

Amsterdam, June 2019

CALL FOR PROPOSALS

Call no. IEA 06/11-2019

IEA Research for Education Series: Assessing the Stability and Comparability of Constructs over Time in TIMSS

1. Introduction

The International Association for the Evaluation of Educational Achievement (IEA) invites proposals for a report based on secondary analysis of IEA data. The general theme for this report is to use IEA's Trends in International Mathematics and Science Study (TIMSS) data to propose and use methods to examine the stability and measurement comparability of student background constructs over time. The deliverable for this project will be an **80- to 150-page book**, to be published by Springer as part of IEA's Research for Education Series (see <https://link.springer.com/bookseries/14293>). The book will include, in addition to the main text: tables, graphs, cited references, and relevant appendix materials.

2. Study background and objectives

Much applied educational research that uses TIMSS data emphasizes mathematics and science achievement as the outcomes of interest, with a particular focus on understanding sources of variation in achievement (see, for example, Breton, 2014; Marsh et al., 2014; Rutkowski, Rutkowski, & Engel, 2013). In general, modeling strategies aim to estimate differences in means for selected contextual variables (e.g., sex, or immigration status). A second, connected interest lies in relating constructs, such as attitudes toward learning, classroom climate, or learning orientation, to achievement, either via standard correlations or through a modeling framework. Modeling and comparing achievement levels in many heterogeneous countries raises questions about the assumption of equivalent measures. Further, given that there are 20 years of data available, equivalence over time is also a matter for scrutiny.

An important criterion for using scale scores in cross-cultural research is that the latent variable of interest is understood and measured equivalently across all groups. This property is usually referred to as measurement invariance (Meredith, 1993), lack of bias (Lord, 1980), or an absence of differential item functioning (e.g., Hambleton & Rogers, 1989; Mellenbergh, 1982; Swaminathan & Rogers, 1990). Although not exhaustive, typical approaches for establishing measurement invariance at the scale level involve multiple-groups confirmatory factor analysis (MG-CFA; Jöreskog, 1971), the alignment method (Asparouhov & Muthén, 2014), and Bayesian invariance methods (Verhagen & Fox, 2013), among others. Often, prototypical examples of invariance analyses involve the comparison of just two groups (Bollen, 1989; Millsap, 2011, p. 3); however, in the context of TIMSS, comparisons involve many populations in terms of educational achievement, as well as other non-achievement domains (e.g., values, beliefs, and attitudes). This complexity poses challenges for making valid comparisons, as

measurement invariance should hold for each considered population, despite substantial language, cultural, and geographic differences across populations.

In spite of a growing body of methodological work around measurement invariance, few of these methods have been applied to TIMSS in a systematic way, either across countries or over time. To that end, IEA's TIMSS serves as a rich source of information for understanding issues around cross-cultural and temporal measurement invariance. The current call invites proposals from researchers investigating issues around measurement invariance in TIMSS background scales. Proposals can be empirical analyses on some selection of TIMSS background scales using a rigorously justified method, methodological (proposing and evaluating a new approach), or some combination. Preference will be given to proposals that include an analysis of invariance over time, as well as country. Further, proposing a rationale and method for grouping countries for comparison is desirable.

3. Possible topics

Interesting aspects that could be explored in the report include, but are not limited to:

- Among affective constructs, which are most invariant across culture and time?
- What differences emerge in a comparison between traditional multi-group CFA and the alignment method?
- Which, if any, scales produce cultural invariance patterns?
- By what criteria can we group and compare countries based on similar levels and types of scale invariance?

4. Data

IEA's TIMSS has assessed student achievement at grades four and eight in mathematics and science at four-year intervals since 1995. In 2015, nationally representative samples of students from 57 countries and seven benchmarking entities (regional jurisdictions of countries, such as states) participated in the study. In total, more than 600,000 students took part in TIMSS 2015. The various international reports on TIMSS provide a detailed overview of the study's results (see for example, Mullis et al., 2012, 2016).

TIMSS not only collects achievement data, but, additionally, the TIMSS database encompasses rich background information from students, their teachers, and principals. Since IEA studies are based on a curriculum model, where the intended, implemented and achieved curriculum are considered, TIMSS also includes information on the intended curriculum in participating countries from national centers and what is implemented in the schools from teachers, principals and students. Over the 20-year span of data collection, more than 2.5 million students from about 100,000 schools worldwide have participated, making the total database a rich source for examining system-level change over time.

5. General guidelines for proposal submission

Proposals must be submitted in English.

Please ensure the proposal demonstrates familiarity with the proposed research by including a sound literature review. Ensure that the contribution of the proposed thematic report to this literature is explicit, especially in terms of its potential to expand the current state of research and knowledge.

When preparing a proposal, please clearly specify the research relevance and the policy relevance of the research questions and methods selected. This specification needs to expand on and augment or complement the outline ideas set out in this call for proposals.

The proposal must furthermore describe the general analytical framework that will guide not only analyses of the IEA data but also interpretation of the results of those analyses. The description of the framework must be such that it clearly shows how the proposed analysis will address the policy-relevant research questions. The description should therefore identify:

- (i) which IEA data (study, questionnaire items, indices, or constructs from questionnaires) you intend to use,
- (ii) any non-IEA data sources that will be included, and
- (iii) any additional data collection that is deemed necessary (such as system-level characteristics).

Please make sure that a clear and complete description of the types of quantitative or qualitative analyses to be used is included. The degree to which the methods are suited to answer the research question is an important evaluation criteria for all proposals.

In addition, the proposal must include a detailed timeline for all analyses and report-writing activities, and a well-considered budget proposal to complete the project.

When developing timelines, assume a start date of 1 December 2019 and an end date of 1 January 2020; the final manuscript of the book must be supplied to IEA for print production by 1 September 2020. Although there may be a certain degree of flexibility in the timeline, it must make provision for (i) submission of a complete draft report by 1 June 2020 for review by IEA, and (ii) adequate time for subsequent revision and language editing of the report prior to print production. The corresponding author must be available for consultation with Springer Publishers during the print production period.

Budgets must include the expected number of work days needed to complete each activity related to the project and a total budget in euros or US dollars. The total budget should not exceed 25,000 euros.

The call is open to all researchers, excluding teams from IEA International Study Centers. For the latter, direct assignments are possible.

The proposal should be no more than 10 pages in length¹. Please also provide a short (500-word maximum) biographical note on each person in the team tendering for the project. Please highlight the relevance of each person's experience to the proposed activities.

IEA will review all proposals according to their methodological quality, research and policy relevance, and budget. All tenderers will be informed of the outcome of these deliberations by 1 November 2019.

¹ Times New Roman, Arial or similar, 12 point type, double spaced.

Proposals may be submitted by post, courier, or email.

The deadline for proposals is 1:00 p.m., Friday, 30 August 2019.

Send the proposal by post to:

International Association for the Evaluation of Educational Achievement, Keizersgracht 311, 1016 EE Amsterdam, The Netherlands

or by email to secretariat@iea.nl.

References

- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495–508. Retrieved from <https://doi.org/10.1080/10705511.2014.919210>
- Bollen, K. (1989). *Structural equations with latent variables*. New York: Wiley.
- Breton, T. R. (2014). Evidence that class size matters in 4th grade mathematics: An analysis of TIMSS 2007 data for Colombia. *International Journal of Educational Development*, 34, 51–57. Retrieved from <https://doi.org/10.1016/j.ijedudev.2013.04.003>
- Hambleton, R. K., & Rogers, H. J. (1989). Detecting potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods. *Applied Measurement in Education*, 2(4), 313–334. Retrieved from https://doi.org/10.1207/s15324818ame0204_4
- Jöreskog, K. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36(4), 409–426. Retrieved from <https://doi.org/10.1007/BF02291366>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New York: Routledge.
- Marsh, H. W., Abduljabbar, A. S., Parker, P. D., Morin, A. J. S., Abdelfattah, F., & Nagengast, B. (2014). The big-fish-little-pond effect in mathematics: A cross-cultural comparison of U.S. and Saudi Arabian TIMSS responses. *Journal of Cross-Cultural Psychology*, 45(5), 777–804. Retrieved from <https://doi.org/10.1177/0022022113519858>
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational and Behavioral Statistics*, 7(2), 105–118. <https://doi.org/10.3102/10769986007002105>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. Retrieved from <https://doi.org/10.1007/BF02294825>
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
- Rutkowski, L., Rutkowski, D., & Engel, L. (2013). Sharp contrasts at the boundaries: School violence and educational outcomes internationally. *Comparative Education Review*, 57(2), 232–259. Retrieved from <https://doi.org/10.1086/669120>
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361–370. Retrieved from <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Verhagen, A. J., & Fox, J. P. (2013). Bayesian tests of measurement invariance. *British Journal of Mathematical and Statistical Psychology*, 66(3), 383–401. Retrieved from <https://doi.org/10.1111/j.2044-8317.2012.02059.x>