Project Name: PISA

Shuai Shao

POLI271 Project

Department of Psychology

Programme for International Student Assessment (PISA) 2018 Revisited:

The Application of Maximum Likelihood Estimation in Psychometrics

Introduction

Evidence-based educational assessments are of growing interest to educators and policymakers. First launched in 1997, Program for International Student Assessment (PISA), a large-scale, policy-oriented, and international project that measures 15-year-old students' mathematical, cognitive, and reading abilities has a profound impact on international education research (Morgan & Shahjahan, 2014; Khorramdel et al., 2017).

Researchers have applied various technics to understanding the PISA test results, such as Item Response Theory (IRT). In psychometrics, IRT is a classic model to evaluate examinees' reactions and item quality in multiple-choice tests (Warm, 1989). More recently, an IRT-based framework was proposed to explore how examinee's effort changes over time (Debeer & Janssen, 2013). Practically, Maximum Likelihood Estimation (MLE) is a handy and widely-used method to estimate parameters in IRT (Johnson, 2007).

In general, PISA tests provide insightful analyses and information on students' academic performance and daily-life problem-solving abilities in different countries. Understanding PISA test data and results have promoted education reform in several countries such as Brazil and Germany and will contribute to a more efficient and advanced education system.

Methods

Dataset The current project uses PISA 2018 dataset retrieved from PISA website, conducted by Organisation for Economic Co-operation and Development (OECD, the organizer of PISA assessment) in 2018 (PISA, 2018). PISA collected multilevel data from students, teachers, principals, and generated multilevel reports. Public and researchers are especially interested in country-level analyses to understand the universality and specificity of education.

Measurement In each dimension, students read several materials and complete related multiple choices. In the reading section, students are evaluated in relation to different reading texts, processes, and contexts. Mathematical skills are assessed by mathematical contents, processes, and situations. Scientific concepts, processes, and situations are tested in the science part. OECD provides a detailed manual to instruct scorers to give credits to students' responses (i.e., no credit, partial credit, or full credit).

Analysis Plan In this project, I will replicate and improve previous work on the basis of item response theory (Johnson, 2007; Monseur et al., 2011). Specifically, using appropriate maximum likelihood estimation (MLE) models, I will compute 1) the item parameters of questions and 2) students' proficiency. Further, I am interested in cross-country and cross-area (mathematics, reading, and science) comparisons and the implication of education policies.

Descriptive Analysis

More than 70 countries or regions attended the 2018 assessment, including developed countries such as the United States and Canada and developing countries such as China (data only collected in Beijing, Shanghai, Jiangsu, and Zhejiang) and Brazil. Among all the countries, China ranked highest on all three subscales (555 on reading, 591 on mathematics, and 590).

Country-level data descriptions are displayed as follows, generated by R Studio (codes can be provided by request).

It is also worth noticing that students in the United States performed above average in reading (505, compared to the world average of 487) and science (502, compared to the world average of 489). However, Americans achieved lower scores (478) than the world average (489).

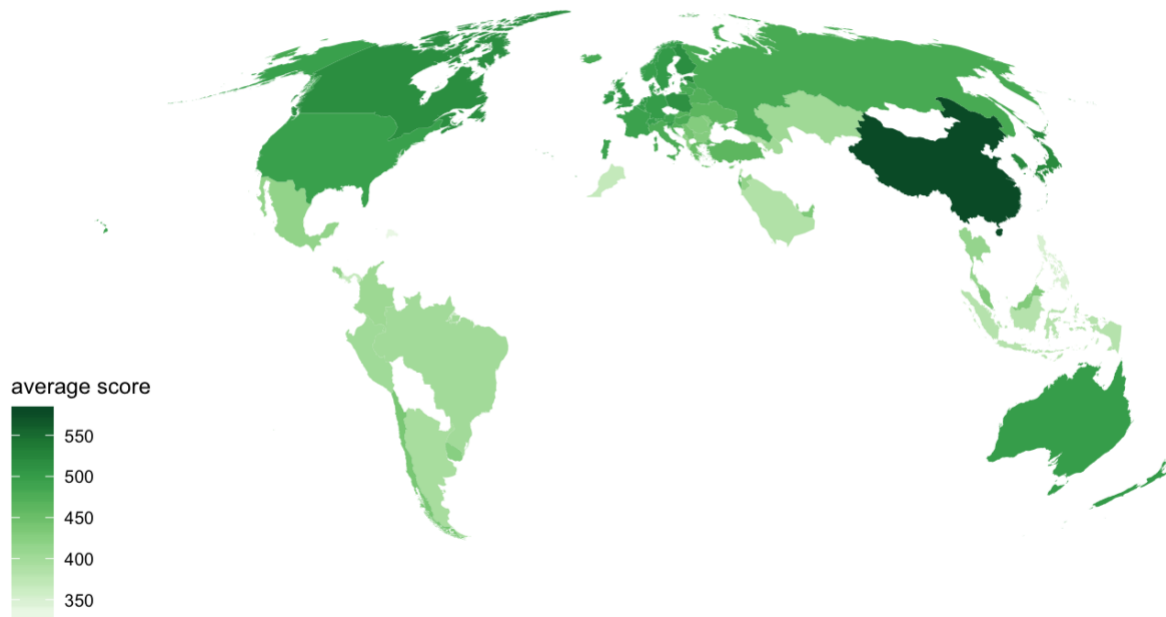Further analyses will be presented in the future.

Figure 1. Country-Level PISA Average Scores

PISA Test Mathematics Scores, 2018
data sources: OECD

math score

500

400
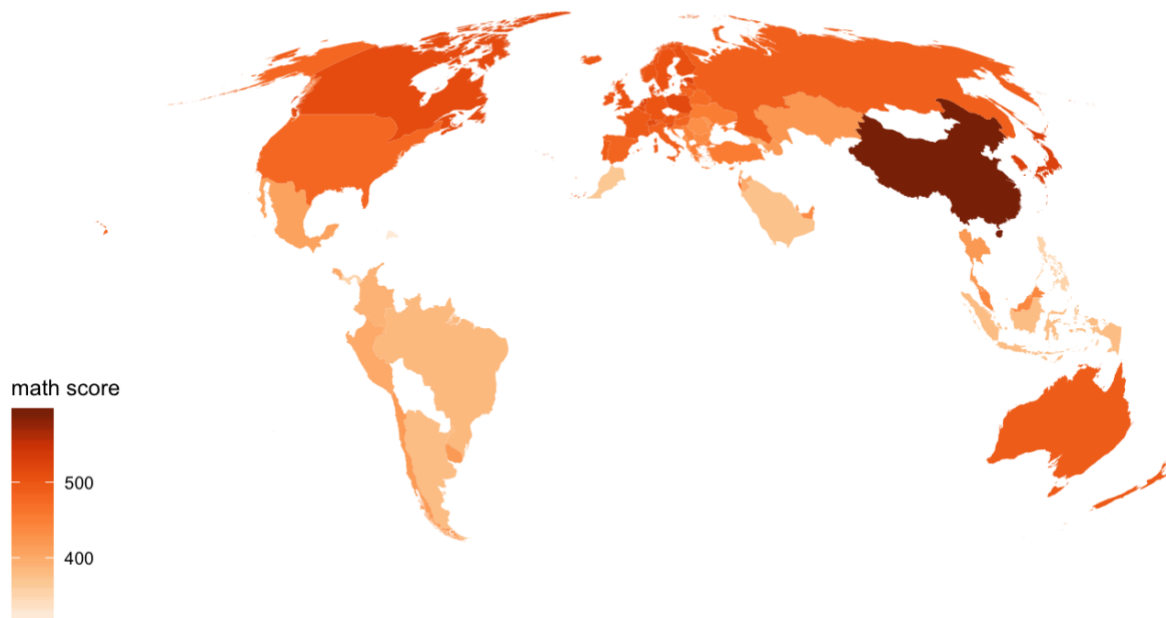
Figure 2. Country-Level PISA Math Scores
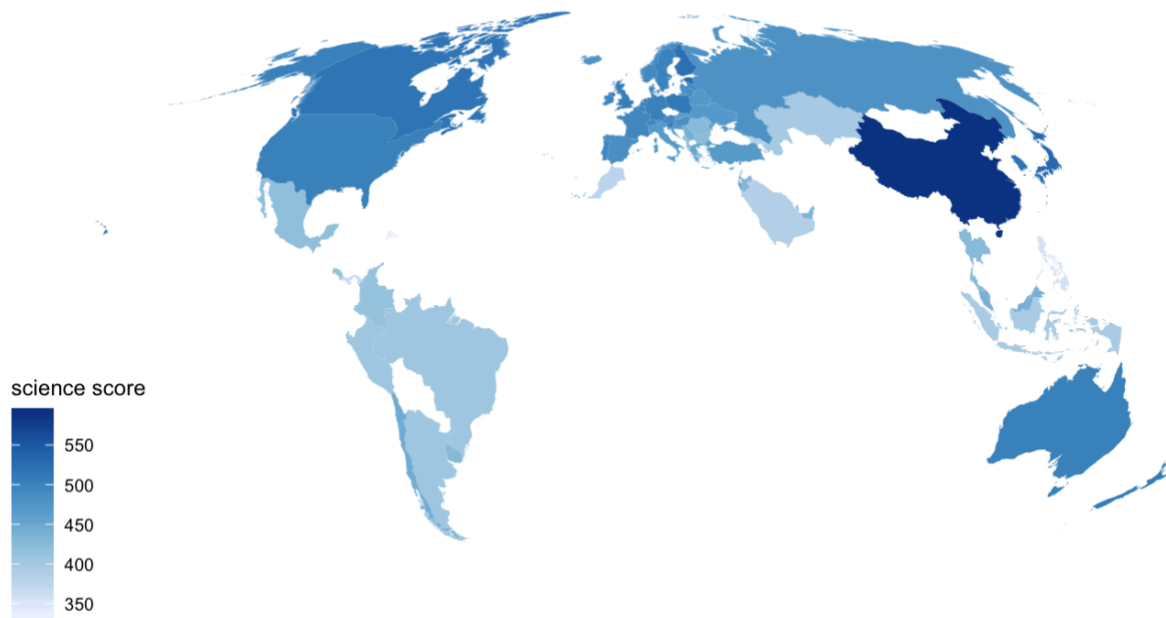
PISA Test Science Scores, 2018
data sources: OECD

science score

550

500

450

400

350

Figure 3. Country-Level PISA Science Scores

PISA Test Reading Scores, 2018
data sources: OECD

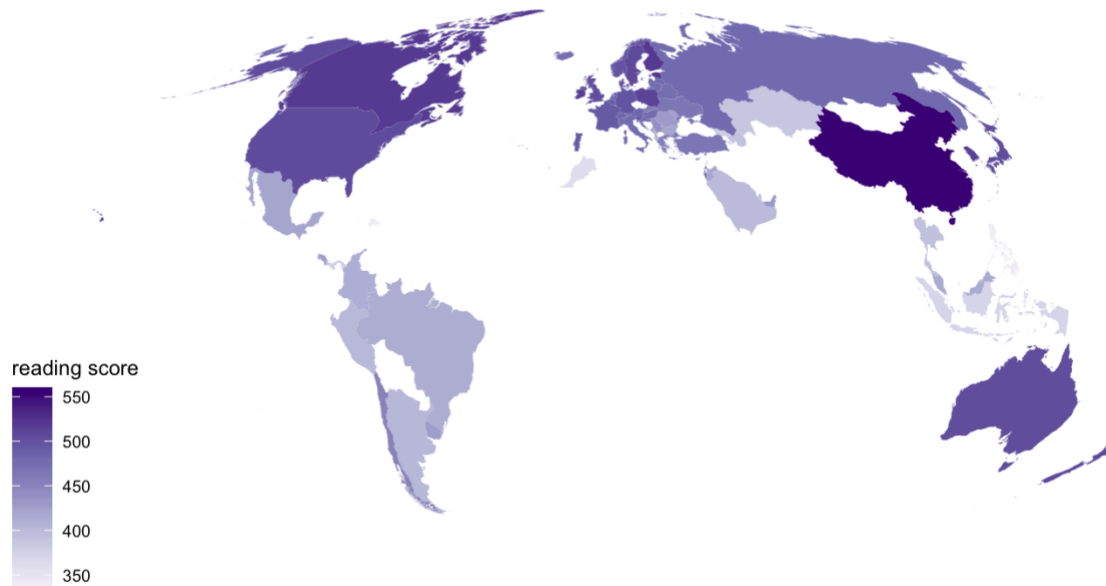

reading score

550
500
450
400
350

Figure 4. Country-Level PISA Reading Scores

References

Debeer, D., & Janssen, R. (2013). Modeling item-position effects within an IRT framework. *Journal of Educational Measurement*. https://doi.org/10.1111/jedm.12009

Johnson, M. S. (2007). Marginal maximum likelihood estimation of item response models in R. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v020.i10

Khorramdel, L., von Davier, M., Bertling, J., Roberts, R., & Kyllonen, P. (2017). Recent IRT Approaches to Test and Correct for Response Styles in PISA Background Questionnaire Data: A Feasibility Study. *Psychological Test and Assessment Modeling*.

Monseur, C., Baye, A., Lafontaine, D., & Quittre, V. (2011). PISA test format assessment and the local independence assumption. *IERI Monograph Series*.

Morgan, C., & Shahjahan, R. A. (2014). The legitimation of OECD's global educational governance: examining PISA and AHELO test production. *Comparative Education*. https://doi.org/10.1080/03050068.2013.834559

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*. https://doi.org/10.1007/BF02294627