An Inquiry into Girls' Achievement in Science: Horizontal Inequalities, Socio-economic Gaps and Disparities Using 50 Years of International Large-scale Assessment Data

Yuriko Kirilovna Sosa Paredes

Centre for Educational Measurement, University of Oslo

Supervisor: Prof. Dr. Sigrid Blömeke

Co-supervisor: Associate Prof. Björn Andersson

January 24, 2022

An Inquiry into Girls' Achievement in Science: Horizontal Inequalities, Socio-economic Gaps and Disparities Using 50 Years of International Large-scale Assessment Data

This research project aims to measure and analyse the Science achievement gaps of fourth-grade girls assessed in International Large-Scale Assessments (ILSA), using 50 years of TIMSS data.

Therefore, three significant steps are proposed: (1) construct a common Science scale, using a new approach such as latent regression models focus on the evaluation of (non)invariance measurement, (2) propose and build theoretically and methodologically a socioeconomic status scale (SES), using multilevel models; and (3) use the information from (1) and (2) to estimate disparities from the perspective of horizontal inequalities, using the confluence of two characteristics: SES and gender.

Background

Since the beginning of the 20th century, international comparative large-scale studies have become one of the most prolific areas in education. The data generated have the potential to provide nuanced snapshots of characteristics of different school systems; that is, that data should be used with caution and awareness of their context (Johansson, 2016). The International Association for the Evaluation of Educational Achievement (IEA) is currently one of the influential organizations conducting such studies through several ILSAs.

The ILSAs have the advantage of providing mainly two key elements, the information on the students' achievement in a specific area (e.g., reading, mathematics, science) and the knowledge about constructs associated with performance. These can be background variables, scales, and indicators related to students, parents, and school characteristics, and they allow examining drivers and mechanisms responsible for student performance. Results contribute to the understanding base of educational science and the design of evidence-based educational policies (Majoros, 2021).

Significantly studies have utilized the richness of the many ILSAs carried out

since 1964 for examining cross-national or country-specific research questions. In particular, for this inquiry, some of these investigations can be classified into two major contributions. The first one refers to linking cognitive outcomes over time to observe the long-term cross-national trends. Moreover, the second contribution presents the study of global trends focused on socioeconomic status; that is, the analysis of achievement gaps changes due to this characteristic and the possible factors that influence this variation over time.

1. Majoros (2021) and Strietholt and Rosén (2016); propose and compare approaches to link recent and older studies conducted by IEA onto a common scale to study long-term trends within and across studies. They perform a concurrent calibration employing Rasch and Item Response Theory (IRT) models to achieve this aim, considering the four criteria proposed by Kolen and Brennan (2014) to evaluate the degree of similarity between tests: Inferences, constructs, populations, and measurement characteristics/conditions. They focused on mathematics (FIMS, SIMS, TIMSS) for ILSAs administered between 1964 and 2015; and reading (RCS, RLS, PIRLS) for ILSAs administered between 1970 and 2011, respectively. However, one of the limitations of these studies is that they did not consider sample characteristics in person scoring, like latent regression with plausible values does. As reported in the documentation of the TIMSS studies, to enhance the reliability of the student scores, their scaling approach uses a process known as conditioning (von Davier, 2020).

   Another vital aspect to consider in this new approach is the focus on assessing measurement invariance not only between groups but over time (Desa et al., 2019).

2. Chmielewski (2019) describes the global trend and cross-national variation in SES achievement gaps, as well as identifies the possible causes of this variation. The author studies achievement gaps between the 90th and 10th percentiles of the SES composite index: parents' education, parents' occupation, and the number of books in the home. Although, she used standardized measures in order to evaluate the differences.

The data was pooled from 30 IEA and OECD ILSAs of reading, mathematics, and science over 50 years.

Based on this work, Salmela-Aro and Chmielewski (2019) examined the trends in socioeconomic inequality of student outcomes in Finland. Notably, they focus on SES achievement gaps, defined as disparities in academic achievement between students from low- and high-parental education backgrounds. The average performance has decreased in PISA, but inequality in achievement appears to be increasing.

In both investigations, one possibility to expand the study of the disparities in academic performance derived from the SES is further to specify the gaps within the group of girls. That is, reviewing inequalities through categorizations such as low- and high-SES within girls and the variables that could influence these changes over time. After a theoretical and methodological review, it would be possible if a common SES over time is made.

Objectives

Drawing on the aforementioned work, this research project has three main objectives, of which three papers will be proposed:

- **Study 1. Establishing comparable trend measures of science in fourth-grader students assessed by IEA.**

  Over time, the study of long-term trends can be expanded by including student performance in science. For this, the ILSAs data will be used, specifically, those corresponding to science competence. In particular, the considered assessments are FISS (1970/1971), SISS (1983/1984), and TIMSS (1995-2019) of fourth-graders students.

  In this analysis, the considerations about linking and calibration of the previous studies will also be taken into account. The research proposal includes testing the fourth scenario of parameter calibration (Majoros, 2021): latent regression model with plausible values using contextual variables of the students. This approach

can be viewed as an imputation model for the unobserved proficiency distribution that aims at obtaining unbiased group-level proficiency distributions (von Davier, 2020).

Mainly, this research will appraise the (non)invariance measurement of the items across the groups and over the 50 years of ILSAs, using multilevel and multigroup models (Brown, 2015).

- **Study 2. Establishing comparable trend measures of Socioeconomic scale in ILSAs, using data from TIMSS over 50 years.**

  Although the social sciences have extensively studied socioeconomic status, many authors agree on the complexity of arriving at a single definition, mainly because of its multidimensional nature and close relationship with the context of the reality it aims to assess or measure (Minedu, 2018).

  Moreover, the degree of linear or non-linear relationship found with academic performance will depend on how SES is constructed (Minedu, 2018).

  Therefore, a theoretical review is proposed based on the research carried out and taking into account Chmielewski's approach. From this, we will seek to construct a single scale comparable over time, using multilevel models.

- **Study 3. Describing the global trend and cross-national variation in horizontal inequality gaps (SES and gender), and also investigating the possible causes of this variation.**

  Stewart in 2008 defines horizontal inequalities as "inequalities in economic, social or political dimensions or cultural status between culturally defined groups". We treat "culturally defined groups" generally equivalent to "ethnic" groups.

  However, it can also consider ascriptive groups in a broader sense to include gender (Canelas & Gisselquist, 2019). After studies 1 and 2, it will propose to examine the achievement gap defined by the confluence of two characteristics: SES and gender. Therefore, from studying horizontal inequalities, expanding the definition and implications of what is understood as inequality. It would not be presented only between differentiated groups (e.g., boys vs girls, vulnerable groups

vs privileged groups) but also within each of these groups, showing specific dynamics in the gaps for each defined context.

Research questions

1. Can the lack measurement invariance be confirmed, across the objective groups, and over time, in older and recent ILSAs applied in fourth-graders students, conducted by IEA?

2. How the countries' average achievement scores in science has changed over time, using older and recent ILSAs applied in fourth-grader students, conducted by IEA?

3. Considering the new scale compared to the "official" science scale from TIMSS between 1995 and 2019, are there important differences in the trend of average achievement scores?

4. Which is the best model for parameter estimating, between the traditional estimation only considering the responses to the items (e. g., Rasch, 2PL, and 3PL model) and the estimation using contextual variables of the students (latent regression with plausible values)?

5. How has the SES gaps achievement in science within the group of girls changed over time, using older and recent ILSAs conducted by IEA?

6. What are the most effective strategies to improve the study of horizontal inequalities in science education with ILSAs conducted by IEA?

Research method

Overall, the methodology proposed initially for this research plan will follow the subsequent steps:

1. **Data collection.** The necessary databases and documents are collected from the IEA data repository. Moreover, confidential documentation (e.g., books, instruments) will be provided especially for this project by the IEA.

2. **Data pre-processing.** It is necessary to start with the process of cleaning and consistency of the data, which implies, the revision of the variables through indicators: codes, labels, number of categories, types of variables. Afterwards, the unique matrix is prepared given all the information, missing data are identified, quantified, and explored. Finally, the relevant variables are selected for the analysis, including the number of countries in the investigation. It is relevant to mention that it requires a great power computational to do that.

3. **Identification of anchoring items.** Quantify the number of items in common between periods of tests.

4. **Concurrent calibration and linking over time.** Perform the linking of the measures considering the entire matrix response and background standard variables over time, in addition, to taking into account the considerations to evaluate the degree of similarity between the tests.

5. **Investigation of measurement invariance of the items in ILSAs.** It requires statistical modelling to inspect the measurement (non)invariance of the pertinent items across groups and over the 50 years of assessments.

6. **Construction of scales for the analysis of cross-national disparities of horizontal inequalities gaps over time (SES and gender)** At first, it is suggested to use the SES scale proposed by filtering the study countries suggested in the first objective. However, the relevance of a new SES configuration will be explored. Secondly, it is necessary to generate a new variable that mixes the categories “low SES and high SES” within the group of girls.

7. **Computing and interpretation of the achievement gaps and possible influential factors.** It is proposed to set a three-level multilevel model: gaps, study years, and countries; to assess dependence and appropriately analyze nested data.

Expected results

We will better understand the trend measures of science over time, the inspection of the variation or not of difficulty items measures in ILSAs over time, and the comparison between (selected) countries. Moreover, the analysis of variation in horizontal inequality gaps, related to SES and gender, and the possible causes. Our results should be communicated in conferences, workshop proceedings, and journal manuscripts. Furthermore, the developed scripts and databases can be stored in repositories that other researchers can access. The findings of this research project can be applied to other educational and economic research and can be used to make better decisions in social and public policies.

References

Brown, T. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). The Guilford Press.

Canelas, C., & Gisselquist, R. (2019). Horizontal Inequality and Data Challenges. *Soc Indic Res*, *143*, 157–172. https://doi.org/10.1007/s11205-018-1932-1

Chmielewski, A. (2019). The Global Increase in the Socioeconomic Achievement Gap, 1964 to 2015. *American Sociological Review*, *84*, 517–544. https://doi.org/10.1177/0003122419847165

Desa, D., van de Vijver, F. J. R., Carstens, R., & Schulz, W. (2019). Measurement invariance in international large-scale assessments : Integrating theory and method. In T. P. Johnson, B.-E. Pennell, I. A. Stoop, & B. Dorer (Eds.), *Advances in Comparative Survey Methods*. John Wiley; Sons Inc.

Johansson, S. (2016). International large-scale assessments what uses what consequences. *Educational Research*, *58*(2), 139–148. https://doi.org/10.1080/00131881.2016.1165559

Kolen, M., & Brennan, R. (2014). *Test equating, scaling and linking* (3rd ed.). Springer. https://doi.org/10.1007/978-1-4939-0317-7

Majoros, E. (2021). Measures of long-term trends in mathematics: Linking large-scale assessments over 50 years. *Educational Assessment, Evaluation, and Accountability*, *33*, 71–103. https://doi.org/10.1007/s11092-021-09353-z

Minedu. (2018). *Desafíos en la medición y el análisis del Estatus Socioeconomico de los estudiantes peruanos*. Springer.

Salmela-Aro, K., & Chmielewski, A. (2019). Socioeconomic Inequality and Student Outcomes in Finnish Schools. In Volante, Schnepf, Jerrim, & Klinger (Eds.), *Socioeconomic Inequality and Student Outcomes* (pp. 153–168). Springer. https://doi.org/10.1007/978-981-13-9863-6_9

Strietholt, R., & Rosén, M. (2016). Linking large-scale reading assessments: Measuring international trends over 40 years. *Measurement: Interdisciplinary Research and Perspectives*, *14*(1), 1–26. https://doi.org/10.1080/15366367.2015.1112711

von Davier, M. (2020). TIMSS 2019 Scaling Methodology. In M. Martin, M. von Davier, & I. Mullis (Eds.), *Methods and Procedures: TIMSS 2019 TECHNICAL REPORT*. TIMSS & PIRLS International Study Center.