Overview

The code in this replication package focuses on primary data collected in South Punjab, Pakistan from children, observations in pre-primary and Grade 1 classrooms and interviews with teachers. It also includes additional secondary data analysis on UNCIEF MICS data and Pakistan census data for background statistics on the region. The code uses Stata and R to run all analysis and provide all figures and tables in the paper. R file `Primary R Script.R` is the main R script that should be run first. `Primary dofile.do` is the main Stata .do file that should be run after the R code. The replicator should expect all code to execute within 15 minutes.

Because we cannot share data with identifiers with the replicator, the code has been modified to omit the stage at which PII is stripped and unique IDs are created. The code to do this (`load_clean.R`) is left in the files for reference but is commented out in the Primary R script.

Data Availability and Provenance Statements

Statement about Rights

- ☑ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

- ☐ I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

Summary of Availability

- ☐ All data **are** publicly available.

- ☐ Some data **cannot be made** publicly available.

- ☑ **No data can be made** publicly available.

Details on each Data Source

Data.Name	Data.Files	Location	Provided	Citation
"Punjab MICS"	`ch_hl_hh.dta`	rawdata/	FALSE	Bureau of Statistics Punjab, Planning & Development Board, Government of the Punjab &

				UNICEF. (2018)
"PHCIP School Data"	`"PHCIP 3400 Schools with Baseline Sample 900 Schools Tagged.xlsx"`	rawdata/	FALSE	Pakistan Bureau of Statistics. (2017)
"South Punjab Census"	`"dataset for HCI Districts having primary portion K-5 only.xlsx"`	rawdata/	FALSE	Pakistan Bureau of Statistics. (2017)
"South Punjab Primary Data Collection"	`"teach_labelled.dta"` `"Pak_Punjab2022_clean.dta"` `"Module 6 -teacher dataset_d.dta"`	rawdata/	FALSE	N/A

The data for this project are confidential and cannot be shared.

Details on data files

Data.Files	Description
`ch_hl_hh.dta`	This is data from the Multiple Cluster Indicator Survey (MICS) 2017-2018 conducted in Punjab. This is publicly available data at the MICS website after registration (**https://www.unicef.org/pakistan/reports/multiple-indicator-cluster-survey-2017-18-punjab**) but was provided to the authors merged after use in Tomlinson et al. (2023). This is used for Table 1.
`"dataset for HCI Districts having primary portion K-5 only.xlsx"`	An extract from Punjab census data listing all primary-level schools in Punjab. This is used to create Table 2.
`"PHCIP 3400 Schools with Baseline Sample 900 Schools Tagged.xlsx"`	A full list of all the 3400 schools part of the PCHIP and 900 schools indicated for the sample along with additional school-level information. This is used to create Table 2.
`"Pak_Punjab2022_clean.dta"`	Child-level data on the Anchor Items for Measurement of Child Development – Direct Assessment (AIM-ECD). This includes the item-level data along with some basic demographic descriptors. Observations are identified with the combination of `merge_class_id` (class unique identifier) and `childid` (child identifier). This data is used for all plots and tables about the AIM-ECD DA.
`"teach_labelled.dta"`	Classroom observations using the Teach ECE tool. This includes the item-level data of the observations and some

	observations of classroom resources. Observations are identified by `merge_class_id` (class unique identifier). This data is used for plots and tables about the Teach ECE.
`"Module 6 -teacher dataset_d.dta"`	Teacher-level from a teacher survey. This includes the item-level data along with some basic demographic descriptors. Observations are identified by `merge_class_id` (class unique identifier). This data is used for plots and tables regarding teacher characteristics.

Tables and Figures

The below table outlines exactly where in the code each table and figure (which uses reproducible data) is generated.

Table/Figure number	File	Line
Table 1	analysis.do	76
Table 2	analysis.R	90
Table 6	analysis.R	179
Table 7	analysis.R	199
Figure 5	analysis.R	232
Figure 6	analysis.R	258
Figure 7	analysis.R	295
Figure 8	analysis.R	330
Figure 9	analysis.R	371, 389
Figure 10	analysis.R	412
Figure 11	analysis.R	441
Figure 12	analysis.R	469

Figure 13	analysis.R	505
Figure 14	analysis.R	545
Figure 15	analysis.R	571
Figure 16	analysis.R	597
Figure 17	analysis.R	623
Figure 18	analysis.R	648
Figure 19	analysis.R	675
Figure 20	analysis.R	700
Table 8	analysis.do	136
Figure 21	analysis.do	182
Figure 22	analysis.do	213
Figure 23	analysis.do	249
Appendix C	analysis.R	524
Appendix D	analysis.do	162, 164

Instructions to run the code

R 4.2.1 and Stata 17 were used to reproduce the results. "Primary R Script.R" is the main R script that should be run first. Primary dofile.do is the main Stata .do file that should be run after the R code.

If the main script is opened using RStudio, it will automatically detect missing libraries to run the code. The libraries used in this reproducibility package are: pacman, psych, readxl, writexl, lavaan, tidyverse, and readstata13. Ado files of Stata dependencies are included in the folder "dofiles/ado". These are loaded by the main the main do-file and do not need to be installed by the replicators.

An optional environment file for *renv* is also included. This will allow replicators to recreate the exact R environment in which results are guaranteed to be reproduced. To enable the optional environment, follow these steps:

1. Open the RStudio project file "RR_PAK_2024_120.Rproj" using RStudio,

2. Run the code: *renv::restore()* . If prompted to install any package, reply Yes.

Version 1.0.3 of renv was used to create the environment, though other version might also work. The installation of RTools might be needed for installing some packages depending on the replicator's operating system.

References

Bureau of Statistics Punjab, Planning & Development Board, Government of the Punjab & UNICEF. (2018). Multiple Indicator Cluster Survey Punjab, 2017-18, Survey Findings Report. Bureau of Statistics Punjab, Planning & Development Board, Government of the Punjab. https://mics-surveys-prod.s3.amazonaws.com/MICS6/South%20Asia/Pakistan%20%28Punjab%29/2017-2018/Survey%20findings/MICS%20SFR_Final_English.pdf

Pakistan Bureau of Statistics. (2017). *Population Census 2017: Population by single year age, sex and rural/urban*. Pakistan Bureau of Statistics. https://www.pbs.gov.pk/sites/default/files/population/2017/tables/pakistan/Table04n.pdf

Tomlinson, H. et al., 2023. Fostering Early Childhood Development (English), World Bank Group. United States of America. Retrieved from https://policycommons.net/artifacts/3742149/fostering-early-childhood-development-english/4548056/ on 15 Apr 2024. CID: 20.500.12592/1mb9sh.