1 Instructions for Replication

These files are the code to replicate "Focused Interventions and Test Score Fade-Out" by Michael Gilraine and Jeffrey Penney. No data are provided as it is not publicly-available. This document describes the code (in Stata) that is available to replicate the paper. All files are available from the Review of Economics and Statistics Dataverse at https://dataverse.harvard.edu/dataverse/restat.

1.1 Obtaining the Data

Unfortunately, all of the data used for the paper are from the North Carolina Education Research Data Center (NCERDC) at Duke University. The URL for their website is: https://childandfamilypolicy.duke.edu/north-carolina-education-research-data/. The current (as of October 2022) requirements to obtain these data are as follows:[1]

1. Signed Data Use Agreement and confidentiality agreements for the Principal Investigator and all research assistants.
2. IRB approval from the Principal Investigator's institution.
3. Completed proposal for use of the data, including the data security plan and statement of benefits for the Data Center.
4. Agreement to give copies of all research products to the Data Center and to the Department of Public Instruction.
5. Compensation of the Data Center for access to the data and time spent on the project by Data Center personnel.
6. For students only - Faculty sponsorship of the proposal, indicated by a letter of reference.

The raw data files provided by NCERDC come in SAS form. These were converted to STATA (.dta) using **R** using the "read_sas" and "write_dta" commands (requires the "haven" package). Files names were kept the same and so the .dta files are just their .sas7bdat counterparts on the NCERDC server. All analysis was done on STATA 15.1.

[1] The procedures for obtaining the data are found here: https://childandfamilypolicy.duke.edu/wp-content/uploads/sites/2/2022/02/Procedures-for-Obtaining-Data_2.16.22.pdf

2 Analysis Dataset and .Do Files Explanation

All .do files have a number in front of them. If you run the do files in the sequence of the numbers all the results from the paper will be created (.do files with the same number but differentiated with a letter can be run in any sequence). We have placed the .do files into two folders: the "Code to Generate Analysis Dataset" are .do files that create data, while the ".Do Files" folder contains .do files that run the RD analysis or the structural model. We now go through what each .do file does and which results it generates.

2.1 Dataset Generating .Do Files

The dataset used for the main analysis is "student_mb_2009to2017.dta." The dataset is constructed using the NCERDC data and so cannot be shared. The .do file constructing this dataset is:

1. "1 - Data_get_2009_17.do": constructs the data set "student_mb_2009to2017.dta". This file is constructed using the masterbuild files provided by the NCERDC.

Additional datasets are then constructed for each of the long-run outcomes that we analyze (see Table 3). These files and the corresponding .do file to create them are:

1. "3a - RD_long_run_sat_gen.do": constructs the data set "sat_2009_2019.dta". This file is constructed using the SAT files provided by the NCERDC.

2. "3b - RD_long_run_psat_gen.do": constructs the data set "psat_2013_2019.dta". This file is constructed using the PSAT files provided by the NCERDC.

3. "3c - RD_long_run_grad_gen.do": constructs the data set "long_run_grad_data.dta". This file is constructed using the graduation files provided by the NCERDC.

2.2 Analysis .Do Files

The following .do files generate all results in the paper. We state each analysis .do file and what Table/Figure it creates here:

1. "2a - RD_results_Math.do": Creates all the RD results when the running variable is math test scores relative to the retest threshold. Includes results in Figures 2, C.1, C.2, and C.4 and Tables 2, A.1, A.2, and C.1. The only input required is the main analysis data set "student_mb_2009to2017.dta."

2. "2b - RD_results_English.do": Reruns all the above analysis, but for when the running variable is English test scores relative to the retest threshold. Creates the English results in all the aforementioned Tables/Figures and also creates Figures A.2 and C.3.

3. "4 - RD_long_run_Table3.do": This code creates Table 3. It uses the main analysis dataset as well as the long-run outcome datasets.

4. "5a - Structural_Gain_Math.do": Creates the structural results for math, creating Figure 1(a) and Table 1. The only input required is the main analysis data set "student_mb_2009to2017.dta."

5. "5b - Structural_Gain_English.do": Reruns the structural model for English, creating Figure 1(b) and Table 1.

6. "5a.2 - Structural_Gain_Math_bootstrap" and "5b.2 - Structural_Gain_English_bootstrap": These are sister files to the structural files above. All these files do is place the structural model in a bootstrap loop which is used to generate standard errors. Note: As these files must run through a substantial number of loops, they take a **long** time to run. On my server, each .do file took approximately 1 month to run.