TIMSS, SACMEQ, PIRLS and Matric
A reconciliation of the trends

Martin Gustafsson
Department of Basic Education (also ReSEP at University of Stellenbosch)
20 December 2017

1 Introduction

This report presents a reconciliation of three sources of data relating to South Africa's educational outcomes:

- 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).

Many questions are raised about whether the above sources 'speak to each other', whether they point to the same general trends. Such questions are obviously crucial in the education policy debates. If all sources point in the same general direction, fairly strong conclusions can be drawn about the general effectiveness or ineffectiveness of our current education strategies. If the sources contradict each other, then one must perhaps accept more uncertainty with regard to what has been happening in the basic education sector. The recent release of the 2016 PIRLS results has appeared confusing to some, as they point to a lack of progress, where other sources had pointed to improvements. The reliability of the SACMEQ, but even the TIMSS, results have been questioned.

The methods applied in this report draw from an emerging body of literature (referred to below, to some extent) dealing with issues such as comparisons of trends across different testing systems using standard deviations, gauging grade-on-grade learning gains within single schooling systems, and risks which can affect the reliability of country scores in international testing systems.

2 An account of key national statistics

The table appearing below provides details on ten vital statistics on educational performance in the schooling system in the period 2002 to 2016. For TIMSS and SACMEQ, the focus was on mathematics, and not on the other subject in these programmes: science in TIMSS, reading in SACMEQ. Had these other subjects been used instead, it is unlikely that the analysis would change substantially. PIRLS focusses only on reading.

In **TIMSS**, mathematics in Grade 9 was tested in 2002, 2011 and 2015. The South Africa country averages point to a large improvement over this period, of around .95 of a standard deviation, or .07 of a standard per year (counting an improvement from 289 to 372, and a standard deviation of 87). How large or small is a .07 standard deviation a year improvement? It is not large to the extent of being unbelievable, yet it is large enough to be impressive. Brazil, a particularly rapid improver, achieved an annual average gain in its PISA[1] mathematics scores of around .05 standard deviations a year over the twelve-year period 2010 to 2012[2]. Hanushek and Woessman (2007: 4) moreover consider a .05 standard deviation a

[1] Programme for International Student Assessment (PISA).
[2] See Carnoy *et al* (2015: 9). Note one PISA point is around .01 of a Brazilian standard deviation.

year improvement as ambitious but obtainable for developing nations[3]. South Africa's .07 a year improvement can thus be described as considerable and impressive, but not beyond the realm of the believable. We can probably consider the 2002 score of 289 a point of departure representing a long period of no improvement. For the years 1995, 1999 and 2002, TIMSS tested mathematics in *Grade 8* (in 2002 both grades 8 and 9 were tested). The Grade 8 trend across the years 1995 to 1999 to 2002 was virtually flat, the averages being 278, 275 and 264[4]. What is confusing is that a very high value for South Africa of 354 for 1995 was published in a *pre-final* version of the TIMSS 1995 report (Beaton, Mullis *et al,* 1996: 22). This figure of 354 has been repeated in Howie (1997: 37). That value was clearly a mistake, and was revised to 278 in subsequent reports[5]. The Grade 9 score of 2002 is thus considered a 'base' in the table below.

South Africa's 2007 **SACMEQ** score of 495 in Grade 6 mathematics can probably also be considered the base, or a low-level point of departure. A similar score of 486 was obtained in SACMEQ 2000[6]. Between 2007 and 2013, however, there was clearly a substantial improvement. The Department of Basic Education's (DBE) final report on SACMEQ 2013 points to an increase of 495 to 552, around .57 of a standard deviation, or .10 a year. In part due to some confusion brought about by the release of pre-final 2013 SACMEQ scores which were considerably higher than the final ones[7], the DBE checked the 2007 to 2013 trend using the raw test data for South Africa submitted to the SACMEQ office. Specifically, trends using just 23 questions (items) shared across 2007 and 2013 were analysed. This analysis pointed to an overall improvement of .34 standard deviations between 2007 and 2013, or .06 a year. This .06 of a standard deviation a year is lower than the .10 a year obtained using the item response theory (IRT) scores calculated by the international SACMEQ office. Such a large discrepancy between classical scores and IRT scores should not occur. Reasons for the difference will hopefully become clearer once access is obtained to the technical documentation relating to the SACMEQ 2013 process. Despite these issues, what remains clear is that large improvements occurred in South Africa's SACMEQ scores between 2007 and 2013. Below, the impact of using each of the two annual gain statistics will be considered.

The monitoring of progress according to **PIRLS** is made complex by the fact that the 2011 score for Grade 4 reading was recalibrated recently to make it comparable to both 2006 *Grade 5* test results and 2016 *Grade 4* results. (A further complexity relating to PIRLS is that some testing occurred only in sub-national samples of schools. This is ignored here. Only results which are representative of the country as a whole are discussed.) What made headlines in 2017 was the fact that South Africa's Grade 4 reading results appeared not to have improved between 2011 and 2016 – the scores for the two years were 323 and 320[8] (the difference here is too small to be considered a decline, given the sample-based nature of the data). The 2011 score of 323 is of course the recalibrated score. Yet between 2006 and 2011, there was clearly considerable improvement. The 2011 *Grade 4* learners performed around .15 standard deviations better than the 2006 *Grade 5* learners (who obtained the 302 seen in the table). This means the 2011 Grade 4 learners would have performed better than the 2006 *Grade 4* learners by a year's worth of learning, *plus* .15 standard deviations. There are in fact estimates of South Africa's PIRLS Grade 4 results for 2006, as Grade 4 was in fact tested then, though the results have hardly been published given that the level of performance is below a level at which the PIRLS system would reliably distinguish between learners. This is discussed in the next section.

[3] The authors' value of 0.5 was divided by 10 years, the shortest possible period according to the authors.
[4] Mullis *et al* (2000: 36) and Mullis *et al* (2004: 44).
[5] Thanks to Dr Miriam Altman for pointing out these discrepancies to me.
[6] Makuwa, 2010.
[7] Department of Basic Education, 2016.
[8] Howie *et al*, 2017.

KEY NATIONAL STATISTICS 2002-2016								
	Test year	*Grade*	*Class of…*	*Series*	*Score*	*Standard deviation*	*Base?*	*Sources and comments*
1	2002	9	2005	TIMSS math.	289	91	B	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.
2	2006	5	2013	PIRLS reading	302	136	B	See Mullis *et al* (2007: 37, 329).
3	2007	6	2013	SACMEQ math. IRT	495	99	B	See Department of Basic Education (2017: 27). The standard deviation is calculated from the microdata.
4	2007	6	2013	SACMEQ math. classical	39.3	17.9	B	Calculated from the microdata.
5	2011	9	2014	TIMSS math.	352	86		Mullis, Martin, Foy and Arora (2012: 488).
6	2011	4	2019	PIRLS reading	323	n/a		See Howie *et al* (2017: 6). What is not made explicit in this report, but has been confirmed with the authors, is that 2011 Grade 4 results, using prePIRLS tests, were rescaled to the regular PIRLS scale, shortly prior to the release of this 2017 report. Hence the national prePIRLS average of 461 seen in Mullis, Martin, Foy and Drucker (2012: 39) was converted to the 323 average seen here. What this means is that the Grade 4 average of 323 is comparable to the Grade 5 average of 302. Grade 4 learners in 2011 scored 21 points better than Grade 5 learners in 2006, using a comparable scale.
7	2013	6	2019	SACMEQ math. IRT	552	101		See Department of Basic Education (2017: 27). The standard deviation is calculated from the microdata.
8	2013	6	2019	SACMEQ math. classical	45.3	17.6		Calculated from the microdata.
9	2015	9	2018	TIMSS math.	372	87		See Mullis *et al* (2016: Appendix G.4).
10	2016	4	2024	PIRLS reading	320	106		See Mullis *et al* (2017: 321).

3 Differences expressed in years-worth of learning

Improvements in learning outcomes in schools are increasingly being expressed in years-worth of learning, as this is easy to understand. Thus if one gets learners in Grade 5 to perform as well as the Grade 6 learners ten years previously, one could say one has obtained a gain equal to a years-worth of learning over ten years.

Estimates of how much better learners in one grade perform, relative to learners one grade down, in a particular year, are obviously needed if one is to talk about years-worth of learning. For developed countries where national standardised tests allow for comparison of the competencies of learners across grades in one point in time, estimates have been published, for instance those of Hill *et al* (2008: 173) for the United States. The United States estimates point to a general pattern: gains are larger in lower grades than higher grades. This is consistent with the notion if we want Grade 12 results to improve, the most effective strategy is to work hard at improving grade-on-grade gains in the foundation grades. The US estimates for mathematics range from a gain of 1.03 standard deviations between grades 1 and 2, to a gain of just .25 standard deviations between grades 9 and 10. (The standard deviations used were, for instance, the average standard deviation amongst grades 1 and 2 learners when gauging the grades 1 to 2 gain.)

In South Africa, we have very few opportunities to gauge 'business-as-usual' grade-on-grade gains, but there are two in particular.

Firstly, we can compare the Grade 5 PIRLS score of 302 for reading in 2006, to the corresponding average score if we look at the data of the 2006 Grade 4 learners (who wrote the same test as the Grade 5 learners). The Grade 4 average was 253, a figure which can be found in two publicly available reports: Howie *et al* (2008: 19) and Department of Education (2008: 34). The difference between 302 and 253 comes to .36 of a standard deviation (using the 2006 Grade 5 standard deviation of 136). In Hill *et al* (2008), the grades 4 to 5 gain in mathematics for the United States is said to be .56 of a standard deviation. One can expect grade-on-grade gains to be smaller in developing countries such as South Africa, compared to developed countries. The vertical gap between the two PIRLS points in Figure 1, of 1.78, is calculated as follows (there are actually three PIRLS points, but the top two overlap given the closeness of the 2011 and 2016 results). It was assumed that the grade-on-grade gain between Grade 4 and Grade 5 comes to 43.4 PIRLS points. The value 43.4 is the average across three separate estimates: (1) An estimate of 42.4 uses the standard deviation from Hill *et al* and combines this with the 106 PIRLS points standard deviation seen in the above table; (2) An estimate of 36.0 uses the two measures of PIRLS performance in 2006 of 289 and 253 for grades 5 and 4 respectively; (3) An estimate of 51.9 draws from our National School Effectiveness Study (NSES) data of 2007 to 2009, which points to the average annual grade-on-grade gain between grades 2 and 4 being 0.49 of a standard deviation[9]. The average of 43.4 was then found as a fraction of the 302 minus 253 difference, giving 0.78. To 0.78 was added 1.00, representing the difference between Grade 5 and Grade 4, given that Grade 5 performance in one year is being compared to Grade 4 performance in a later year.

The second opportunity for gauging grade-on-grade gains in South Africa is presented by the difference between Grade 8 and Grade 9 performance in the TIMSS mathematics tests in 2002. This difference was 21 TIMSS points (285 minus 264). This informs the vertical gaps between the 2002, 2011 and 2015 TIMSS points in the graph.

Figure 1 essentially takes the 'Score' values from eight of the ten rows in the above table – the SACMEQ classical scores were excluded – and brings them together in a picture where on the whole the assumption is made that any progress was distributed equally across years. Each

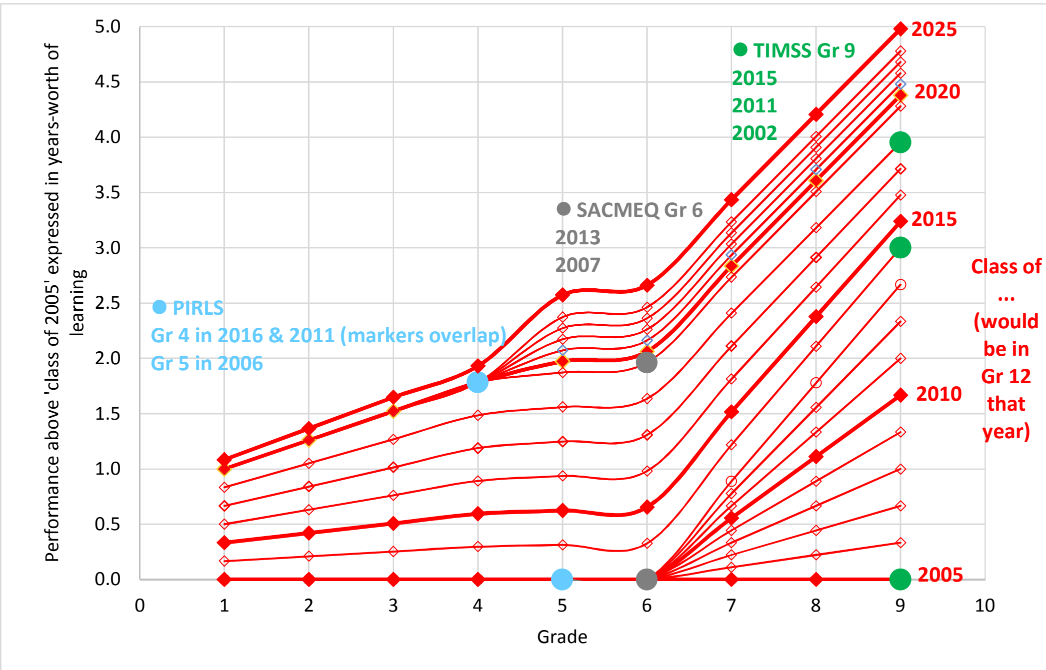
[9] Values from two different tables in the book Taylor *et al* (2013) were used for this.

curve represents a 'Class of …', meaning a cohort of learners reaching Grade 12 in a particular year. The class of 2005 was assumed to be the base year, as the discussion so far has suggested that before this cohort, there was little change. What seems to have happened is major initial progress at the *secondary* level (specifically Grade 9), with little progress at the *primary* level, at least up to the 'Class of 2013'. The 'Class of 2014' seems to be the first cohort to benefit from primary-level improvements indicated by both the SACMEQ and PIRLS results. Importantly, the 'Class of 2013' was tested by both PIRLS (in 2006 in Grade 5) and again in SACMEQ (in 2007 in Grade 6). Both these results appeared no higher than earlier results (as discussed above), suggesting that the subsequent 'improvement wave' would at the earliest have been felt by the 'Class of 2014'.

Two key assumptions used in the graph should be explained. Firstly, a grade-on-grade gain figure for Grade 6 was chosen that lay between the gains suggested separately by PIRLS (for grades 4 and 5) and TIMSS (for Grade 9). Importantly, this assumption about grade-on-grade gains is in terms of *standard deviations*. Secondly, it was assumed that year-on-year gains at Grade 1 were felt initially by the 'Class of 2014' (who would have been in Grade 1 in 2002), and that up to the 'Class of 2019' (in Grade 1 in 2007) Grade 1 performance improved by one year's worth of learning (so less than in Grade 4, according to PIRLS). The 'Class of 2020' experienced no Grade 1 improvement, relative to the 'Class of 2019', in line with the flat trend seen in the newly released PIRLS Grade 4 trend.

An optimistic assumption was made that for performance *above* Grade 4 for the more recent 'Classes'. It was assumed that in Grade 5 and above, every year learners would perform a bit better, given interventions above Grade 4, in part aimed at counteracting the weak (or non-existent) improvement in the lower grades in recent years. Finally, it was assumed, rather optimistically, that the 'Class of 2025' would benefit from a small uptick in Grade 4 performance post-2016.

Figure 1: Learning outcome trends by Matric cohort (I)



Importantly, the relative consistency of the eight performance points seen in the above graph holds only partially if one uses the SACMEQ improvements seen in the classical scores. Using these scores reduces the 2007 to 2013 SACMEQ improvement from .57 to .34 standard deviations, meaning one ends up with an improvement that is 40% smaller. The following graph is like the previous one, but with the use of the SACMEQ classical scores. As discussed above, greater clarity is needed around the inconsistencies between the SACMEQ classical and IRT scores, clarity which should become possible when the technical documentation on the calculation of the IRT scores becomes available.

Figure 2: Learning outcome trends by Matric cohort (II)

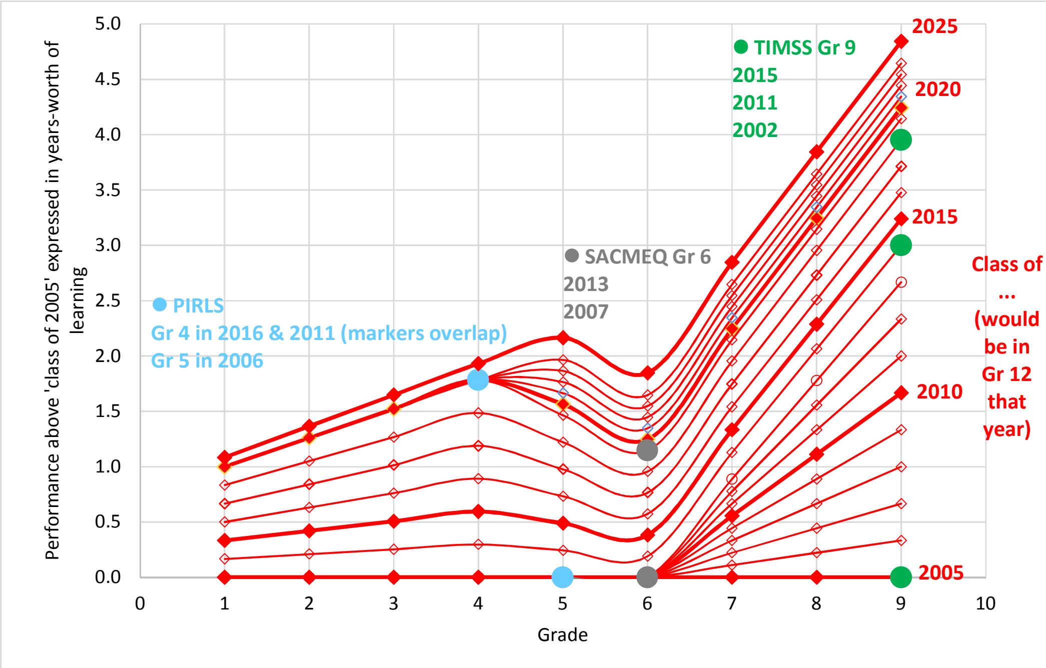
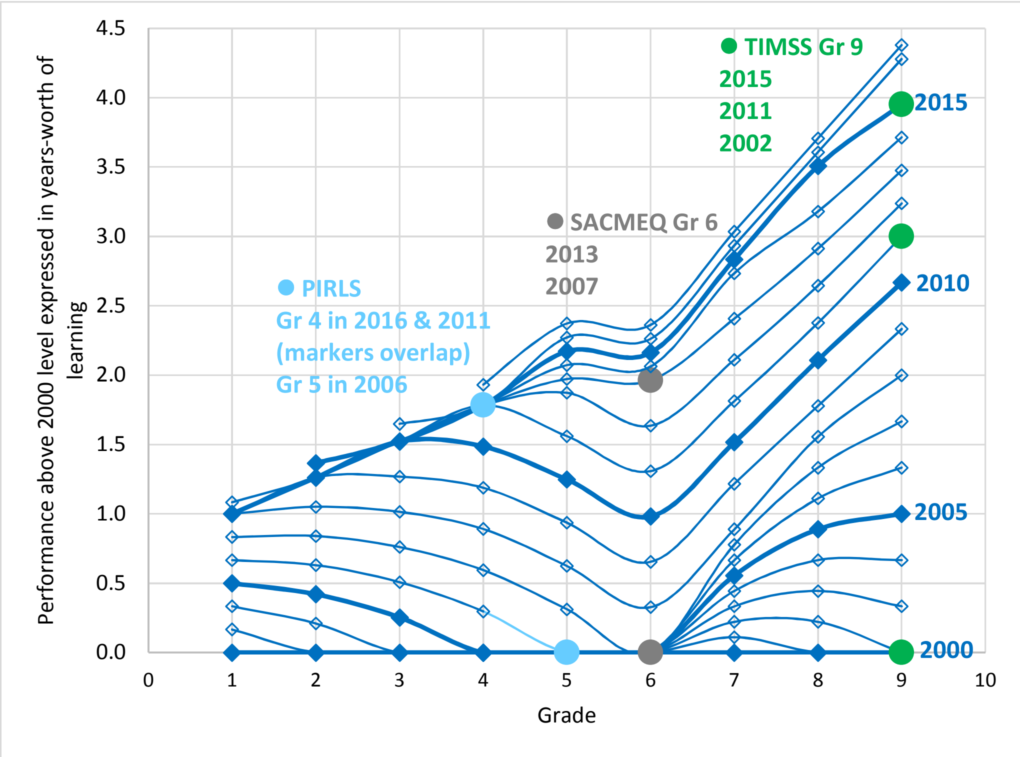


Figure 3 below uses exactly the same data points as Figure 1, meaning SACMEQ IRT scores are used, but here curves represent not cohorts of learners, but simply the year in which learning took place. Analysing cohorts is important. In particular, the cohort approach used above forced the modelling to always avoid unrealistic patterns whereby a cohort's historical advantage in one grade is *lower* than what applied in the previous grade. Given the cumulative nature of learning, one would not expect learning to regress in this manner. However, Figure 3 is probably easier for most people to understand. Generally we think about, for instance, the quality of learning in Grade 4 in 2011, and not about the quality of Grade 4 learning experienced by those reaching Grade 12 in 2019.

Figure 3: Learning outcome trends by grade



References

Beaton, A.E., Mullis, I.V.S., Martin, M.O. & Gonzalez, E.J. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill: Boston College. Available from: <http://timssandpirls.bc.edu/timss1995i/TIMSSPDF/BMathAll.pdf> [Accessed February 2017].

Carnoy, M., Khavenson, T., Fonseca, I. & Costa, L. (2015). Is Brazilian education improving? Evidence from Pisa and Saeb. *Cadernos de Pesquisa*, 45(157). Available from: <http://www.scielo.br/scielo.php?pid=S0100-15742015000300450&script=sci_arttext&tlng=en> [Accessed May 2017].

Department of Basic Education (2016). *SACMEQ IV study results: Portfolio Committee on Basic Education [13 September 2016]*. Pretoria. Available from: < https://pmg.org.za/committee-meeting/23250> [Presentation, accessed November 2017].

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].

Hanushek, E.A. & Woessman, L. (2007). *The role of school improvement in economic development*. Washington: National Bureau of Economic Research. Available from: <http://papers.nber.org/papers/w12832.pdf?new_window=1> [Accessed June 2007].

Hill, C.J., Bloom, H.S., Black, A.R. & Lipsey, M.W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives,* 2(3): 172-175.

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. (1997). *Mathematics and science performance in the middle school years in South Africa*. Pretoria: HSRC. Available from: <http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/15/a1/04.pdf> [Accessed March 2010].

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].

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., Kennedy, A.M. & Foy, P. (2007). *PIRLS 2006 international report*. Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed April 2007].

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].

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., Gonzalez, E.J. & Chrostowski, S.J. (2004). *TIMSS 2003 international mathematics report*. Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed January 2008].

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J. & Gregory, K.D. (2000). *TIMSS 1999 international mathematics report*. Chestnut Hill: Boston College. Available from: <http://timss.bc.edu> [Accessed August 2005].

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].

Taylor, N., Van der Berg, S & Mabogoane, T. (ed.) (2013), *Creating effective schools*. Cape Town: Pearson.