



Reproducible Research Repository

Federal Research Funding and STEM Education

First Submission: RR_USA_2026_618

Eric Sabiti

reproducibility@worldbank.org

May 26, 2026

This review verifies the reproducibility of the exhibits included in the paper “*Federal Research Funding and STEM Education*”.

Contents in this review:

1. Main findings
2. List of exhibits and reproducibility status
3. Reproduction Environment

Main findings

- **Every exhibit in the main section of the paper and 10 randomly selected exhibits in the appendix were reproduced accurately**
- The code was successfully executed on a new computer after:
 1. Updating the working directory in line 9 of the do-file "main.do", and running it.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 50 minutes to run.
- We conducted our reproducibility analysis based on the paper shared by the authors.
- **Verification Process and Data Handling:**
 - The reproducibility package relies on 2 types of data: open, and limited-access (available to data users at ICPSR member institutions).
 - Open data is included in the public reproducibility package.
 - The ICPSR HEGIS microdata used in the first stage of the analysis are restricted to users at ICPSR member institutions and could not be accessed by the reviewers. The first part of the verification was therefore conducted virtually: the authors ran the scripts that construct the intermediate HEGIS datasets and shared the resulting processed files with the reviewers.
 - *data_hash_report.csv* lists the SHA256 hashes of all files in the Data folder to support data integrity checks. Users who acquire the open data independently can use this file to verify that the data has not been altered.

- *data_hash_report_author.csv* lists the SHA256 hashes of all files in the "HEGIS_Degrees" folder to support data integrity checks. Users who acquire the limited-access data can use this file to verify that the data has not been altered.
- **Reproducibility Summary:**
 - **Data:** Some data is limited-access and has not been included in the reproducibility package. For more details, please refer to the README file.
 - **Code:** All code files (from cleaning to analysis) are included in the reproducibility package.
 - **Outputs:** All outputs are generated by code included in the reproducibility package.
 - **Dependencies environment:** The reviewers created a new environment using the latest versions of dependencies available at the moment of the review.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Figure 1 (a)** Reproduced.
- **Figure 1 (b)** Reproduced.
- **Figure 1 (c)** Reproduced.
- **Figure 1 (d)** Reproduced.
- **Figure 1 (e)** Reproduced.
- **Figure 1 (f)** Reproduced.
- **Figure 2 (a)** Reproduced.
- **Figure 2 (b)** Reproduced.
- **Figure 2 (c)** Reproduced.
- **Figure 2 (d)** Reproduced.
- **Figure 2 (e)** Reproduced.
- **Figure 2 (f)** Reproduced.
- **Figure 3 (a)** Reproduced.
- **Figure 3 (b)** Reproduced.
- **Figure 3 (c)** Reproduced.
- **Figure 3 (d)** Reproduced.
- **Figure 3 (e)** Reproduced.
- **Figure 3 (f)** Reproduced.

- **Table 1** Reproduced.
- **Table 2** Reproduced.
- **Table 3** Reproduced.
- **Table 4** Reproduced.

Results in the Annex

For the Appendix, we did not review every exhibit. Instead, we randomly selected 10 exhibits from the remaining datasets to assess the appendix. Our review was based on those 10 exhibits. Since they were chosen randomly, we are operating under the assumption that if all randomly selected exhibits are reproducible, then the rest should be as well. The seed used to generate the random selection was 787461, the Stata version used was 19.5, and the exhibits selected were: Table A1, Table A2, Table A3, Table A5, Table A6, Table A7, Table A8, Table A9, Figure B1, Table A11

- **Table A1** Reproduced.
- **Table A2** Reproduced.
- **Table A3** Reproduced.
- **Table A5** Reproduced.
- **Table A6** Reproduced.
- **Table A7** Reproduced.
- **Table A8** Reproduced.
- **Table A9** Reproduced.
- **Figure B1** Reproduced.
- **Table A11** Reproduced.

Reproduction Environment

Paper exhibits were reproduced on a computer with the following specifications:

- OS: Windows 11 Enterprise
- Processor: INTEL(R) XEON(R) PLATINUM 8562Y+ (2.80 GHz) (2 processors)
- Memory available: 32 GB
- Software version: Stata 19.5 MP