use of the HSBT. The HSBT is a nationally representative probability sample of U.S. public high school biology teachers. It was conducted by the Survey Research Center of Penn State University in 2007 and focuses on the biology teachers' approach to teaching evolution (and creationism) in the classroom, as well as their educational background and personal attitudes on evolution.

I use individual-level data of the HSBT, as made available by the Association of Religion Data Archives (www.theARDA.com). To access the data, I entered an agreement about the data usage.

I am able to post all of the code used for the data analysis, however, I do not have the permission to post the data on a repository. Researchers interested in obtaining the restricted-use data license for HSBT need to agree with the terms and conditions of data usage of the ARDA (https://www.thearda.com/data-archive?fid=HSBIOL&tab=3) and download the dataset here: https://www.thearda.com/data-archive?fid=HSBIOL. I will assist with any reasonable replication attempts for two years following publication. Further information can also be obtained here:

E-mail: support@thearda.com

Questions by phone can be made by calling The Center for the Study of Religion and American Culture: Phone: (317) 274-8409

Computational Requirements

The code contained in this replication package is run in Stata. The code was last run with Stata version 18. The program "master_file.do" will automatically install all packages locally that need to be installed manually. These include event_plot, csdid, estout, coefplot, honestdid, ftools, reghdfe (all in the versions as of 2024-05-01).

Note: The dataset "word_counts.xlsx" is generated from texts using Python. I do not have the permission to provide the underlying texts for replication (see "Data Availability and Provenance Statements"), but I provide the Python file used to generate "word_counts.xlsx" from the underlying texts. It is stored here: "Replication Package\code\text_analysis\word_count.ipynb". It is a Jupyter Notebook that can be run in the Anaconda Navigator environment, using a Jupyter Notebook of Version 6.5.4.

The approximate time needed to reproduce the analyses on a standard 2024 desktop machine amounts to roughly 6 weeks. The code to generate Figure A.VI (Placebo Tests: Density Plot of Placebo Estimates of Effect of Evolution Coverage in Science Standards on Probability of Working in Life Sciences, Based on 1000 Permutations) takes roughly 5 weeks to run, the code to generate all other Figures and Tables takes roughly 1 week to run. The code was last run on a Laptop with 11th Gen Intel(R) Core(TM) i7-1185G7 @ 3.00GHz, 32.0 GB (31.7 GB usable) installed RAM, and Windows 10 Education.

Instructions to Replicators

A replicator should be able to submit requests for the restricted-use data licenses of the individual level datasets (NAEP, GSS, ACS; HSBT) to relevant agencies described in the section "Data Availability and Provenance Statements" of this document. With all the relevant files saved in the corresponding folders, a replicator should run the programs in "0_master.do" which converts the raw datafiles to the analysis datafiles, and generates the tables and figures of the paper. Before running the file, the replicator should update the file path where the replication package is stored in line 22 of the file "0_master.do".

As visible in "master_file.do", the setup and installation of Stata packages is followed by the analysis. The analysis is conducted in eight steps corresponding to eight datasets. For a given step using dataset X, the file "X_prepare.do" converts the raw datafiles to the analysis datafile, and the file "X_output.do" generates the tables and figures that appear in the paper. All of these files are referenced to in "master_file.do".

List of Tables, Figures and Programs

To facilitate the replication, the following table provides an overview of all tables and programs figures of the paper, the program used to generate these tables and figures, and the specific line number within the program.

Table\Figure #	Program	Line Number
Table I, Panel A	code/naep_analysis/naep_output.do	9
Table I, Panel B	code/gss_analysis/gss_output.do	9
Table I, Panel C	code/acs_analysis/acs_output.do	9
Table II, Panel A	code/naep_analysis/naep_output.do	31
Table II, Panel B	code/gss_analysis/gss_output.do	25
Table II, Panel C	code/acs_analysis/acs_output.do	26
Table A.I	code/vaccination_analysis/vaccinations_output.do	9
Table A.II	code/vaccination_analysis/vaccinations_output.do	15
Table A.III	code/naep_analysis/naep_output.do	84
Table IV, Panel A	code/naep_analysis/naep_output.do	112
Table IV, Panel B	code/gss_analysis/gss_output.do	69
Table IV, Panel C	code/acs_analysis/acs_output.do	68
Table A.V	code/naep_analysis/naep_output.do	132
Table A.VI	code/naep_analysis/naep_output.do	150
Table A.VII	code/gss_analysis/gss_output.do	88
Table A.VIII	code/gss_analysis/gss_output.do	141
Table A.IX, Panel A	code/gss_analysis/gss_output.do	179
Table A.IX, Panel B	code/gss_analysis/gss_output.do	194
Table A.X, Panel A	code/gss_analysis/gss_output.do	269
Table A.X, Panel B	code/gss_analysis/gss_output.do	283
Table A.XI	no code	-

Table A.XII	code/naep_analysis/naep_output.do	164
Table A.XIII	code/gss_analysis/gss_output.do	360
Table A.XIV	code/acs_analysis/acs_output.do	87
Table A.XV	code/gss_analysis/gss_output.do	576
Table A.XVI	code/text_analysis/text_output.do	9
Table A.XVII	code/states_analysis/states_output.do	9
Table A.XVIII	code/googletrends_analysis/googletrends_output.do	9
Table A.XIX	code/naep_analysis/naep_output.do	293
Table A.XX	code/naep_analysis/naep_output.do	302
Table A.XXI	code/gss_analysis/gss_output.do	486
Table A.XXII	code/gss_analysis/gss_output.do	494
Table A.XXIII	code/gss_analysis/gss_output.do	502
Table A.XXIV	code/gss_analysis/gss_output.do	510
Table A.XXV	code/gss_analysis/gss_output.do	518
Table A.XXVI	code/acs_analysis/acs_output.do	201
Table A.XXVII	code/hsbt_analysis/hsbt_output.do	9
Table A.XXVIII	code/acs_analysis/acs_output.do	213
Table A.XXIX	code/gss_analysis/gss_output.do	540
Table A.XXX	no code	-
Table A.XXXI	no code	-
Table A.XXXII	code/naep_analysis/naep_output.do	314

Table A.XXXIII	code/gss_analysis/gss_output.do	555
Table A.XXXIV	code/acs_analysis/acs_output.do	252
Table A.XXXV	code/states_analysis/states_output.do	29
Figure I	no code	-
Figure II, Panel A	code/naep_analysis/naep_output.do	331
Figure II, Panel B	code/naep_analysis/naep_output.do	433
Figure III, Panel A	code/gss_analysis/gss_output.do	595
Figure III, Panel B	code/gss_analysis/gss_output.do	698
Figure IV	code/acs_analysis/acs_output.do	268
Figure V, Panel A	code/acs_analysis/acs_output.do	288
Figure V, Panel B	code/acs_analysis/acs_output.do	390
Figure A.I	code/hsbt_analysis/hsbt_output.do	29
Figure A.II, Panel A	no code	-
Figure A.II, Panel B	no code	-
Figure A.III, Panel A	code/gss_analysis/gss_output.do	807
Figure A.III, Panel B	code/naep_analysis/naep_output.do	540
Figure A.IV	code/naep_analysis/naep_output.do	552
Figure A.V	code/gss_analysis/gss_output.do	819
Figure A.VI	code/acs_analysis/acs_output.do	498
Figure A.VII, Panel A	code/naep_analysis/naep_output.do	762
Figure A.VII, Panel B	code/naep_analysis/naep_output.do	771
Figure A.VIII, Panel A	code/gss_analysis/gss_output.do	1033
Figure A.VIII, Panel B	code/gss_analysis/gss_output.do	1042
Figure A.IX, Panel A	code/acs_analysis/acs_output.do	707

Figure A.IX, Panel B	code/acs_analysis/acs_output.do	716
Figure A.X, Panel A	code/naep_analysis/naep_output.do	780
Figure A.X, Panel B	code/naep_analysis/naep_output.do	882
Figure A.XI, Panel A	code/gss_analysis/gss_output.do	1051
Figure A.XI, Panel B	code/gss_analysis/gss_output.do	1153
Figure A.XII, Panel A	code/acs_analysis/acs_output.do	725
Figure A.XII, Panel B	code/acs_analysis/acs_output.do	827
Figure A.XIII	no code	-
Figure A.XIV	code/naep_analysis/naep_output.do	984
Figure A.XV	code/gss_analysis/gss_output.do	1255

Statement about Rights

☑ I certify that the author of the manuscript has legitimate access to and permission to use the data used in this manuscript.

References

Barro, Robert J., "Vaccination rates and COVID outcomes across U.S. states," Economics & Human Biology, 47 (2022), 101–201

Chetty, Raj, John N Friedman, Michael Stepner, et al., "The economic impacts of COVID 19: Evidence from a new public database built using private sector data," Quarterly Journal of Economics, (2023)

Davern, Michael; Bautista, Rene; Freese, Jeremy; Herd, Pamela; and Morgan, Stephen L.; General Social Survey 1972-2016. [Machine-readable data file]. Principal Investigator, Michael Davern; Co-Principal Investigators, Rene Bautista, Jeremy Freese, Pamela Herd, and Stephen L. Morgan. NORC ed. Chicago, (2024)

Mead, Louise and Anton Mates, "Why Science Standards are Important to a Strong Science Curriculum and How States Measure Up," Evolution: Education and Outreach, 2 (2009), 359–371

Plutzer, Eric and Michael B. Berkman, "The Association of Religion Data Archives (ARDA)," National Survey of High School Biology Teachers (2007)

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek, “Integrated Public Use Microdata Series USA: Version 10.0 [dataset],” Minneapolis, MN: IPUMS (2020)