Instructions for replicating the results from Yasuyuki Sawada, Minhaj Mahmud, Mai Seki, and Hikaru Kawarazaki (2023)

This read-me file contains instructions for replicating all tables and figures from Yasuyuki Sawada, Minhaj Mahmud, Mai Seki, and Hikaru Kawarazaki (2023), "Fighting the Learning Crisis in Developing Countries: A Randomized Experiment of Self-Learning at the Right Level," *Economic Development and Cultural Change*.

We, the authors of the manuscript, certify that we have legitimate access to and permission to use the data used in this manuscript.

1. Software requirements

These files require the following ado files:

- rwolf (version 2.2.0)
- leebounds (version 1.5)
- parmest (version 16.0)
- cgmwildboot (version of 2015-03-10)
- psmatch2 (version 4.0.12)
- distplot (version of 2.1.0)

We modified the last command, cgmwildboot, as cgmwildboot_2 to report the p-value as e(Pvals). This ado file is located in the ado file folder; other features of cgmwildboot have not been changed. Also, the command rwolf has been updated, so rwolf (version 2.2.0) is saved in the ado file as the name of "rwolf_2."

All code is run on Stata 16.1 MP. By commenting out "set processor 1" in "0_all.do", the code can be executed in Stata other than MP.

2. Folder structure

The replication package contains five folders. The "ado" folder contains the aforementioned "cgmwildboot_2" ado files. The "data" folder contains the original data. We describe the data set below.

"do-files" contains all the do-files needed to replicate the results. All the file names indicate output by the code.
"outputs" folder contains the output tables and figures created by the above code.
"temp" folder contains the temporally files created.

After opening the zip file, you need to make empty folders "outputs" and "temp" as a subfolder of "replication_package." Additionally, you need to make another "temp" folder under the folder of "data."

3. Data sets

The data folder contains the following data set.

Name	Data description
main_data.dta	The main data set used for all tables and most figures. It includes baseline and endline outcome variables, treatment status, as well as demographic information such as gender and age.
psc_score.dta	This data set includes the results of Primary School Certificate, which are necessary to examine the long-term impact.
learning_curve.dta	This is a daily record of the worksheet. For more information, see Kawarazaki et al. (2023).
Kumon_CBA_IRR_v11.xls	This is a record of cost and benefits.

The "temp" folder in the "data" folder stores the intermediate files created the code.

4. Replication instructions

- Please create subfolders, following Section 2 above.
- "0_all.do" contains all the path settings and random seeds, needed to run the other codes.
- Please change the global path in the line 19 of "0_all.do" to the location of your own folder to run the code. Specifically, for example, you would put this replication folder in the location of your own computer as "/Users/username/drive/replication/replication_package".
  Then you set the global path as:
  ```
  global path_data "/Users/username/drive/replication/replication_package"
  ```
  This way, Stata can recognise the location of the files in your own computer.
- The code corresponding to each table or figure is named accordingly. Once the cleaning code is run, each code can be run independently.

5. References

Kawarazaki, Hikaru, Minhaj Mahmud, Yasuyuki Sawada, and Mai Seki. "Haste Makes No Waste: Positive Peer Effects of Classroom Speed Competition on Learning." *Oxford Bulletin of Economics and Statistics* (2023). https://doi.org/10.1111/obes.12545.