www.etn-occam.eu
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

The impact of measurement error in background data on PISA scale scores

Laura Zieger
University College London
ESR 6

Motivation

www.etn-occam.eu
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

- OECD conducts the international large-scale assessment PISA every three years, since 2000
- 81 countries and economies have participated in PISA until 2015, especially most OECD member countries (OECD, 2016a)

→ Largest and most media-present test of this kind

www.etn-occam.eu
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

- Country‘s results in PISA get a lot of media coverage and sometimes can result in public uproars as well as policy reforms (e.g. Waldow, 2009; bpb, 2013)

→PISA results influence the teaching practice and education

→Valid, sensible and comparable scale scores are crucial

PhD project

www.etn-occam.eu
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

- Examine if PISA scale scores can be used for valid international comparisons
- Investigate if changes in methodology have an influence on comparability
- (If applicable) Quantify the impact of bias

Research issues

www.etn-occam.eu
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

1. *Replication of the scale scores*
   Shine light on used statistical methodology, changes in methodology and the exact set of used variables
2. *Identification of measurement error and invariance in background data*
   Fair computation can only be granted if influencing variables function the same in all countries

www.etn-occam.eu
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

3. *Investigation of the effects*
   If bias or incomparability is found, the effects of those on the PISA scale scores need to be identified

   → Simulation study to quantify the direction and extent of the effect as well as the affected group (e.g. all countries, Asian countries, Spanish speaking countries)

www.etn-occam.eu
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

Data

Focus on PISA 2012 and PISA 2015 data, because those are the two most recent studies and major changes in testing mode and statistical methodology took place in 2015 (OECD, 2016a; OECD, 2014a). Also including field trial data from 2015, if possible (OECD, 2016b).

Sample: PISA 2012:65 countries
PISA 2015: 72 countries (58 computer-based/ 14 paper-based)
→Overlap of 62 countries

Methodology

www.etn-occam.eu
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

Scale score computation

Two-step approach which is roughly the same in all cycles (OECD, 2016b; OECD, 2014b):

i. Estimation of item difficulties using item response theory framework (Baker & Kim, 2014)

ii. Scores are computed using a population model (mixture of IRT model and latent regression)

The exact models in (i) vary depending on the item and also between 2012 and 2015 (change in methodology).

www.etn-occam.eu
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

Differential item functioning

Evaluate if single items affect people of a certain group (e.g. country, sex, age) to answer in a certain way independently of their true ability or value (Magis et al., 2010). This can be done via multiple methods which can be IRT-based or parameter free.

Measurement invariance

Tests if a constructed measure (e.g. reading practice) functions the same for different groups. A multi-group confirmatory factor analysis is fitted and more and more restrictive layers are applied (Steenkamp & Baumgartner, 1998)

Scope

www.etn-occam.eu
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

→Validate achievement measures in the international context

→Shape the way people think about surveys and their comparability

→Results and statements drawn from PISA scores will be better informed and carefully considered

→Show survey makers the importance of background measures

ETN OCCAM
www.etn-occam.eu
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 765400.

Bibliography

- Baker, F. B., & Kim, S.-H. (2014). *Item response theory: Parameter estimation techniques*, 2nd edition, CRC, Boca Raton.
- bpb [Bundeszentrale für politische Bildung] (2013). PISA Studie: Deutsche Schüler über dem Durchschnitt. Retrieved from http://www.bpb.de/politik/hintergrund-aktuell/174546/pisa-studie
- Magis, D., Béland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning, *Behavior Research Methods*, 42(3), 847-863.
- OECD (2014a). *PISA 2012 results: What students know and can do – Student performance in mathematics, reading and science* (Volume I, Revised edition, February 2014), PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264201118-en
- OECD (2014b). *PISA 2012 technical report,* PISA, OECD Publishing, Paris.
- OECD (2016a). *PISA 2015 results (Volume I): Excellence and equity in education*, PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264266490-en
- OECD (2016b). *PISA 2015 technical report,* PISA, OECD Publishing, Paris.
- Steenkamp, J.-B., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research, *Journal of Consumer Research*, 25, 78-90.
- Waldow, F. (2009). What PISA did and did not do: Germany after the PISA shock, *European Education Research Journal,* 8(3), 476-483.