

# What I really want- policy maker views on education in East Asia Pacific

Reproducibility Package for “What I really want - policy maker views on education in East Asia Pacific”, by Noah Yarrow, Paul Cahu, Mary Breeding and Rythia Afkar.

## Overview

The code in this replication packages constructs the tables and figures as well as the statistics presented in the paper (2023) using stata. One main file runs all the code. It is called “WhatIreallyWant\_replication\_script.do” and it is located in the root directory of the package. The replicator should expect the code to run for only a few minutes.

## Package structure

The root directory contains 4 types of files.

1. A stata script to run the code.
2. Some excel file that contains input data as well as outputs for the paper.
3. A pdf file of the paper denoted.
4. Several stata .dta files that contains the input data.

## License

The data are licensed under a Creative Commons/CC-BY-4.0 License.

## Summary of Availability

- All data are publicly available.
- Some data cannot be made publicly available.
- No data can be made publicly available.

## Data sources

Name	File	Source	Citation
World Development Indicators - GDP	External_data.xlsx > GDP	Downloaded from: <a href="https://data.worldbank.org/">https://data.worldbank.org/</a>	World Bank (2023). World Development Indicators
World Development Indicators – Total population	External_data.xlsx > pop-WDI	Downloaded from: <a href="https://data.worldbank.org/">https://data.worldbank.org/</a>	World Bank (2023). World Development Indicators
Learning Poverty for EAP countries	External_data.xlsx > Fig2 Figures_and_tables.xlsx > Fig 23	Downloaded from : <a href="https://www.worldbank.org/en/topic/education/publication/state-of-global-learning-poverty">https://www.worldbank.org/en/topic/education/publication/state-of-global-learning-poverty</a> For Pacific countries, see Figure A.4 in : <a href="https://documents1.worldbank.org/curated/en/099051523071076581/pdf/P1777340ebd5720e50b2bd023a031d7f955.pdf">https://documents1.worldbank.org/curated/en/099051523071076581/pdf/P1777340ebd5720e50b2bd023a031d7f955.pdf</a>	World Bank (2022). The State of Global Learning Poverty: 2022 Update. World Bank (2023) Raising Pasifika Strengthening Government Finances to Enhance Human Capital in the Pacific
Education returns by level of education	External_data.xlsx > Patrinos&Montenegro	Computed from sources downloaded from: <a href="https://datacatalogfiles.worldbank.org/ddh-published/0059847/DR0083929/Comparable%20Returns%20to%20Education%20Database.xlsx?versionId=2021-11-08T20:12:50.5747747Z">https://datacatalogfiles.worldbank.org/ddh-published/0059847/DR0083929/Comparable%20Returns%20to%20Education%20Database.xlsx?versionId=2021-11-08T20:12:50.5747747Z</a>	Montenegro, C.E. and Patrinos, H.A. (2021), "A data set of comparable estimates of the private rate of return to schooling in the world, 1970–2014", International Journal of Manpower, Vol. ahead-of-print No. ahead-of-print. <a href="https://doi.org/10.1108/IJM-03-2021-0184">https://doi.org/10.1108/IJM-03-2021-0184</a>
Women, Business and the Law	External_data.xlsx > Fig13	Available at: <a href="https://wbl.worldbank.org/en/wbl-data">https://wbl.worldbank.org/en/wbl-data</a> Included in the package as: "WBL-1971-2023-Dataset.xlsx"	World Bank. 2023. Women, Business and the Law 2023. Washington, DC: World Bank."
Budget share of education spending	External_data.xlsx > Fig15	Downloaded from: <a href="http://sdg4-data.uis.unesco.org/">http://sdg4-data.uis.unesco.org/</a>	UNESCO 2023. Sustainable Development Goal version 4.
Government expenditure on education as a percentage of GDP (%)	External_data.xlsx > UIS-sdg4	Downloaded from: <a href="http://sdg4-data.uis.unesco.org/">http://sdg4-data.uis.unesco.org/</a>	UNESCO. SDG 4 September 2023 Release
ISO country codes	External_data.xlsx > isocode	Country codes to merge UNESCO and WDI. Manual input from: <a href="https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3#Officially_assigned_code_elements">https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3#Officially_assigned_code_elements</a>	
CGD survey 2020	CGD_2020.dta	CGD (2020). Can be downloaded at: <a href="https://www.cgdev.org/sites/default/files/education-policy-maker-preferences-codes-and-data.zip">https://www.cgdev.org/sites/default/files/education-policy-maker-preferences-codes-and-data.zip</a>	Lee Crawford, Susannah Hares, Ana Minardi, Justin Sandefur, 2021. "Understanding Education Policy Preferences: Survey Experiments with Policymakers in 35 Developing Countries." CGD Working Paper 596. Washington, DC: Center for Global Development.
CGD survey 2022	CGD_2022.dta	CGD (2022).	
SEA-PLM 2019	SEA-PLM2019_students.dta	Raw data can be downloaded at: <a href="https://www.seaplum.org/index.php/seaplum-2019/datasets-and-questionnaires-2019">https://www.seaplum.org/index.php/seaplum-2019/datasets-and-questionnaires-2019</a>	UNICEF 2020, SEA-PLM 2019 assessment

## Data Description

The main data used in the paper is taken from a couple of surveys realized by the Center for Global Development (CGD) in 2020 and 2022. All the statistics derived from both these surveys are directly generated by the script. These are:

1. "CGD\_2020.dta"
2. "CDG\_2022.dta"
3. "CGD\_DCE\_2022.dta"
4. "CGD\_DCE\_joint.dta"
5. SEA-PLM2019\_students.dta

Some other statistics have been used for the paper and are contained in a dedicated excel file denoted:

6. "External\_data.xlsx"

There are other stata data files which are created by the script:

7. "CDG\_2022\_for\_figures.dta"
8. "perchildspending.dta"
9. "countries.dta"
10. "lpguess.dta"
11. "temp.dta"

Finally, there is one file that is not directly used by the script but has been used to compile the excel sheet denoted "External\_data.xlsx":

12. "WBL-1971-2023-Dataset.xlsx"

The files "CGD\_2020.dta" and "CGD\_2022.dta" contain the microdata from both surveys of policymakers.

The file "CGD\_DCE\_2022.dta" contains the microdata of the Discrete Experiment where respondents were randomly presented to a set of options that they had to choose from.

The file "CGD\_DCE\_joint.dta" allows matching the "CGD\_DCE\_2022.dta" with the respondent's characteristics and their other answers to the survey.

The file "SEA-PLM2019\_students.dta" allows computing figures of Table 2.

The file "External\_data.xlsx" contains all the other data necessary to run the analysis. It has been compiled from external sources, see table above. The sheet "Patrinos&Montenegro" is the only one which contains data that were not directly downloaded.

## Instruction to replicators

- Clone the repository to your local machine.
- Customize line 26 of the script to machine the place on your machine where you put the package.
- Run the script by using stata, version 14 or higher.
- If you want to replicate the data sources compiled in the "external\_data.xlsx" file, you will find the links to download the primary data in the table above.

- In the sheet “Patrinos&Montenegro”, the columns represent the ratios of the mean labour income for workers that differ only by their education level.
- The latest survey for each country was used unless specified in the excel sheet.
- The formula used is simply using the average Mincer coefficient specified in the database and a conventional length of schooling of 6 years for secondary and 4 years for tertiary in accordance with the typical Mincer model.
- Figure 5/14 represent the mean value and confidence intervals of the probabilities of respondents to have selected a given investment project based on its type. It is based on probit modeling. According to such model, the probability of selection of a project  $j$  by an official  $i$  of the region  $k$  is modelled as follows, where  $\Phi(\cdot)$  is the cumulative standard normal distribution,  $x_{ij}$  a vector of characteristics for the officials and the project,  $\alpha, \beta$  and  $\gamma$ , vectors of estimates and  $type = j$  and  $region = k$  a set of binary variables indicating respectively the type of the project and whether the official is from the EAP region.

$$p_{ijk} = \Phi(\beta x_i + \alpha_j(type = j) + \gamma_j(type = j) \times (region = k)) \quad (Eq. 1)$$

- To compare the selection probabilities by types of projects  $j$  regardless of the characteristics of the officials  $j$  and the other characteristics of the project (such as level of spending and implementation delays), we consider a “typical” official and a “typical” type of project, denote by the vector  $\bar{x}$ .
- This allows computing the probabilities for only the types and the region as:
 
$$p_{ik} = \Phi(\beta \bar{x} + \alpha_j(type = j) + \gamma_j(type = j) \times (region = k)) \quad (Eq. 2)$$
- The values of the multipliers  $\beta, \alpha$  and  $\gamma$  are given by the probit regression.
- The features of the “typical” project and officials are specified in the excel sheet in notes. We consider that the official is 50% female, 42% high-level, that the project will have rapid results at 50% and that the level of spending is intermediate, the type of choice is also set at 50% and the officials are set at 50% working in MoF and 50% in MoE.
- The (Eq. 2) gives the formula for the mean probabilities, computed in line 24, columns G to T.
- To get the confidence intervals, computed in lines 22 and 23 (same columns), we add to the quantity within the normal distribution the 5% values of the parameters  $\alpha_j$  and  $\gamma_j$  by adding or subtracting two standard deviations of these estimated parameters.

## Software

Stata 14 and higher.

## Other sources

The source of Figure 23 is figure 3.11 in World Bank (2023) Raising Pasifika Strengthening Government Finances to Enhance Human Capital in the Pacific, which can be downloaded at:

<https://documents1.worldbank.org/curated/en/099051523071076581/pdf/P1777340ebd5720e50b2bd023a031d7f955.pdf>

The figures in table 1 can be extracted from the Learning Poverty Update 2022 and from Table 3.1 in World Bank (2023) Raising Pasifika Strengthening Government Finances to Enhance Human Capital in the Pacific (see below for link).