

README for the Reproducibility Package

Quality Matters: Estimating Economic Impacts of Improving Education Quality in the Kingdom of Saudi Arabia

1. Overview

This reproducibility package contains the data, code, model workbook, outputs, and manuscript required to reproduce the empirical estimations and simulation exhibits reported in the paper.

The empirical analysis uses an unbalanced country panel and estimates the relationship between education quality, measured using Harmonized Learning Outcomes (HLO) in mathematics, and GDP per capita growth. The simulation components apply the estimated coefficients to Saudi Arabia, GCC, and MENA reform scenarios.

2. Package contents

Path	Purpose
README.docx	Replication instructions, output mapping, and Data Availability Statement.
LICENSE.txt	Package-level reuse statement and pointer to original data licenses.
data/HLO Original Data.dta	Analysis-ready Stata dataset used for Appendix Tables B1-B3.
data/data_sources.csv	Machine-readable Data Availability Statement.
data/data_dictionary.csv	Variable names, Stata labels, data types, and nonmissing counts.
code/stata/estimations.do	Stata master file for Appendix Tables B1-B3.
code/r/WB Projection method.R	R master file for Appendix Table B4.
outputs/excel/gcc_mena_ksa_2025.xlsx	Excel simulation model for Appendix Tables B5 and C1-C3; includes a Sources sheet.
outputs/tables/Appendix_Table_B4.csv	Pre-generated wide Appendix Table B4 output.
outputs/tables/Appendix_Table_B4_long.csv	Pre-generated long-form Appendix Table B4 output.
manuscript/	Final manuscript supplied for verification.

3. Start-to-finish replication order

1. Run code/stata/estimations.do to reproduce Appendix Tables B1-B3. The script writes logs/stata.log. If asdoc is installed, it also writes formatted system-GMM output to outputs/tables/Appendix_Table_B2.doc.
2. Run code/r/WB Projection method.R from the package root to reproduce Appendix Table B4 and write outputs/tables/Appendix_Table_B4.csv and outputs/tables/Appendix_Table_B4_long.csv.
3. Open outputs/excel/gcc_mena_ksa_2025.xlsx, keep calculation set to Automatic, and review the specified sheets/ranges to reproduce Appendix Tables B5 and C1-C3.

4. Instructions for replicators

Step 0: Folder paths

Run all scripts from the package root. If not possible, edit only the root path line near the top of each script.

Stata: in code/stata/estimations.do, edit global ROOT "." to the absolute path of the package root.

R: in code/r/WB Projection method.R, edit root <- normalizePath(".", mustWork = FALSE) to the absolute path of the package root.

Step 1: Reproduce Appendix Tables B1-B3 (Stata)

Open Stata and run: do code/stata/estimations.do

Input: data/HLO Original Data.dta

The do-file declares a panel with 3-year spacing using xtset country year, delta(3). The six-year HLO lag is represented by L2.HLO.

Expected output: logs/stata.log. Formatted asdoc output for Appendix Table B2 is produced only if asdoc is installed; otherwise the model output is still captured in the log.

Step 2: Reproduce Appendix Table B4 (R)

From the package root, run: Rscript "code/r/WB Projection method.R"

No external data file is required for this step. Parameters are embedded in the script and documented in the Data Availability Statement.

Expected outputs: outputs/tables/Appendix_Table_B4.csv and outputs/tables/Appendix_Table_B4_long.csv.

Step 3: Reproduce Appendix Tables B5 and C1-C3 (Excel)

Open outputs/excel/gcc_mena_ksa_2025.xlsx in Microsoft Excel. Set calculation to Automatic and recalculate formulas.

Appendix Table C1: sheet "0.06_bez izraela i iranu"; column blocks G-L (+25), M-R (+50), and S-X (+100).

Appendix Table C2: sheet "0.06+45 every 5years"; block "Instant increasing the average achievement by 180 points", columns Y-AD.

Appendix Table C3: sheet "0.06+45 every 5years"; columns G-L for linear, M-R for backloaded, and S-X for frontloaded.

Appendix Table B5: sheet "0.06+45 every 5years". Upper bounds use coefficient 0.06; lower bounds are $\text{Upper} * (0.009 / 0.06) = \text{Upper} * 0.15$.

5. Mapping of outputs to files

Manuscript exhibit	Script/workbook	Input(s)	Output / where to verify
Appendix Table B1	code/stata/estimations.do	data/HLO Original Data.dta	logs/stata.log; estimates stored as B1_1-B1_3
Appendix Table B2	code/stata/estimations.do	data/HLO Original Data.dta	logs/stata.log; if asdoc installed, outputs/tables/Appendix_Table_B2.doc
Appendix Table B3	code/stata/estimations.do	data/HLO Original Data.dta	logs/stata.log; estimates stored as B3_1-B3_4
Appendix Table B4	code/r/WB Projection method.R	Embedded simulation parameters	outputs/tables/Appendix_Table_B4.csv and Appendix_Table_B4_long.csv
Appendix Table B5	outputs/excel/gcc_mena_ksa_2025.xlsx	WDI 2021 GDP levels and model parameters	Workbook sheet "0.06+45 every 5years"; rows 129-150 and columns G, M, S with lower bound = upper * 0.15
Appendix Tables C1-C3	outputs/excel/gcc_mena_ksa_2025.xlsx	WDI 2021 GDP levels and model parameters	Workbook sheets "0.06_bez izraela i iranu" and "0.06+45 every 5years"

6. Computational requirements and runtime

Stata: Stata 14 or later. asdoc is optional and used only for formatted output; install with `ssc install asdoc` if desired.

R: R 4.0 or later. The revised R script uses base R only and does not require external R packages.

Excel: Microsoft Excel 2019 or Microsoft 365 recommended. Use Automatic calculation and recalculate the workbook before checking results.

Expected runtime: Stata estimations < 5 minutes; R simulations < 1 minute; Excel recalculation < 1 minute.

7. Data Availability Statement

Summary: all external raw data sources used in this package are publicly available. The package includes the analysis-ready Stata dataset and the Excel simulation workbook used for verification. The raw public sources, direct URLs or access instructions, access dates, citations, and licenses/access conditions are listed below and in `data/data_sources.csv`.

Rights statement: the authors certify that they had legitimate access to and permission to use the data used in the manuscript. Third-party datasets remain subject to their original licenses and terms of use.

Note on HLO Original Data.dta: this is an analysis-ready dataset assembled by the authors from the public sources listed below. No confidential or proprietary data are included.

Package file(s)	Dataset or model input	Source/provider	URL or access instructions	Access date	Citation	License / access conditions
data/HLO Original Data.dta	Harmonized Learning Outcomes (HLO), mathematics; country, assessment, year, subject, grade, score, and HLO variables	World Bank, Harmonized Learning Outcomes (HLO) Database / Data360	https://data360files.worldbank.org/data360-data/data/WB_HLO/WB_HLO_WIDEF.csv ; metadata: https://datacatalog.worldbank.org/search/dataset/0038001/harmonized-learning-outcomes-hlo-database	2 October 2022	World Bank. Harmonized Learning Outcomes (HLO) Database. Data360/Data Catalog. Accessed 2 October 2022. Related publication: Angrist, Noam, Simeon Djankov, Pinelopi K. Goldberg, and Harry A. Patrinos. 2021. "Measuring human capital using global learning data." <i>Nature</i> 592: 403-408. https://doi.org/10.1038/s41586-021-03323-7 .	Public. Creative Commons Attribution 4.0 International (CC BY 4.0). Redistribution permitted with attribution.
data/HLO Original Data.dta	World Development Indicators controls: GDP per capita growth; GDP per capita, PPP constant 2017 international dollars; gross capital formation; labor force participation; life expectancy; expense; trade; FDI net inflows; natural resource rents; government expenditure on education	World Bank, World Development Indicators (WDI)	Indicator API links: NY.GDP.PCAP.KD.ZG, NY.GDP.PCAP.PP.KD, NE.GDI.TOTL.ZS, SL.TLF.ACTI.ZS, SP.DYN.LE00.IN, GC.XPN.TOTL.GD.ZS, NE.TRD.GNFS.ZS, BX.KLT.DINV.WD.GD.ZS, NY.GDP.TOTL.RT.ZS, SE.XPD.TOTL.GD.ZS. Use https://api.worldbank.org/v2/en/indicator/{indicator}?downloadformat=excel .	2 October 2022	World Bank. World Development Indicators. Washington, DC: World Bank. Accessed 2 October 2022.	Public. World Bank datasets are generally licensed under CC BY 4.0 unless an indicator is specifically labeled otherwise. Redistribution permitted with attribution and World Bank dataset terms.
data/HLO Original Data.dta	Expected years of	World Bank WDI;	https://api.worldbank.org/v2/en/indicator/SE.SCH.LIFE?downloadformat	2 October	World Bank. World Development Indicators, indicator	Public.

	schooling (variable: expectedyears of schooling; WDI code SE.SCH.LIFE)	original source: UNESCO Institute for Statistics	=excel ; metadata: https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SE.SCH.LIFE	2022	SE.SCH.LIFE, Expected years of schooling. Original source: UNESCO Institute for Statistics. Accessed 2 October 2022.	World Bank dataset terms/CC BY 4.0 unless indicator metadata specifies otherwise; source is UNESCO Institute for Statistics.
data/HLO Original Data.dta	Mean years of schooling (ISCED 1 or higher), population 25+ years, both sexes; used for average/years-of-schooling control where applicable	UNESCO Institute for Statistics (UIS), UIS Data Browser, Other Policy Relevant Indicators - Education (OPRI)	https://databrowser.uis.unesco.org/browser/EDUCATION/UIS-EducationOPRI/mys ; OPRI bulk/data hub: https://data.unesco.org/explore/dataset/uis006/export/	2 October 2022	UNESCO Institute for Statistics (UIS). Mean years of schooling (ISCED 1 or higher), population 25+ years, both sexes. Other Policy Relevant Indicators - Education, UIS Data Browser. Source: UNESCO Institute for Statistics (UIS), https://databrowser.uis.unesco.org/browser/EDUCATION/UIS-EducationOPRI/mys , extracted 2 October 2022.	Public. UIS Data Browser content is licensed under Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). Redistribution on permitted with attribution and date of extraction.
outputs/excel/gcc_mena_ksa_2025.xlsx	2021 country GDP levels in current US dollars, used as baseline GDP values in USD billions for Saudi Arabia, GCC, and MENA country simulations	World Bank, World Development Indicators (WDI)	https://api.worldbank.org/v2/en/indicator/NY.GDP.MKTP.CD?downloadformat=excel	2 October 2022	World Bank. World Development Indicators, indicator NY.GDP.MKTP.CD, GDP (current US\$). Washington, DC: World Bank. Accessed 2 October 2022.	Public. World Bank datasets are generally licensed under CC BY 4.0 unless specifically labeled otherwise. Redistribution on permitted with attribution and World Bank dataset terms.
outputs/excel/gcc_mena_ksa_2025.xlsx	SAR/USD conversion used for Saudi GDP in SAR billions where shown in the workbook; workbook applies SAR 3.75 per US\$ 1	Saudi Central Bank (SAMA), Exchange Rate page	https://www.sama.gov.sa/en-US/Currency/FinExc/Pages/Currency.aspx	June 2026 for documentation; the rate is a fixed model conversion used in the workbook	Saudi Central Bank (SAMA). Exchange Rate. Accessed June 2026.	Public website information. Used only to document the fixed conversion factor; not redistributed as a dataset.
outputs/excel/gcc_mena_ksa_2025.xlsx; code/r/WB Projection method.R	Education-quality improvement scenarios: +25, +50, +100, +180	Authors' simulation assumptions, based on the model and scenarios described	No external data download. Values are embedded in the Excel workbook and R script; see manuscript Table 1 and Appendix Tables B4-C3.	Not applicable	Authors' own simulation assumptions and calculations. See Table 1 and Appendix Tables B4-C3 in the manuscript.	Authors' work included for reproducibility

	one-time improvements; linear, backloaded, and frontloaded paths over 20 years; target score and baseline score parameters embedded in workbook/script	in the manuscript				ity verification. No third-party data license applies to these assumptions.
outputs/excel/gcc_mena_ksa_2025.xlsx; code/stata/estimations.do	Growth coefficients used in simulations: 0.009, 0.034, and 0.06	Authors' econometric estimates generated from data/HLO Original Data.dta using code/stata/estimations.do	Generated by the included Stata code. Coefficients correspond to pooled OLS, Arellano-Bond, and Arellano-Bover/Blundell-Bond estimates in Appendix Tables B1-B3.	Not applicable	Authors' own estimations. See Appendix Tables B1-B3 in the manuscript.	Generated output included for reproducibility verification.
outputs/excel/gcc_mena_ksa_2025.xlsx	Country group definitions for MENA and GCC totals	Authors' classification used in the workbook	No external data download. MENA countries in reported Arab-country totals: Algeria, Bahrain, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libya, Morocco, Oman, Qatar, Syria, Tunisia, United Arab Emirates, Yemen, Saudi Arabia. GCC countries: Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, United Arab Emirates.	Not applicable	Authors' own country grouping for simulation reporting.	Authors' classification included for reproducibility verification.

8. Notes for the reviewer

This README points to the actual Stata and R scripts included in the package. The Stata script is code/stata/estimations.do and the R script is code/r/WB Projection method.R.

The Excel model is documented as a simulation workbook rather than a raw-data file. Its key external input is 2021 country GDP in current US dollars from WDI; the remaining values are model parameters, authors' assumptions, or coefficients generated by the Stata estimations.

The Excel workbook includes a new Sources sheet summarizing raw inputs, model parameters, citations, access dates, and licenses/access conditions.