# From Numbers to Meaning through Comparisons within Countries over Time

Rolf Vegar Olsen & Sigrid Blömeke

Most of the international large-scale assessments are repeated in regular intervals. PISA is conducted every three-years, TIMSS every fourth year, and PIRLS is conducted with five-year intervals. Accordingly, this allows for comparisons within countries over time, with the objective to uncover patterns or trends and to predict future development. The achievement scores are linked over time by having a relatively large number of test questions that are repeated. This makes it possible to anchor subsequent test scores with the previous ones. In addition, sections of the same background questionnaires are repeated over time to also capture changes in the learning context, demographics etc.

The possibility to compare education systems over time overcomes one of the challenges of comparisons between countries (see our previous blog[1]): Comparisons over time are made within the same system, and hidden cultural or other non-observed differences can be regarded as controlled for.

Figure 1 provides an example where all average achievement results for Norway in PIRLS, PISA and TIMSS over 20 years from 1995 to 2015 are collected into one picture.

The three studies included cover three broadly defined domains. The figure illustrates the time series in reading with red lines, mathematics with blue lines and science with green lines. The figure also captures three different populations. The trend lines for the 15-year-olds are shown by solid lines, 8th graders with dashed lines and 4th graders with dotted lines. All the three studies report the achievement

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

[1] https://international-education.blog/en/from-numbers-to-meaning-through-comparisons-between-countries/
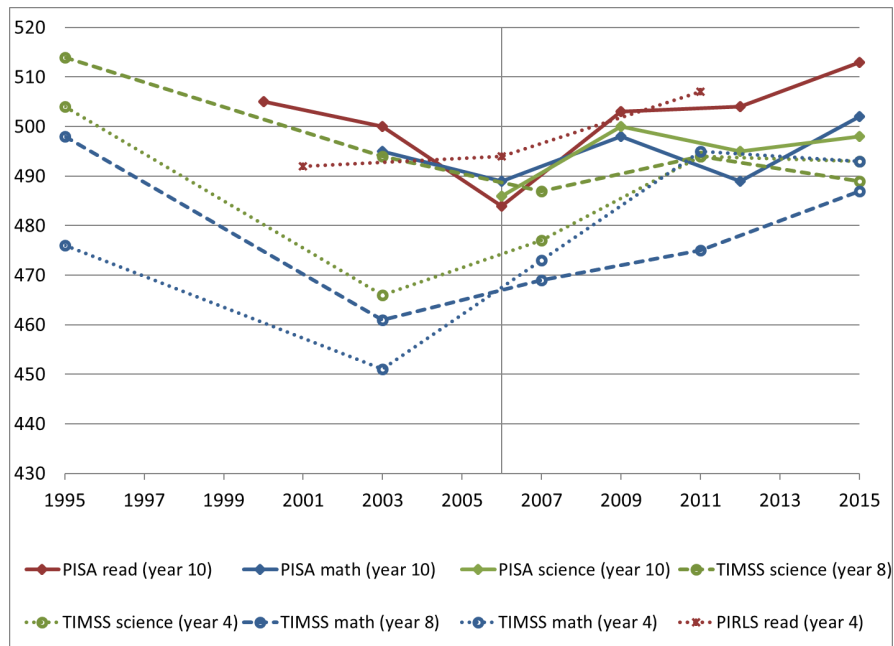
**Fig. 1** Achievement scores for Norwegian students for three ILSAs, in three domains for three student populations in the period 1995-2015. Take care not to interpret differences between the lines – they are largely meaningless (see text).

results on a scale where the international average in the first year of the series is set to 500 (and one standard deviation is set to 100). And here is the obvious weakness of the figure: Even if all the studies make use of what appears to be the same scale, direct comparisons cannot be made across the lines in the figure. The studies are not formally linked with each other, and the international average reflects a different composition of countries for each of the studies (e.g. OECD countries participate in PISA but only some of them in TIMSS and PIRLS). Nevertheless, a more holistic interpretation of features in the figure reveals a rather consistent feature of Norwegian students' development in the period:

- There was a huge decline in performance in the first half of the period, no matter which domain or student age we are talking about. For some subjects and studies the decline was close to 40 points on the scale (or 0.4 of a standard deviation). Another way of stating the same finding is that students who started their schooling in the mid- to late-1990ies for some reasons performed much worse than the previous cohorts of students.

- There is an almost equally strong trend of improvement in the last half of the 20-year period–despite a doubling of the number of students with immigrant backgrounds during this period. In particular, the improvement for the 4th grade

students largely make up for the decline in the early period, but the 8th grade students are still lagging a little bit behind in 2015 compared to 1995.

This figure, and the complexities of results they represent, were used in an evaluation of an educational reform in Norway. This reform was labeled as the "Knowledge Promotion" to reflect the fact that improving students' performance across all ages and many domains by strengthening basic knowledge acquisition and clearly defining learning outcomes was one of the major ambitions of the new policy. A large research-based evaluation was organized, but none of these studies could address the issue of whether or not this major ambition of the reform had been realized. Fortunately, data from the international studies were available and could be used to describe how students learning outcomes had changed in the past 20 years. A more detailed analysis of one of these changes (TIMSS 8th grade mathematics from 2003 to 2015) showed that the most important factors related to the positive change was an improved learning environment and school climate.

The above example illustrates the potential usefulness of the time series reported by the international large-scale studies. Not surprisingly, the trend series are increasingly emphasized in the national and international reporting from the studies. Furthermore, the time series design opens up for other ways of analyzing data than merely describing trends or patterns within countries. Features at the system level can be studied by so called differences-in-differences analyses, where changes in a predictor (e.g. a change in classroom size) are related to changes in an outcome (e.g. mathematics achievement) at the country level. Such analyses benefit from the same methodological advantages like panel analyses where individuals are observed repeatedly.

Olsen, R. V., & Blömeke, S. (2019). From numbers to meaning through comparisons within countries over time. Retrieved from https://international-education.blog/en/from-numbers-to-meaning-through-comparisons-within-countries-over-time