

Functional specification:

Background:

Equity is a fundamental value and guiding principle of education policy, but it is not necessarily actualized in education systems around the world. In order to investigate educational equity in the difference of academic performance between gender groups, females and males, we collect international education data from the Internet. PISA is an international assessment of 15-year-old students' capabilities in reading literacy, mathematics literacy, and science literacy as well as measuring a wider range of factors including students' interest, attitudes, and motivation. It is distributed across over 60 countries, and student background questionnaires and teacher questionnaires are collected as part of PISA assessment. Student scores vary greatly for different student subpopulations within a jurisdiction – e.g., gender. In this project, we are specifically interested in describing and explaining gender differences in students' performances scores across countries. We hope to identify countries that have “equitable” schooling outcomes and their characteristics in education systems, including student and school culture, teacher educational practices, and societal factors. That said, we caution against simplistic international comparisons of country performances using correlates of higher performance in higher ranking countries for projecting causal generalizations and statements.

User profile:

Public users such as parents, teachers, and students

To use PISA-WB analytic tool, public users such as parents, teachers, and students should be able to have familiarity with operating graphical user interfaces (point, click, drop-down filters for interactive graphs).

International organizations such as World Bank, UNESCO, International Association for Evaluation of Educational Achievement (IEA), Educational Testing Service (ETS)

Test experts, teachers, researchers specified in educational measurement and policies, and enthusiastic parents who are data-driven decision makers are welcome to use our analysis tool. Users are expected to know the basic commands of computing, especially in Jupyter Notebook, and make right responses for commands and interpretation. They should have some encounter with Jupyter Notebook which is optimal to enable sharing of notebooks with other colleagues for more in-depth analysis. The benefit of using Jupyter is that users can see the code as well as the actual results of running the code. It is especially helpful for the users to see some text/instructions mixed with the code in a notebook.

Education researchers and methodology experts

Education researchers and measurement experts should have in-depth knowledge of interpreting and using modeling results from hierarchical linear modeling. They should master a certain level of data analysis in python language and common statistic packages.

Data sources:

Datasets are downloadable from the following link:

PISA international: <http://www.oecd.org/pisa/data/2015database/>

World Bank: <http://datatopics.worldbank.org/education/indicators>

Original data sources:

- Student (STU): Student questionnaire data file in the most recent PISA 2015 dataset.
- School (SCH): School questionnaire data file in the most recent PISA 2015 dataset.
- Teacher (TEA): Teacher questionnaire data file in the most recent PISA 2015 dataset.
- Economic social-cultural status (ESCS): Rescaled indices of economic, social and cultural status for countries for trend analysis. Available in PISA 2012 dataset.
- Country equity indicator (CEI): Education equity indicators at the country level from World Bank data sets including Multiple Indicator Cluster Survey and Urban Informal Settlement Survey in the year 2015. Selected series are education equality indicators such as total net attendance rate, lower secondary, gender parity index (GPI).

Three aspects of the design of PISA need careful attention in any analysis. The first stems from the sample design. Schools and students had unequal but known probabilities of selection. As a consequence, to generalize to the population sampled, analyses will need to apply the sampling weights provided in the file.

The second aspect to be considered also stems from the sampling design and bears on the calculation of standard errors. Since the sample design is a two-stage, stratified cluster Design, we decide to do hierarchical linear modeling to account for the clusterness of the data.

Thirdly, PISA has special designs for student scores. PISA datasets include sets of five “plausible values” for each student for each overall subject area score and each subscale score. The plausible values are intended to represent the estimated distribution of scores of students similar to the given student in terms of responses to the assessment items and background questionnaire items. What this means for analyses is that, when it comes to statistics involving the achievement scores we use the approach of doing it five times, once for each plausible value, and then average the results.

A metadata of PISA student variable names is shown in the following:

	NAME	VARLABEL	TYPE	FORMAT	VARNUM	MINMAX	VAL
1	CNTRYID	Country Identifier	NUM	3.0	1	8 – 974	
2	CNT	Country code 3-character	CHAR	\$3.	2		
3	CNTSCHID	Intl. School ID	NUM	8.0	3	800001 – 97400238	
4	CNTSTUID	Intl. Student ID	NUM	8.0	4	800001 – 97407472	
5	CYC	PISA Assessment Cycle (2 digits + 2 character Assess...	CHAR	\$4.	5		
6	Region	Region	NUM	5.0	7	800 – 97000	
7	STRATUM	Stratum ID 7-character (cnt + region ID + original str...	CHAR	\$7.	8		
8	SUBNATIO	Adjudicated sub-region code 7-digit code (3-digit co...	CHAR	\$7.	9		
9	LANGTEST_COG	Language of Assessment	NUM	3.0	22	113 – 622	
10	BOOKID	Form Identifier	NUM	3.0	25	1 – 99	
11	ST123Q01NA	<this academic year>: My parents are interested in m...	NUM	1.0	65	1 – 4	
12	ST123Q02NA	<this academic year>: My parents support my educat...	NUM	1.0	66	1 – 4	
13	ST123Q03NA	<this academic year>: My parents support me when I...	NUM	1.0	67	1 – 4	

Use cases:

Objective: Describing and explaining gender differences in students' performances across countries

Public users such as parents, teachers and students

To use PISA-WB analytic tool, public users such as parents, teachers and students would be able to see patterns visually described in gender performances in subjects across countries, and also navigate the interactive graph to see descriptive scores for specific attributes relating to student, teacher, school characteristics across countries. For example, they can see Science subject scores in male and females across countries. In some way they will see the measures of different educational systems as well in mean, 10 percentile, 90 percentile, etc. for visually interacting with project results to understand international comparison of education outcomes and systems.

Policy makers, researchers and experts on methodology

Policy makers and education researchers across the world can compare country performances and country characteristics identifying attributes of its education systems for generating insights about effective policy and practice strategies that are associated with learning outcomes.

To use this analytic tool, researchers and measurement experts should have in-depth knowledge of interpreting and using modeling results from hierarchical linear modeling. They should master certain level of data analysis in python language and common statistic packages. With this tool, they can rank-order countries by using either graphical user interfaces, or computing commands to generate tables for data retrieval. They will be able to see the exemplar analysis and visualization, understanding the project result, and perform further analysis aligned with similar inquiries, such as educational equity in Social Economic Status (SES), social mobility, racial, ethnic, and national groups.