

From Numbers to Meaning Through Comparisons: An Introduction

Rolf Vegar Olsen & Sigrid Blömeke

Large-scale international assessments studies (ILSA) are regarded as important sources for monitoring educational quality in many countries. The results from the studies are frequently cited in policy documents and they are regularly used as warrants or rebuttals in political debates. Over the last 20 years or so, ILSAs have thus established themselves as powerful knowledge sources. In this series of four blog-posts we will present and discuss two of the main reasons for why these studies have gained this position: they support interpretations of findings from two comparative perspectives – comparisons between educational systems and comparison within one system over time. In the fourth and final post we will identify and discuss some of the challenges for these two comparative perspectives for future ILSAs.

Quantifying education

Measuring qualities in education is hard because the phenomena we are trying to capture are not directly observable. Usually we are interested in making inferences about psychological attributes of persons – how motivated are students to learn in school, how satisfied are teachers with their work environment, or how proficient are students in reading, to mention a few. These attributes cannot be observed or measured directly. We do not have a “reading-proficiency-ruler” or some electronic sensor, which immediately can report a value on a scale. Instead, we have to rely on indirect procedures where one or (usually) several observations of a person are used to establish numbers on a scale. Fortunately, substantive theory in collaboration with test theory (or “psychometrics”) provides us with powerful tools to develop reliable

Rolf Vegar Olsen

Centre for Educational Measurement, University of Oslo, e-mail: rolfvo@cemo.uio.no

Sigrid Blömeke

Centre for Educational Measurement, University of Oslo, e-mail: sigrid.blomeke@cemo.uio.no

and valid measures or indicators of such phenomena in the social and psychological realm.

Nevertheless, even if measures have documented good anchoring in substantive theory and psychometric quality, the numerical values themselves are usually hard to interpret, e.g. as exemplified through questions like:

- What is an acceptable level of achievement for a system?
- What value on a bullying scale represents a level of concern?
- At which values should we conclude that a regression coefficient, correlation coefficient, difference between two groups etc. is substantively meaningful?

International comparisons provide a frame of reference

Answers to questions like these are largely normative or political because we are not only lacking substantive theory for deriving answers of this kind, we also may disagree normatively: Someone can therefore just stipulate thresholds and try to add meaning to them. However, politically decided or stipulated thresholds also need backing, and ILSAs provide several ways of creating a more rational basis for such backings by allowing for comparisons.

The choice of a comparison, a criterion or norm, is not a neutral activity either. Rather, it forms the argumentative core or warrant for the interpretations to be made of the data. The following blog posts will discuss how comparisons with other countries aid interpretation of data from ILSAs, but also why such comparisons can be deceiving and methodologically flawed.

This text has been posted on the blog international-education.blog and it is available in different languages on international-education.blog

Olsen, R.V. & Blömeke, S. (2019). From Numbers to Meaning Through Comparisons: An Introduction. Retrieved from <https://international-education.blog/en/2019/10/10/from-numbers-to-meaning-through-comparisons-an-introduction/>

The blog International-Education.Blog is an outcome of the project OCCAM. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 76500.

