TIMSS, SACMEQ and PIRLS
Data issues

Martin Gustafsson
22 January 2019

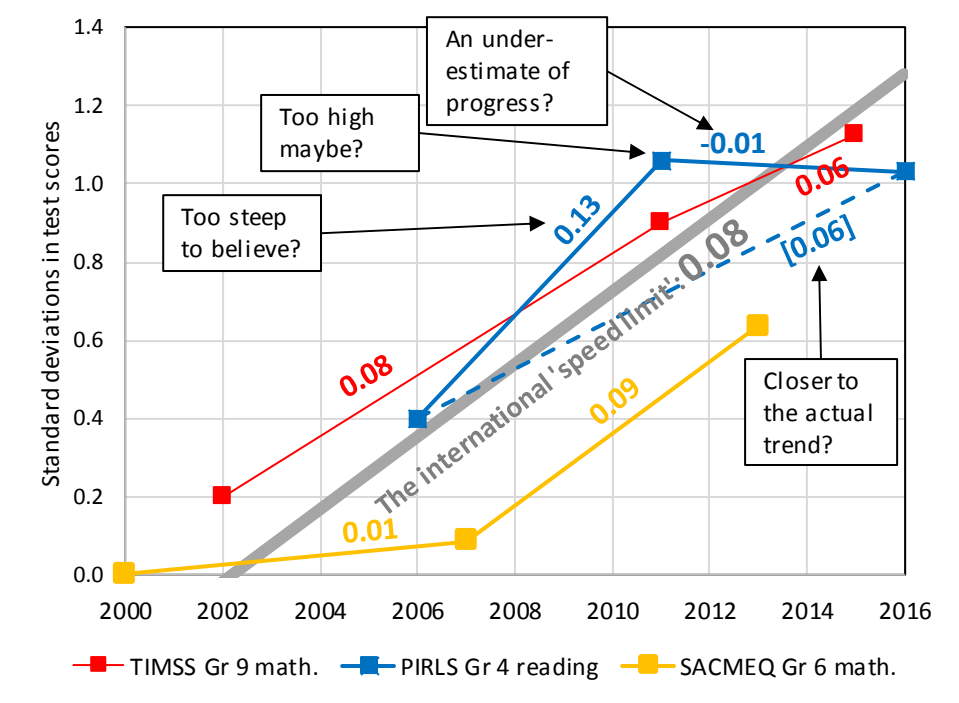
The current document draws from work done for the Department of Basic Education and for ReSEP at the University of Stellenbosch.

This report provides background details relating to the following graph, a graph intended to summarise improvements seen in three testing programmes:

- Trends in International Mathematics and Science Study (**TIMSS**).
- Southern and Eastern Africa Consortium for Monitoring Educational Quality (**SACMEQ**).
- Progress in International Reading Literacy Study (**PIRLS**).

The graph looks as follows. The starting points for each of the three programme-specific series are fairly arbitrary, specifically placed to facilitate the reading of the graph. However, trends over time reflect actual changes in terms of South Africa-specific standard deviations. The values shown in the graph are standard deviation improvements per year, which describe the steepness of the slope. A key intention with the graph is to raise questions around the 2011 PIRLS point and to suggest that it may be incorrectly high, meaning the 2006 to 2016 trend is in more consitent than the figures would suggest at face value. It is completely legitimate to question the values for these international testing programmes. They are mostly right, but not always. The most prominent and convincing paper on specific problems in these programmes is arguably one by John Jerrim (2013), who disputes PISA and TIMSS trends for England.

Figure 1: Trends for South Africa for three testing programmes



Where does the international 'speed limit' come from? It is a term I have come up with, in part because the matter of 'speed limits' in educational improvement barely feature in the literature. What is meant by this is the maximum improvement per year, expressed in national standard deviations, seen by the fastest improving countries in the world in the last few decades. I reflect a value of 0.08 standard deviations a year in the graph. That represents a slightly more permissive figure than the 0.06 I arrived at in my PhD (which has some references to some work by others in this area)[1]. I decided to go beyond the 0.06 a bit in order to acknowledge that Brazil might not reflect the absolute best a developing country could reasonably achieve (large improvements are nearly always seen in developing countries, as these countries have not reached 'natural ceilings' yet). The importance for policy purposes of having some grasp of the 'speed limit' found its way into a UIS report I wrote[2].

The following table shows, among other things, where the values for the nine test score points, three for each programme, come from.

[1] Gustafsson, 2014: 136.
[2] UNESCO: UIS, 2018: 30.

Test year	*Score*	*South African standard deviation*	*Sources and comments*
TIMSS Grade 9 mathematics			
2002	289	91	The average score in Reddy *et al* (2012: 3) is 285. This figure *excludes* independent schools. This was verified by examining the microdata. In TIMSS 2011, the difference between the average score *with* independent schools of 352 (this is the figure appearing in Reddy *et al* [2012: 3]) and the average score *without* independent schools of 348 (Reddy *et al* [2012: 5]) is 4 TIMSS points. This is what explains the raising of 285 by 4 to produce the 289 seen here. The standard deviation is calculated from the microdata. It seems the *Grade 9* TIMSS microdata are not available via the international TIMSS-PIRLS website, though the *Grade 8* data are. The HSRC's TIMSS team would have the data and can probably share it – see http://www.timss-sa.org.za.
2011	352	86	Mullis, Martin, Foy and Arora (2012: 488). Microdata available via the international TIMSS-PIRLS website.
2015	372	87	See Mullis *et al* (2016: Appendix G.4). Microdata available via the international TIMSS-PIRLS website.
PIRLS Grade 4 reading			
2006	253	n/a	The 253 value appears in two publicly available documents: Howie *et al* (2008: 19) and Department of Education (2008: 34). This figure represents the performance of Grade 4 learners in the same Grade 4 test written by learners in the rest of the world. It did not receive the attention of the Grade 5 result, which was 302, and was published in the international reports. The report Howie, Combrinck, Roux, Tshele, Mokoena and McLeod Palane (2017: 10) refers briefly to technical difficulties surrounding the 253 value. I imagine this is a reference to the fact that learners perform so far below the level for which the test was designed. This is indeed a problem, but I do not believe the 253 value can be far from an accurate reflection of Grade 4 reading performance in 2006, in part because the difference between this 253 and the published Grade 5 results is roughly what one would expect. The microdata for Grade 4 (unlike that for Grade 5) is not available on the international TIMSS-PIRLS website. The Centre for Evaluation and Assessment (CEA) at the University of Pretoria would have the microdata.
2011	323	n/a	See Howie, Combrinck, Roux, Mokoena and McLeod Palane (2017: 6), also Howie, Combrinck, Roux, Tshele, Mokoena and McLeod Palane (2017: 84). What is not made explicit in these reports, but has been confirmed with the authors, and appears on the TIMSS-PIRLS website, is that 2011 Grade 4 results, using prePIRLS tests, were rescaled to the regular PIRLS scale, shortly prior to the release of these 2017 reports. Hence the national prePIRLS average of 461 seen in Mullis, Martin, Foy and Drucker (2012: 39) was converted to the 323 average seen here. The microdata with the original and rescaled scores appear to be available on the TIMSS-PIRLS website.
2016	320	106	See Mullis *et al* (2017: 321).
SACMEQ Grade 6 mathematics			
2000	486	108	See Makuwa (2010). The standard deviation was calculated from the microdata, which is available on request through the Department of

Test year	*Score*	*South African standard deviation*	*Sources and comments*
			Basic Education.
2007	495	99	See Makuwa (2010). The standard deviation was calculated from the microdata, which is available on research through the Department of Basic Education.
2013	552	101	See Department of Basic Education (2017: 27). The standard deviation is calculated from the microdata. I have had privileged access to the microdata as a result of my involvement in an enquiry, during 2016, into the calculation of preliminary scores by the SACMEQ office. Unfortunately, preliminary and incorrect SACMEQ 2013 scores were communicated from the central SACMEQ office to the Department of Basic Education, and then presented to Parliament. The incorrect scores pointed to 2007 to 2013 improvements for South Africa, and other SACMEQ countries, which were improbably large, and way in excess of the 'speed limit' referred to above. The annual improvement for South Africa was 0.13 standard deviations. This received much media attention – see in particular Nic Spaull's blog post titled 'Serious technical concerns about SACMEQ IV results presented to parliament' at https://nicspaull.com/2016/09/14/serious-technical-concerns-about-sacmeq-iv-results-presented-to-parliament. The decision was taken to conduct a formal enquiry into the 2013 SACMEQ scores. These scores were then recalculated and became the scores seen in Department of Basic Education (2017). I have signed an agreement that I would not reveal the details of what was found in the enquiry. However, what I can say is that I was sufficiently satisfied that the recalculated scores were a reflection of reality.

Finally, I decided to use just three standard deviation values, one for each programme. They were as follows:

- TIMSS: 99. This is the average across the three values in the above table.
- SACMEQ: 104. Again, the average across three values.
- PIRLS: 106. I used the only value I had. It is unlikely that this would change much if I had calculated the average across three values.

References

Department of Basic Education (2017). *The SACMEQ IV project in South Africa: A study of the conditions of schooling and the quality of education*. Pretoria. Available from: < http://www.education.gov.za > [Accessed November 2017].

Department of Education (2008). *Education for All: Country report: South Africa*. Pretoria. Available from: <http://unesdoc.unesco.org/images/0022/002218/221801e.pdf> [Accessed December 2017].

Gustafsson, M. (2014). *Education and country growth models*. Stellenbosch: University of Stellenbosch. Available from: <http://scholar.sun.ac.za/handle/10019.1/86578> [Accessed April 2014].

Howie, S., Combrinck, C., Roux, K., Mokoena, G.M. & McLeod Palane, N. (2017). *PIRLS literacy 2017: South African highlights report*. Pretoria: Centre for Evaluation and Assessment. Available from: <http://www.up.ac.za/media/shared/164/ZP_Files/pirls-literacy-2016-hl-report-3.zp136320.pdf> [Accessed December 2017].

Howie, S., Combrinck, C., Roux, K., Tshele, M., K., Mokoena, G. & McLeod Palane, N. (2017). *PIRLS literacy 2017: Progress in International Reading Literacy Study 2016: South African Children's Reading Literacy Achievement*. Pretoria: Centre for Evaluation and Assessment. Available from: <https://www.up.ac.za/media/shared/164/ZP_Files/pirls-literacy-2016_grade-4_15-dec-2017_low-quality.zp137684.pdf> [Accessed October 2017].

Howie, S., Venter, E., Van Staden, S., Zimmerman, L. & Long, C. (2008). *PIRLS 2006 summary report: South African children's reading literacy achievement*. Pretoria: Centre for Evaluation and Assessment. Available from: <https://nicspaull.files.wordpress.com/2011/04/howie-et-al-pirls-2006-sa-summary-report.pdf> [Accessed December 2017].

Jerrim, J. (2013). The reliability of trends over time in international education test scores: Is the performance of England's secondary school pupils really in relative decline? *Journal of Social Policy*, 42(2): 259-279.

Makuwa, D.K. (2010). Mixed results in achievement. *IIEP Newsletter,* XXVIII(3). Available from: <http://www.iiep.unesco.org> [Accessed October 2010].

Mullis, I.V.S., Martin, M.O., Foy, P. & Hooper, M. (2017). *PIRLS 2016 international results in reading*. Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed December 2017].

Mullis, I.V.S., Martin, M.O., Foy, P. & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed December 2012].

Mullis, I.V.S., Martin, M.O., Foy, P. & Drucker, K.T. (2012). *PIRLS 2011 international results in reading*. Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed June 2014].

Mullis, I.V.S., Martin, M.O., Foy, P. & Hooper, M. (2016). *TIMSS 2015 international results in mathematics*. Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed December 2016].

Reddy, V., Prinsloo, C., Arends, F. & Visser, M. (2012). *Highlights from TIMSS 2011: The South African perspective*. Pretoria: HSRC. Available from: <http://www.hsrc.ac.za> [Accessed January 2013].

UNESCO: UIS (2018). *Costs and benefits of different approaches to measuring the learning proficiency of students (SDG Indicator 4.1.1)*. Montreal. Available from: <http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2019/01/IP53-costs-benefits-approaches-measuring-proficiency-2018-en.pdf> [Accessed January 2019].