Alignment between South African mathematics assessment standards and the TIMSS assessment frameworks

2 authors:

Mdu Ndlovu
University of Johannesburg
53 PUBLICATIONS **188** CITATIONS

SEE PROFILE

Andile Mji
Tshwane University of Technology
81 PUBLICATIONS **606** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Integrating Bring Your Own Devices in Higher Institution View project

Access to Higher Education STEM programmes View project

Alignment between South African mathematics assessment standards and the TIMSS assessment frameworks

Authors:
Mdutshekelwa Ndlovu[1,2]
Andile Mji[2]

Affiliations:
[1]Institute for Mathematics and Science Teaching (IMSTUS), Department of Curriculum Studies, University of Stellenbosch, South Africa

[2]Faculty of Humanities, Tshwane University of Technology, South Africa

Correspondence to:
Mdutshekelwa Ndlovu

Email:
mcn@sun.ac.za

Postal address:
Faculty of Education (IMSTUS), University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

South Africa's performance in international benchmark tests is a major cause for concern amongst educators and policymakers, raising questions about the effectiveness of the curriculum reform efforts of the democratic era. The purpose of the study reported in this article was to investigate the degree of alignment between the TIMSS 2003 Grade 8 Mathematics assessment frameworks and the Revised National Curriculum Statements (RNCS) assessment standards for Grade 8 Mathematics, later revised to become the Curriculum and Assessment Policy Statements (CAPS). Such an investigation could help to partly shed light on why South African learners do not perform well and point out discrepancies that need to be attended to. The methodology of document analysis was adopted for the study, with the RNCS and the TIMSS 2003 Grade 8 Mathematics frameworks forming the principal documents. Porter's moderately complex index of alignment was adopted for its simplicity. The computed index of 0.751 for the alignment between the RNCS assessment standards and the TIMSS assessment objectives was found to be significantly statistically low, at the alpha level of 0.05, according to Fulmer's critical values for 20 cells and 90 or 120 standard points. The study suggests that inadequate attention has been paid to the alignment of the South African mathematics curriculum to the successive TIMSS assessment frameworks in terms of the cognitive level descriptions. The study recommends that participation in TIMSS should rigorously and critically inform ongoing curriculum reform efforts.

Introduction

Hencke, Rutkowski, Neuschmidt and Gonzalez (2009) make the important remark that the Trends in International Mathematics and Science Study (TIMSS) examines the effectiveness of curriculum and instruction in relation to student achievement. There is increasing global interest in and attention paid to the resultant rankings of participating countries, making the very participation in TIMSS a high-stake local decision. As a consequence of the heightened (political and educational) stakes, the relevance of the tests to local curricula has come under sharp scrutiny, which makes the issue of alignment of the South African (SA) curriculum with TIMSS important for educators, curriculum workers, test developers and policymakers. Hencke et al. (2009) concede upfront that whilst TIMSS assessments were developed to represent an agreed-upon framework with as much in common across countries as possible, it was inevitable that the match between test and curriculum would not be identical in all countries. However, the more aligned a national curriculum is to what is common across countries the greater the chance of that country's students performing well. In other words, rather than reject the common core assessments as irrelevant it might be beneficial to investigate in depth what discrepancies exist between SA's curricula and TIMSS, with special focus on the overlapping content.

Mullis, Martin, Ruddock, O'Sullivan and Preuschoff (2009) refer to the TIMSS curriculum model as consisting of an *intended curriculum*, an *implemented curriculum* and an *attained curriculum*, all of which are familiar terms in curriculum theory. For instance, Porter (2004, p. 1) suggests that a curriculum can be divided into four aspects: the intended, enacted, assessed and learned curriculum. The enacted curriculum refers to instructional events in the classroom whereas the assessed curriculum refers to student *achievement tests*. Mullis et al.'s (2009) attained curriculum refers to *student achievement* in those tests. For cross-national tests such as TIMSS to be valid, it is critical that their assessed curricula correspond with the intended national curricula. Moreover, assessments aligned with the assessment standards can guide instruction and raise achievement (Martone & Sireci, 2009; Polikoff, Porter & Smithson, 2011). In view of the foregoing it is expected that, in order to be relevant, cross-national studies or tests should provide curriculum information that can help countries to improve the quality of their education systems on the basis of benchmarking performance (Reddy, 2006). This makes curriculum matching analysis a logical starting point.

In bemoaning the absence of extensive use of alignment research in the classroom, Martone and Sireci (2009) point out lost opportunities to help policymakers, assessment developers and educators to make refinements so curriculum, assessment and instruction support each other in achieving what is expected of students. In an attempt to bridge this gap the aim of the study was to analyse the alignment between SA's Grade 8 Mathematics curriculum and TIMSS by means of the Porter (2002) procedure. To achieve this goal the remainder of this article gives the theoretical background to alignment studies in general and shows why the Porter index was chosen. Thereafter we spell out the research questions guiding the study and outline the procedure for determining the index before presenting and discussing the results. The article concludes with summary observations and recommendations.

Theoretical framework of alignment studies

Definition of alignment and scope of alignment studies

For purposes of comparing the Grade 8 Revised National Curriculum Statements (RNCS) for Mathematics and the TIMSS assessment we analyse measures of curricula and assessment alignment based on research that has developed methods for judging the extent and nature of alignment (e.g. Porter, 2002; Porter & Smithson, 2001; Webb, 2005). Alignment can be defined as the degree of agreement, match or measure of consistency between curriculum content (content standards) for a specific subject area and the assessment(s) used to measure student achievement of these standards (Bhola, Impara & Buckendahl, 2003; Näsström, 2008; Näsström & Henricksson, 2008). A major feature of alignment studies is the development of common languages of topics and categories of cognitive demand for describing content in different subject areas such mathematics, reading and science (Berends, Stein & Smithson, 2009, p. 4). The underlying logic is that if standards specify what and how well students should be learning and tests measure what they know and can do, then the two ought to be synchronised (Herman & Webb, 2007, p. 1). In other words, the language of the assessment items must match the language of the outcomes stated in the RNCS or its successor, Curriculum and Assessment Policy Statements (CAPS) (Department of Basic Education, 2011a). Similarly, the content and cognitive domain language of the CAPS should match that of the TIMSS assessment frameworks as closely as possible. Alignment, thus, has both content and consequential validity in terms of the knowledge and skills prescribed and tested (Bhola et al., 2003, p. 21) Although the alignment between standards and assessment has been most commonly studied (e.g. Bhola et al., 2003; Herman & Webb, 2007), the alignments between standards and instruction as well as between instruction and assessment have also been studied (e.g. Porter, 2002). In curriculum theory and practice, standards have lately come to refer to 'descriptions of what students are expected to know and be able to do' (Näsström, 2008, p. 16), which makes them synonymous with the intended relationship between educational objectives and subject matter content. In SA, the term 'outcomes' has been used widely to frame statements about both subject matter content and anticipated learning behaviours.

Porter's model for evaluating alignment

From the three commonly used primary models of evaluating alignment, that is Webb's (1997, 2005) Depth of Knowledge Procedure, Rothman, Slattery, Vranek, and Resnick's (2002) Achieve Procedure model and Porter's (2002) Surveys of Enacted Curriculum index, we opted for the last one. Unlike the other two approaches, the Surveys of Enacted Curriculum index does not rely on direct comparison of assessments or assessment items with objectives or standards. Instead, content analysts first code the standards and assessments onto a common framework, a content taxonomy, developed by subject matter experts. The taxonomy defines content in terms of two variables: topics or sub-topics and levels of cognitive demand. The two variables compare favourably with Webb's Categorical Concurrence and Performance Centrality. Analysts place assessment items and objectives from standards documents into the taxonomy and the documents are then represented as matrices of proportions, where the proportion in each cell (topic and cognitive demand) indicates the proportion of total content in the document that emphasises that particular combination of topic and cognitive demand. The matrices for standards and assessments are then compared, cell by cell, and an alignment index is calculated. We believe that the Porter procedure achieves in two dimensions what the Webb and Achieve procedures do in four measures. More importantly, the Porter alignment model 'can be applied to analyse the match between any two of curriculum, instruction and assessment' (Liu et al., 2009, p. 795). It was therefore appropriate for our purpose since we wanted to compare two curriculum documents: the TIMSS frameworks and RNCS.

The calculated Porter alignment index ranges from 0 to 1 with 0.5 as its centre since it uses absolute differences, a characteristic that has to be taken into account when interpreting the computed values. Fulmer (2011) has recently provided critical values for the strength of the Porter index of alignment based on the number of cells and number of standard points used. Furthermore, the Porter procedure agrees with Bloom's Taxonomy of educational objectives which has also been used by TIMSS and the RNCS.

Purpose of the study

The purpose of this study was to determine the degree of alignment between the Grade 8 RNCS for Mathematics and the 2003 TIMSS assessment frameworks by means of the Porter index. TIMSS regularly assesses learners at the Grade 4 and 8 levels and SA has previously participated (in 1995, 1999, 2003 and 2011). The Grade 12 level has not been consistently assessed. We chose the Grade 8 curriculum because it is a transitional grade between primary and secondary phases. The 2003 results were the latest available of South Africa's participation in TIMSS because the 2011 results were still pending at the time of this article. The following research questions guided the study:

- What is the structure of the content and cognitive domain matrices for the components of the 2003 TIMSS Grade 8 Mathematics assessment frameworks?
- What is the structure of the content and cognitive domain matrix for the RNCS Grade 8 Mathematics assessment standards?
- What are the computed Porter indices of alignment within and between the components of the 2003 TIMSS assessment frameworks and the RNCS assessment standards for Grade 8 Mathematics?
- What is the structure of discrepancies in emphasis between the RNCS assessment standards and the 2003 TIMSS assessment objectives?
- How do the discrepancies assessment objectives compare with SA's performance in TIMSS 2003?

To help answer these questions we adopted the document analysis methodology in this study.

Methodology
The document analysis methodology

The methodology of document analysis was adopted for this study as it entails systematic and critical examination rather than mere description of instructional or curriculum documents (Center for Teaching and Learning, 2007). Document analysis is also referred to as qualitative content analysis (Daymon & Holloway, 2011), an analytical method used in qualitative research to gain an understanding of trends and patterns that emerge from data. The aim of qualitative document analysis is to discover new or emergent patterns, including overlooked categories (Daymon & Holloway, 2011, p. 321). Statistical reports within a qualitative study should reveal ways in which the data and statistics have been organised and presented to convey the key messages and meanings intended. The qualitative document analysis in this study is organised and presented statistically by means of the Porter alignment procedure to convey messages and meanings about the strength of the alignment between the TIMSS (2003) mathematics assessment frameworks and the RNCS for Grade 8 Mathematics.

For the empirical work the first author worked with two experienced mathematics in-service facilitators for the Senior Phase and the Further Education and Training phase. The Grade 8 RNCS Mathematics assessment standards were compared with the TIMSS assessment objectives and a common template consisting of 110 standard points or fragments as follows: Number: 32 (8 whole number

points, 5 integer points, 10 fractions and decimals points, 2 irrational number and financial mathematics points, and 7 ratio, proportion and percentage points); Algebra: 26 (3 patterns, 7 algebraic expressions, 7 equations and formulas, 9 relationships/functions); Measurement: 16 (3 attributes and units, 13 tools, techniques and formulae); Geometry: 18 (3 lines and angles, 6 two-dimensional and three-dimensional shapes, 3 congruence and similarity, 4 location and spatial relationships, 3 symmetry and transformations); and Data: 17 (4 data collection and organisation, 4 data representation, 5 data interpretation, 4 uncertainty and probability). In the common template, 88% of the RNCS assessment standards were covered whilst 89% of the TIMSS assessment objectives were covered. The facilitators were introduced to the mathematics cognitive domain categories used in TIMSS 2003 (Mullis et al., 2003, pp. 27–33) and asked to code the cognitive domain levels elicited by each standard point in the template according to the verbs used (see Table 1). Following the level descriptors, the two facilitators independently coded the standards and the author allocated marks according to their coding, totalling 1 score point per standard point. Table 1 shows the content and cognitive domain categories, the weightings and the verbs or descriptors that characterise the cognitive levels.

Table 2 shows an example of five selected RNCS standard points coded following Airasian and Miranda's (2002, pp. 251–253) procedure of coding objectives according to the Revised Bloom's Taxonomy of educational objectives.

Assessment standard point 1.1, for example, uses three verbs, the first of which is in the 'knowing' category and the other

TABLE 1: Content and cognitive domains for mathematics used in TIMSS 2003.

TIMSS 2003 content domains	TIMSS 2003 cognitive domain levels
Number (30%) Algebra (25%) Measurement (15%) Geometry (15%) Data (15%)†	Knowing facts and procedures (15%) (recall, recognise/identify, compute, use tools)
	Using concepts (20%) (know, classify, represent, formulate, distinguish)
	Solving routine problems (40%)† (select, model, interpret, apply, verify/check)‡
	Reasoning (25%) (hypothesise/conjecture/predict, analyse, evaluate, generalise, connect, synthesise/ integrate, solve non-routine problems, justify/ prove)‡

†, See Martin, M.O., Mullis, I.V., & Chrostowski, S.J. (2004). *TIMSS 2003 technical report.* Chestnut Hill, MA: International Association for the Evaluation of Educational Achievement. Available from http://timss.bc.edu/PDF/t03_download/T03TECHRPT.pdf
‡, See Mullis, I.V., Martin, M.O., Ruddock, G.J.,O'Sullivan, C.Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks.* Chestnut Hill, MA: International Association for the Evaluation of Educational Achievement. Available from http://timssandpirls.bc.edu/timss2011/downloads/TIMSS2011_Frameworks.pdf for more detailed examples of these verbs.

TABLE 2: Coding of the RNCS assessment standards according to TIMSS 2003 cognitive levels.

Standard point examples	Sample common template standard points: The student …	Knowing facts and procedures	Using concepts	Solving routine problems	Reasoning	Score points
1.1	Recognises, classifies, and represents to describe and compare integers	X	XX			3 × 0.333 = 1
1.2	Recognises, classifies, and represents to describe and compare decimal fractions	X	XX			3 × 0.333 = 1
2.1	Solves problems that involve ratio			X		1
3.1	Designs and uses questionnaires with a variety of possible responses to collect data		X	X		2 × 0.5 = 1
4.1	Investigates and extends numeric and geometric patterns, relationships or rules				X	1
Score points		0.667	1.833	1.5	1	-

two are both in the 'using concepts' category. Standard point 1.2 uses the same three verbs but for a different content sub-topic (fractions instead of integers); standard point 2.1 solves problems and standard point 4.1 investigates and extends at the reasoning level. Reliability was assured by the independent coding of the experts who were given copies of the relevant pages for the classification of TIMSS assessment objectives as they appear in Mullis et al. (2003, pp. 27–33) and implored to adhere to these as closely as possible. (The inter-rater kappa reliability index could not be computed because it applies to items falling in mutually exclusive categories.)

Computation of the Porter index

As already noted, the Porter procedure analyses the extent of alignment between two matrices or matrices of frequencies (Fulmer, 2011, p. 384). It produces a single alignment index, ranging from 0 to 1 to indicate how closely the distribution of points in the first matrix (of standards) aligns with the second matrix (of assessment). The alignment index P is arrived at in four steps as shown in Figure 1:

1. Create matrices of frequencies for the two documents being compared and label these as X and Y.
2. For each cell in matrices X and Y, compute the ratio of points in the cell with the *total* number of points in the respective matrix. Label the matrices of ratios as x and y.
3. For every row j and column k in matrices X and Y (the matrices of ratios), calculate the absolute value of the discrepancy between the ratios in cells x_{jk} and y_{jk}.
4. Compute the alignment index using the formula
$$P = 1 - \frac{\sum_{k=1}^{K}\sum_{j}^{J}|a_{jk} - b_{jk}|}{2}$$, where J is the number of rows, K is the number of columns in each of matrices X and Y, and x_{jk} and y_{jk} are ratios of points in the cells at row j and column k for each of ratio matrices x and y respectively.

Critical values for the strength of the alignment index

A greater number of cells in the matrices will yield a range of likely values that is lower than for matrices with fewer cells. Hence the total number of cells in the A and B matrices can have an effect on the significance of the alignment index. When we also consider that the centre of the distribution of indices is not zero, as noted earlier, we need to assess how far an observed alignment index is from 0.5. Fulmer (2011) generated a matrix of means and critical values for alignment indices with results also demonstrating the expected (mean) distribution of pattern of alignment indices (see sample entries in Table 3).

In addition to matrix-size dependence, the alignment index also depends on the number of curriculum or standards statements or test items being coded. If the total number of cells in the matrix is N ($= J \times K$) then for matrices A and B, $J = 2$ and $K = 2$ yields $N = 4$. In this study we used Fulmer's (2011) estimates of the critical
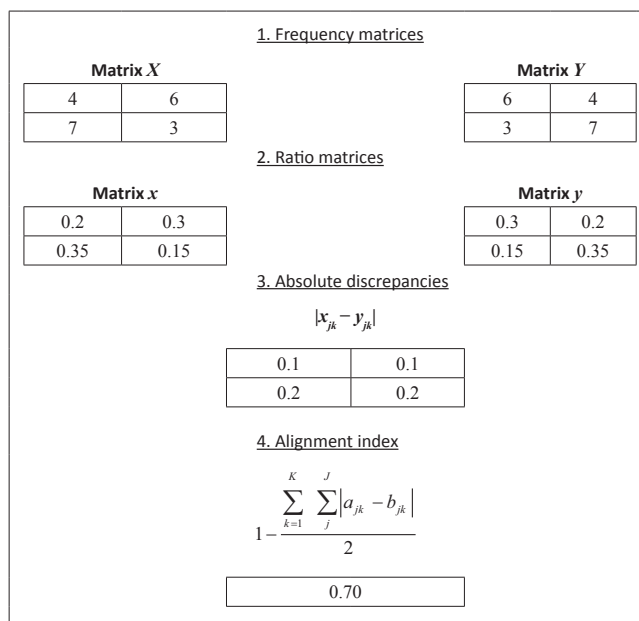

FIGURE 1: Porter alignment index example calculation for 2 × 2 matrices.

TABLE 3: Sample mean alignment indices by number of cells and standard points.

Cells	Standards points			
	30	60	90	120
10	0.9464	0.9782	0.9875	0.9916
20	0.8674	0.9428	0.9635	0.9737
30	0.8054	0.8974	0.9291	0.9438
50	0.6587	0.7958	0.8384	0.8553
70	0.5451	0.7039	0.7478	0.7600
90	0.4614	0.6258	0.6633	0.6716
100	0.4309	0.5908	0.6276	0.6337

Source: Adapted from Fulmer, G.W. (2011). Estimating critical values for strength of alignment among curriculum, assessments and instruction. *Journal of Educational and Behavioural Statistics, 36*(3), 381–402. http://dx.doi.org/10.3102/1076998610381397

values as determined by the number of cells and standard points. Table 4 shows sample reference (or critical) value estimates from the corresponding number of cells and standards points.

From results presented by Porter (2002), for instance, the alignment between the standards of four US states (and the NCTM) and their own assessments ranged from 0.30 to 0.47 for 30 standards points. Six content areas and five cognitive levels were used, which meant 30 squares made up matrices A and B. Table 4 gives a critical value of 0.7372 for the lower quantile ($\frac{0.05}{2} = 0.025$) if a two-tailed test is used at the alpha level of 0.05 (i.e. lower than might be expected by chance). Therefore one can conclude that alignment amongst assessment and standards was very low. Liu et al. (2009) used a coding structure with five content categories and six cognitive levels (hence 30 squares again) to compare the alignment of physics curriculum and assessments for China, Singapore and New York state, China and Singapore had alignments of 0.67, which were significantly lower than the mean ($\frac{0.7372 + 0.8667}{2} = 0.80195$) at the 0.05 level (below the critical value of 0.7372); New York's alignment index of 0.80 was equivalent to the mean.

TABLE 4: Sample reference values for indices of alignment by number of cells and standard points.

Cells	Quantiles for 30 standards points					
	0.025	0.050	0.100	0.900	0.950	0.975
10	0.9167	0.9250	0.9250	0.9667	0.9667	0.9750
30	0.7372	0.7500	0.7643	0.8500	0.8565	0.8667
60	0.5211	0.5342	0.5479	0.6438	0.6571	0.6667
90	0.3684	0.3836	0.4026	0.5176	0.5316	0.5429
	Quantiles for 60 standards points					
20	0.9100	0.9200	0.9288	0.9583	0.9610	0.9625
50	0.7225	0.7412	0.7577	0.8313	0.8394	0.8455
70	0.6137	0.6333	0.6521	0.7548	0.7657	0.7758
100	0.4833	0.5021	0.5230	0.6549	0.6722	0.6859
	Quantiles for 90 standards points					
10	0.9811	0.9822	0.9833	0.9911	0.9922	0.9933
20	0.9302	0.9404	0.9535	0.9727	0.9747	0.9758
50	0.7537	0.7696	0.7894	0.8786	0.8869	0.8931
90	0.5556	0.5773	0.5889	0.7273	0.7421	0.7556
	Quantiles for 120 standards points					
20	0.9333	0.9417	0.9620	0.9821	0.9833	0.9845
60	0.7083	0.7250	0.7495	0.8616	0.8720	0.8801
90	0.5667	0.5833	0.6000	0.7391	0.7551	0.7312
114	0.4833	0.5000	0.5167	0.6500	0.6728	0.6877

Source: Adapted from Fulmer, G.W. (2011). Estimating critical values for strength of alignment among curriculum, assessments and instruction. *Journal of Educational and Behavioural Statistics, 36*(3), 381–402. http://dx.doi.org/10.3102/1076998610381397

Results and discussion

The structure of the content and cognitive domain matrices for the components of the 2003 Grade 8 TIMSS Mathematics assessment frameworks

Three matrices were derived in respect of the three components of the TIMSS assessment frameworks: the TIMSS 2003 assessment objectives, the TIMSS 2003 target percentages and the released TIMSS 2003 test items. The content and cognitive domain matrix for the TIMSS 2003 assessment objectives was derived from the list of objectives given in the TIMSS assessment frameworks document (Mullis et al., 2003, pp. 27–33). Table 5 shows the results of the coding of the 98 (fine-grained) TIMSS objectives. The numerical values form the required matrix. The 98 fine-grained objectives were accorded equal weight guided by the estimated time to be devoted to each of them. Of the 110 fine-grained standard points (assessment objectives or standards) in the template, 12 were not amongst the TIMSS assessment objectives.

The content and cognitive domain matrix for the TIMSS 2003 target percentages (Table 6) was computed by extrapolation from Exhibit 2 (Mullis et al., 2003, p. 9) showing the target percentages of TIMSS 2003 mathematics assessment time devoted to content and cognitive domain for the Grade 8 level. Time devoted was assumed to be equivalent to the importance attached to the respective categories in the frameworks as underpinned by the respective objectives.

The content and cognitive domain matrix (Table 7) for the released TIMSS 2003 test items was derived from tallying the coding of all the released TIMSS 2003 test items. The assumption was that the items were accurately coded and accurately reported on. In the coding process multiple-choice items were allocated one score point each whilst constructed response items were allocated marks, depending on the amount of work to be done, so that at least one third of the assessment came from constructed response items (Martin,

TABLE 5: Results of coding the TIMSS 2003 assessment objectives by content and cognitive domain.

Content domain	Knowing facts and procedures	Using concepts	Routine problem solving	Reasoning
Number	0.141	0.073	0.068	0.000
Algebra	0.096	0.044	0.081	0.018
Measurement	0.016	0.000	0.141	0.000
Geometry	0.117	0.044	0.013	0.008
Data	0.019	0.055	0.043	0.040

TABLE 6: Derived target percentages of TIMSS 2003 mathematics assessment devoted to content and cognitive domain by grade level.

Content domain	Knowing facts and procedures	Using concepts	Routine problem solving	Reasoning
Number	0.045	0.060	0.120	0.075
Algebra	0.075	0.050	0.100	0.063
Measurement	0.023	0.030	0.060	0.038
Geometry	0.023	0.030	0.060	0.038
Data	0.023	0.030	0.060	0.038

TABLE 7: TIMSS 2003 Grade 8 Mathematics content and cognitive domain matrix for test items.

Content domain	Knowing facts and procedures	Using concepts	Routine problem solving	Reasoning
Number	0.070	0.051	0.125	0.032
Algebra	0.060	0.056	0.047	0.084
Measurement	0.042	0.009	0.070	0.037
Geometry	0.032	0.037	0.047	0.042
Data	0.005	0.028	0.065	0.060

Mullis & Chrostowski, 2004, p. 35). The results of the tallying (Table 7) agreed with Exhibit 2.24 in Martin et al. (2004, p. 60) and were converted to proportions.

The structure of the content and cognitive domain matrix for the Grade 8 RNCS for Mathematics

Table 8 shows the content and cognitive domain matrix obtained for SA's Grade 8 Mathematics curriculum. All of the 98 RNCS assessment standards whose content was covered by TIMSS were coded. The resultant score points were

totalled for each content and cognitive domain category and converted to the proportions shown.

Computed Porter indices of alignment for this study

Table 9 shows the calculated raw cell-by-cell differences between RNCS assessment standards and TIMSS assessment objectives. These raw differences were converted to absolute differences, from which the Porter index of alignment was computed using the formula $1 - \dfrac{\sum\limits_{k=1}^{K} \sum\limits_{j}^{J} |a_{jk} - b_{jk}|}{2}$. The computed index was 0.735. The mean-simulated alignment index for a 5 × 4 comparison with 20 cells is 0.9635 (see Table 3). Using a two-tailed test, at the 0.05 alpha level, we looked to the 0.025 and 0.975 quantiles in Table 4. Close to 100 standards points and matrices of 20 squares each were used, so the critical values for 90 standard points are 0.9302 and 0.9758 respectively whilst those for 120 standard points are 0.9333 and 0.9843 respectively. The computed alignment value is well below 0.9302 and 0.9333 in the 0.025 quantiles ($\dfrac{0.05}{2} = 0.025$). The alignment was therefore significantly lower than would be expected by chance at the 0.05 level.

Further iterations were conducted to determine the pair-wise alignment amongst the RNCS assessment standards and the three components of the TIMSS 2003 assessment frameworks. The indices obtained are as shown in Table 10. Surprisingly the pair-wise alignment is significantly lower in all instances, without exception, suggesting a low internal consistency even amongst the TIMSS components themselves.

The structure of discrepancies between TIMSS 2003 framework components and RNCS assessment standards

The first structure of discrepancies investigated was between RNCS assessments standards and the TIMSS assessment objectives. Figure 2 shows the structure of (mis)alignment by content and cognitive domain.

From the graph it is evident that the RNCS assessment standards were stronger than TIMSS assessment objectives in all cases where the bars extend upwards above zero but weaker in those cases where the bars extend downwards below zero. Whilst there is a common criticism of teachers concentrating on knowledge of facts and procedures, the tables shows that the RNCS was weaker than TIMSS objectives in this cognitive level in four of the content domains, namely Number, Algebra, Geometry and Data. The RNCS was stronger with respect to routine problem solving in Number and Data but weaker in Measurement and Geometry. A similarly mixed picture emerged in respect of Reasoning.

The second structure to be investigated was between the RNCS objectives and the TIMSS target percentages. Figure 3 summarises the structure. Surprisingly, the RNCS was stronger on knowledge of facts and procedures in

TABLE 8: Grade 8 RNCS Mathematics content and cognitive domain matrix.

Content domain	Knowing facts and procedures	Using concepts	Routine problem solving	Reasoning
Number	0.141	0.073	0.068	0.000
Algebra	0.096	0.044	0.081	0.018
Measurement	0.016	0.000	0.141	0.000
Geometry	0.117	0.044	0.013	0.008
Data	0.019	0.055	0.043	0.040

TABLE 9: Raw cell-by-cell differences between the RNCS assessment standards and the TIMSS 2003 assessment objectives.

Content domain	Knowing facts and procedures	Using concepts	Routine problem solving	Reasoning
Number	-0.050	-0.003	0.012	0.002
Algebra	-0.077	-0.002	0.002	0.071
Measurement	0.009	0.012	-0.031	0.000
Geometry	-0.049	-0.008	-0.008	0.048
Data	-0.012	-0.047	0.042	0.006

TABLE 10: Porter indices of alignment amongst the TIMSS components and the RNCS.

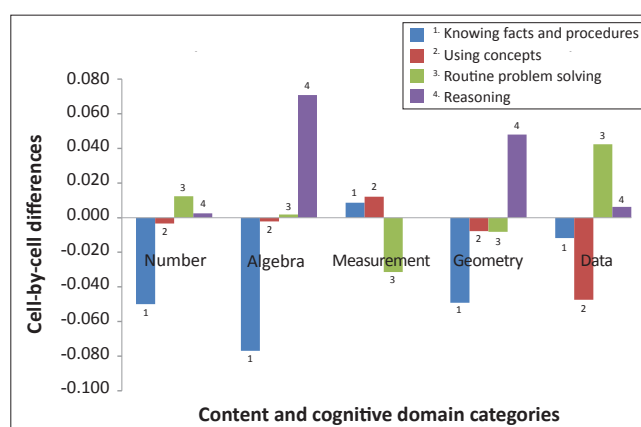| No.	Curriculum Component	1	2	3	4
1	RNCS assessment standards	1.000	0.751	0.758	0.698
2	TIMSS 2003 assessment objectives	0.751	1.000	0.647	0.647
3	TIMSS 2003 test items	0.758	0.647	1.000	0.845
4	TIMSS 2003 target percentages	0.698	0.647	0.845	1.000

FIGURE 2: Discrepancies between RNCS assessment standards and the 2003 TIMSS assessment objectives.
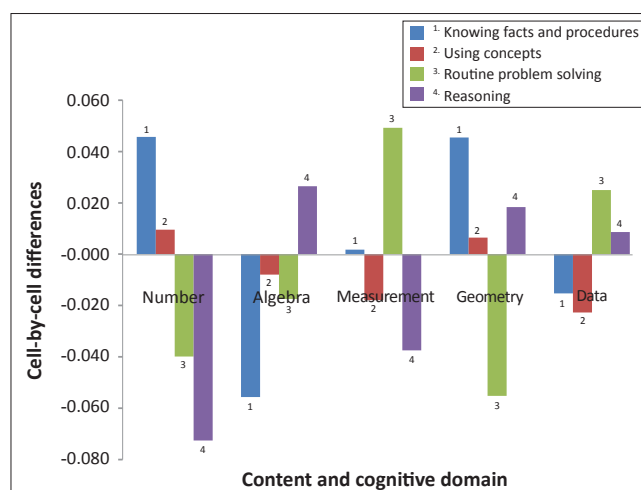

FIGURE 3: Discrepancies between RNCS assessment standards and TIMSS target percentages.

Number, Measurement and Geometry but weaker in Algebra and Data. Taken together, these discrepancies were significant.

A third and final structure to be investigated was that between the RNCS and TIMSS test items. Figure 4 summarises the discrepancies. A marked shift in this comparison is that the RNCS was stronger with respect to knowledge of facts and procedures in Number, Geometry and Data. The RNCS was, however, weaker in respect of Reasoning in all content categories. Routine problem solving was almost evenly split,
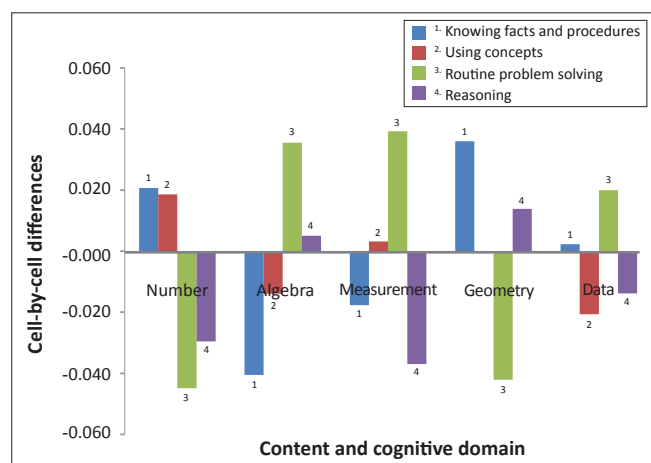
FIGURE 4: Discrepancies between RNCS and assessment standards and TIMSS 2003 test items.
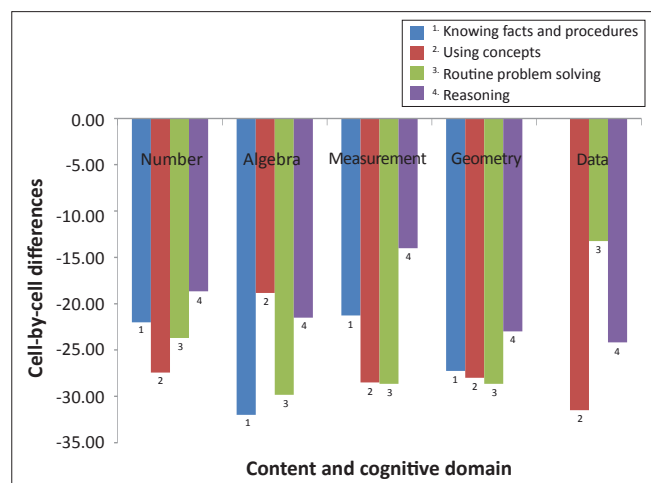

FIGURE 5: South African students' performance by content and cognitive domains in TIMSS 2003 compared to the international average.

above in Algebra, Measurement and Data, but below in Number and Geometry.

The next logical question is whether there was any relationship between the structure of discrepancies and South African students' performance in TIMSS 2003. Given the discrepancies within the TIMSS components themselves, the ultimate question is whether the discrepancies between the RNCS and the TIMSS 2003 test items had any correlation with student performance as they were partially in force at the time of the 2003 TIMSS assessments.

A comparison of the RNCS-TIMSS test discrepancies with South African students' performance in TIMSS 2003

Figure 5 was compiled after extracting South African students' performance in each of the released test categories by content and cognitive domain relative to the international average. It is already well known that South African students performed below the international average across the board (e.g. Reddy, 2006).

Beyond that, however, the intention in this study was to additionally investigate if the pattern of discrepancies between RNCS assessment standards and TIMSS assessment objectives was in any way related to South African students' performance (i.e. the achieved TIMSS curriculum).

Table 11 and Figure 6 attempt to answer that question. It is evident that student performance correlated negatively with discrepancies in Number but positively with discrepancies in Algebra, Measurement, Geometry and Data. That is, the narrower (or positive) the discrepancy was, the closer the performance was to the international average in all content domains except Number.

In Number, SA students performed worst in items on Using Concepts even though this was not the weakest cognitive domain representation in the RNCS assessment standards. In Algebra, SA students performed worst in the Knowledge of Facts and Procedures and this was the weakest category. In Measurement they performed the worst in Routine Problem Solving which was the weakest category of RNCS. In Geometry they performed worst in the Routine Problem Solving category which was the second weakest in the RNCS curriculum. In Data Handling they performed worst

TABLE 11: Comparison of RNCS-TIMSS discrepancy with SA performance in TIMSS 2003.

| Cognitive domain	Discrepancy	Knowledge of facts and procedures	Using concepts	Routine problem solving	Reasoning
Number	RNCS-TIMSS discrepancy	-0.044	0.001	0.017	0.003
	Deviation from international average	-22.000	-27.429	-23.714	-18.667
Algebra	RNCS-TIMSS discrepancy	-0.076	0.000	0.007	0.076
	Deviation from international average	-32.000	-18.833	-29.833	-21.500
Measurement	RNCS-TIMSS discrepancy	0.010	0.013	-0.025	0.000
	Deviation from international average	-21.286	-28.500	-28.667	-14.000
Geometry	RNCS-TIMSS discrepancy	-0.045	-0.006	-0.008	0.051
	Deviation from international average	-27.250	-28.000	-28.667	-23.000
Data	RNCS-TIMSS discrepancy	-0.011	-0.047	0.048	0.009
	Deviation from international average	0.000	-31.500	-13.250	-24.167
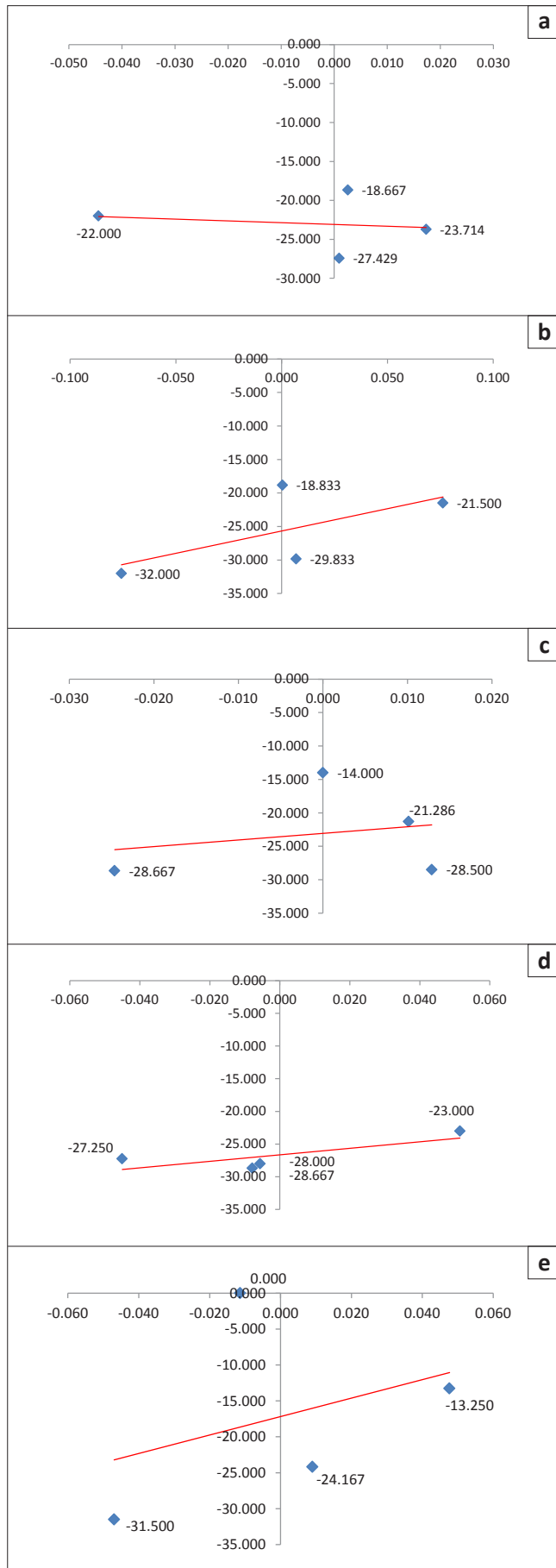
FIGURE 6: Comparison of the RNCS-TIMSS discrepancy in assessment objectives and SA students' performance in TIMSS 2003: (a) Number, (b) Algebra, (c) Measurement, (d) Geometry and (e) Data.

in Using Concepts, which was also the weakest point in the RNCS. Overall, SA learners performed worst in Using Concepts, suggesting little conceptual understanding being achieved by the curriculum. Routine Problem Solving was second worst. This pattern has implications for the intended curriculum which determines what curriculum materials should emphasise and ultimately what teachers should teach in the classroom. Brijlall (2008) notes that the lack of problem-solving skills in SA may be a result of the way it has been taught in schools: individual solution by learners, presentation of abstract problems foreign to learners. There is little doubt that the ultimate answer lies in the implemented curriculum but what feeds into the implemented curriculum is the intended curriculum.

Conclusions

The study reported in this article set out to investigate the alignment of South Africa's RNCS for Grade 8 Mathematics with the TIMSS 2003 Grade 8 Mathematics assessment frameworks. From the results we conclude that the computed Porter index of 0.751 suggests that the misalignment was low enough to warrant urgent attention, from curriculum designers, assessment practitioners, educators, teacher educators and policymakers alike, in order to enhance prospects of improved performance in future participations. In particular there is need to pay attention to the observed discrepancies between the content and cognitive domain emphases. The fact that, even where the RNCS curriculum was stronger than TIMSS, performance was still generally poor suggests the likelihood of a gap between the intended curriculum and the implemented curriculum. Such a gap further suggests a possible mismatch in emphasis between the intended curriculum and the curriculum support materials that actualise it. However, this conjecture requires further investigation. The study also points to the likelihood of a consequential gap between the implemented (SA) curriculum and the attained (TIMSS) curriculum reported by Reddy (2006, p. xiv). From a developing country perspective, what is even more disconcerting is that the three components of TIMSS do not appear to be aligned. That the misalignment is statistically significant calls into question the value-neutrality of TIMSS which currently appears to be a constantly shifting target that only well-resourced, developed countries can cope with.

Finally, participation in TIMSS should not be another bureaucratic ritual. Rather, it should rigorously and reflexively inform curriculum reform and innovation. In an increasingly globalised knowledge economy the school system needs to be globally competitive in the gateway fields of mathematics and science education. A simple illustration of the current disconnect is that, despite SA's participation in previous TIMSS studies, the recently published Curriculum and Assessment Policy Statements (CAPS), which are largely a refinement of the RNCS, proclaim to have been influenced by the cognitive domain levels used in TIMSS 1999 (Department of Basic Education, 2011a, p. 55, 2011b, p. 59). This is so in spite of changes in 2003 (when the country participated in TIMSS for the third time) and further changes

in 2007 (when the country did not participate), which were carried over to 2011. Accordingly, the influence of the Revised Bloom's Taxonomy on TIMSS as evident in the 2007 and 2011 frameworks was apparently not taken into account in the latest curriculum revisions. This suggests that even the newly introduced Annual National Assessments (ANA) for Grades 1–6 and 9, together with the Senior National Certificate examinations, will continue to be guided by out-of-sync domain categories.

By implication, curriculum and assessment will continue to be out of step with international trends resulting in mixed messages for teaching and learning. Bansilal (2011), for example, calls for a closer alignment of curriculum implementation plans with classroom realities. It is ironic that although SA teachers and educationists have complained of rapid curricula changes, the National Curriculum Statement has not changed at the same pace as TIMSS. Accordingly, educators, teacher educators and education researchers should be engaged more constructively in the curriculum and assessment reform processes for sustainable curricula coherence to be achieved. Reddy (2006, p. xiv) reports that during the period of TIMSS 2003, SA teachers consulted disparate curricula documents to determine what and how they taught. As affirmed by Airasian and Miranda (2002, p. 253), severe misalignment of assessment, standards and instruction can cause numerous difficulties. Given the extent to which the misalignment of the SA curriculum has gone relatively unchecked, the school system will continue to buckle for some time to come when subjected to international scrutiny. The latest of such scrutiny is the Global Competitiveness report (Schwab, 2012) and the Southern and East African Consortium for Monitoring Educational Quality report (Spaull, 2011), in which SA ranks very unfavourably.

Acknowledgements

Authors' contributions

M.N. (University of Stellenbosch) was the researcher responsible for the empirical study as well as the writing of the manuscript. A.M. (Tshwane University of Technology), as postdoctoral supervisor, shared his expertise during the main study and the development of the manuscript.

Competing interests

We declare that we have no financial or personal relationship(s) which might have inappropriately influenced our writing of this article.

References

Airasian, P.W., & Miranda, H. (2002). The role of assessment in the revised taxonomy. *Theory into Practice, 41*(4), 249–254. http://dx.doi.org/10.1207/s15430421tip4104_8

Bansilal, S. (2011). Assessment reform in South Africa: Opening up or closing spaces for teachers? *Educational Studies in Mathematics, 78*, 91–107. http://dx.doi.org/10.1007/s10649-011-9311-8

Berends, M., Stein, M., & Smithson, J. (2009). *Charter public schools and mathematics instruction: How aligned are charters to state standards and assessments?* Nashville, TN: National Center on School Choice, Vanderbilt University.

Bhola, D., Impara, J.C., & Buckendahl, C.W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice, 22*(3), 21–29. http://dx.doi.org/10.1111/j.1745-3992.2003.tb00134.x

Brijlall, D. (2008). Collaborative learning in a multilingual class. *Pythagoras, 68*, 52–61. http://dx.doi.org/10.4102/pythagoras.v0i68.67

Center for Teaching and Learning. (2007). *Instructional assessment resources: Document analysis*. Austin, TX: The University of Texas at Austin. Available from http://www.utexas.edu/academic/ctl/assessment/iar/teaching/plan/method/

Daymon, C., & Holloway, I. (2011). *Qualitative research methods in public relations and marketing communications.* New York, NY: Routledge.

Department of Basic Education. (2011a). *Curriculumm and assessment policy statement: Mathematics Grades 7–9.* Pretoria: DBE. Available from http://www.education.gov.za/LinkClick.aspx?fileticket=7AByaJ8dUrc%3d&tabid=672&mid=1885

Department of Basic Education. (2011b). *Curriculum and assessment policy statement: Mathematics Grades 10–12.* Pretoria: DBE. Available from http://www.education.gov.za/LinkClick.aspx?fileticket=QPqC7QbX75w%3d&tabid=420&mid=1216

Fulmer, G.W. (2011). Estimating critical values for strength of alignment among curriculum, assessments and instruction. *Journal of Educational and Behavioural Statistics, 36*(3), 381–402. http://dx.doi.org/10.3102/1076998610381397

Hencke, J., Rutkowski, L., Neuschmidt, O., & Gonzalez, E.J. (2009). Curriculum coverage and scale correlation on TIMSS 2003. In D. Hastedt, & M. von Davier (Eds.), *IERI monograph series: Issues and methodologies in large-scale assessments,* Vol. 2 (pp. 85–112). Princeton, NJ: IEA-ETS Research Institute.

Herman, J.L., & Webb, N.L. (2007). Alignment methodologies. *Applied Measurement in Education, 20*, 1–5.

Liu, X., Zhang, B.H., Liang, L.L., Fulmer, G.W., Kim, B., & Yuan, H.Q. (2009). Alignment between the physics content standards and standardized tests: A comparison among US-NY, Singapore and China-Jiangsu. *Science Educaton, 93*, 777–797. http://dx.doi.org/10.1002/sce.20330

Martin, M.O., Mullis, I.V., & Chrostowski, S.J. (2004). *TIMSS 2003 technical report.* Chestnut Hill, MA: International Association for the Evaluation of Educational Achievement. Available from http://timss.bc.edu/PDF/t03_download/T03TECHRPT.pdf

Martone, A., & Sireci, S.G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research, 19*(9), 11–16. http://dx.doi.org/10.3102/0034654309341375

Mullis, I.V., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks.* Chestnut Hill, MA: International Association for the Evaluation of Educational Achievement. Available from http://timssandpirls.bc.edu/timss2011/downloads/TIMSS2011_Frameworks.pdf

Mullis, I.V., Martin, M.O., Smith, T.A., Garden, R.A., Gregory, K.D., Gonzalez, E.J., et al. (2003). *TIMSS assessment frameworks and specifications 2003.* (2nd edn.). Chestnut Hill, MA: International Association for the Evaluation of Educational Achievement. Available from http://timss.bc.edu/timss2003i/PDF/t03_af_book.pdf

Näsström, G. (2008). *Measurement of alignment between standards and assessment.* Umeå, Sweden: Umeå universitet. Available from http://umu.diva-portal.org/smash/get/diva2:142244/FULLTEXT01

Näsström, G., & Henricksson, W. (2008). Alignment of standards and assessment: A theoretical and empirical study of methods for alignment. *Electronic Journal of Research in Educational Psychology, 6*(3), 667–690.

Polikoff, M.S., Porter, A.C., & Smithson, J. (2011). How well aligned are state assessments of student achievement with state content standards? *American Educational Research Journal, 48*(4), 965–995. http://dx.doi.org/10.3102/0002831211410684

Porter, A.C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3–14. http://dx.doi.org/10.3102/0013189X031007003

Porter, A.C. (2004). *Curriculum assessment.* Nashville, TN: Vanderbilt University. Available from http://datacenter.spps.org/uploads/curricassess.pdf

Porter, A., & Smithson, J. (2001). *Defining, developing, and using curriculum indicators.* Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania. Available from http://www.cpre.org/sites/default/files/researchreport/788_rr48.pdf

Reddy, V. (2006). *Mathematics and Science achievement at South African schools in TIMSS 2003.* Cape Town: Human Sciences Research Council.

Rothman, R., Slattery, J.B., Vranek, J.L., & Resnick, L.B. (2002). *Benchmarking and alignment of standards and testing.* Los Angeles, CA: Center for the Study of Evaluation.

Schwab, K. (2012). *The global competitiveness report 2012–2013.* Geneva: World Economic Forum. Available from http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2012-13.pdf

Spaull, N. (2011). *A preliminary analysis of SACMEQ III South Africa.* Stellenbosch: Department of Economics, University of Stellenbosch. Available from http://www.ekon.sun.ac.za/wpapers/2011/wp112011/wp-11-2011.pdf

Webb, N.L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education.* Madison, WI: National Institue for Science Education.

Webb, N.L. (2005). *Alignment analysis of mathematics standards and assessments, Michigan high schools.* Madison, WI: Author.