


AUTHORS:

Dr S. van Staden¹ Prof. M.A. Graham² Ms J.C. Harvey³ 

AFFILIATION:

¹University of Pretoria, Centre for Evaluation and Assessment²University of Pretoria, Department of Science, Mathematics and Technology Education³Human Sciences Research Council, Inclusive Economic Development research programmeDOI: <http://dx.doi.org/10.18820/2519593X/pie.v38.i2.19>

e-ISSN 2519-593X

Perspectives in Education

2020 38(2): 285-302

PUBLISHED:

04 December 2020

Published by the UFS
<http://journals.ufs.ac.za/index.php/pie>© Creative Commons
With Attribution (CC-BY)

AN ANALYSIS OF TIMSS 2015 SCIENCE READING DEMANDS

ABSTRACT

This study investigated the reading demands of restricted-use items¹ administered to South African Grade 9 learners as part of the Trends in International Mathematics and Science Study (TIMSS) 2015. The method proposed by Mullis, Martin and Foy (2013) was used to categorise items into low, medium and high readability groups. The Knowing domain contained mostly low readability items, the Applying domain was almost equally medium and high readability items, with the Reasoning domain containing mostly high readability items. Results show significant differences between the percentage correctly answered between the low and high categories and between the medium and high categories. However, the full impact of reading demand on performance cannot be fully analysed without cross-reference to English proficiency. Nevertheless, the higher the readability, the greater the chance for learners to answer incorrectly. A continued expected low performance for most South African learners is implied.

Keywords: *language in education; readability; science education; South Africa; Trends in International Mathematics and Science Study (TIMSS) 2015*

1. INTRODUCTION

It is widely known that South Africa has performed very poorly in internationally administered literacy tests over the last couple of years. This is evidenced by the findings of the Progress in International Reading Literacy Study (PIRLS) 2006 cycle (Mullis, Martin, Kennedy, & Foy, 2007), 2011 cycle (Mullis, Martin, Foy & Drucker, 2012) and 2016 cycle (Mullis, Martin, Foy & Hooper, 2017) and by the findings of the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) 2007 cycle, SACMEQ III (Moloi & Chetty, 2011) and 2013 cycle, SACMEQ IV (Department of Basic Education [DBE], 2017). The former, PIRLS, is an international study of reading comprehension of fourth and fifth graders conducted across many countries world-wide and, the latter, SACMEQ is a collaborative network of fifteen ministries of education who periodically conduct standardised surveys in Southern and Eastern

1 The terms “restricted use items” was adopted by the IEA to replace the term “released items”. Whereas the released items were previously freely available from the TIMSS website for further analysis and use, permission to use these, now called restricted use items, have to be sought from the IEA after the TIMSS 2015 cycle.

Africa to assess the quality of education by testing language and mathematics abilities at sixth grade level. Not only did the SACMEQ findings point out concerns regarding reading comprehension of South African learners over the last few years, but it also emphasised problems with mathematics comprehension. Another study that assesses mathematics comprehension and highlighted concerns regarding the South African results, is the Trends in International Mathematics and Science Study (TIMSS); it also assesses science achievement. TIMSS is a large-scale study administered every four years (since 1995) by the International Association for the Evaluation of Educational Achievement (IEA) to assess the mathematics and science knowledge and skills of learners all over the world. Out of approximately 40 participating countries, in both TIMSS 2011 (Mullis, Martin, Foy & Arora, 2012) and TIMSS 2015 (Mullis, Martin, Foy & Hooper, 2016) the Grade 9 South African learners' mathematics performance was ranked second-last, whereas the science performance was ranked second-last in TIMSS 2011 (Martin, Mullis, Foy, & Stanco, 2012) and last in TIMSS 2015 (Martin, Mullis, Foy & Hooper, 2016). South African learner achievement in science in the lower secondary grades (or senior phase)² remains disappointingly low. All three these international studies (PIRLS, SACMEQ and TIMSS) "speak to each other" in the sense that it shows similar trends and highlights major concerns in literacy, mathematics and science comprehension and knowledge by South African learners. Further questions arise such as: If South African learners have very restricted literacy comprehension, how does this affect the understanding of other subjects, for example, a word problem in mathematics or science where the problems often involve a narrative of some sort? Given the vast evidence of poor achievement (evidenced from the international studies such as PIRLS, SACMEQ and TIMSS over the last few years) and contributing contextual factors (such as low family socioeconomic status and poor education quality; this is discussed in more detail in Section 3), the rationale for the current study is to investigate the role that reading demands may play in South African Grade 9 learners' ability to demonstrate an understanding of, and engagement with, restricted use science items from the TIMSS 2015 cycle. The following section is structured according to the IEA's tripartite model of curriculum implementation to provide a cursory contextual understanding of the South African landscape and discusses: 1) the intended science curriculum at lower secondary level, 2) the implemented science curriculum against some contextual background factors, and 3) the attained curriculum as evidenced by the South African Grade 9 science achievement in TIMSS 2015 concludes the section.

2. A CONCEPTUAL FRAMEWORK FOR INTERNATIONAL COMPARATIVE STUDIES

According to Shorrocks-Taylor and Jenkins (2001), the IEA's tripartite model of the curriculum includes: what society would like to see taught in the education system (the *intended curriculum*), what is actually taught (the *implemented curriculum*), and what is learnt (the *attained curriculum*). In his sequential explanatory study of factors connected with science achievement in six countries using TIMSS 1999 data, Reinikainen (2007) refers to the focus on these curriculum manifestations as a broad explanatory factor underlying learner achievement.

Insofar as the *intended curriculum* is concerned, Reddy, Arends, Juan and Prinsloo (2016) summarise the South African science curriculum in terms of three broad subject-specific aims

2 South African school grades in the General Education and Training (GET) band are divided as follows: Foundation Phase (Grades 1–3); Intermediate Phase (Grades 4–6); and Senior Phase (Grades 7–9).

that speak to the purposes of learning science (Mullis, Martin, Goh & Cotter, 2016): doing sciences; knowing the subject content and making connections; and understanding the uses of science. The teaching and learning of natural sciences involves the development of process skills that may be used throughout life and may include, amongst others, the ability to access and recall information, remember relevant facts and key ideas to build a conceptual framework and conduct experiments to test hypotheses. Therefore, the intention of the curriculum is to cover skills increasing in complexity and sophistication by the ninth grade so that, even if a career in a related field is not pursued, learners have scientific skills that translate to various fields (Reddy *et al.*, 2016).

The *implemented curriculum* occurs against a complex contextual background. South Africa continues to tackle injustices stemming from the apartheid legacy that stratified society along racial lines. Transformation has involved improving access to inclusive education for all learners and, once access is ensured, equal quality of education. Contextual factors related to this process and to science achievement were flagged in the diagnostic report of the TIMSS 2015 South African results by Prinsloo, Harvey, Mosimege, Beku, Juan, Hannan and Zulu (2017). For purposes of their analyses, these included issues of language, reading and writing, teacher training, the design of the curriculum, curriculum coverage, availability of laboratory facilities, the Language in Education Policy (LiEP) implementation, and learners' reasoning deficits. The impact of these factors must, therefore, be taken into consideration when evaluating classroom teaching and learner achievement.

Against details of curricular intention and implementation that have been discussed here, it is no surprise that the *attained curriculum* is at persistently poor levels of performance. Results from the TIMSS 2015 study point to Grade 9 learner science performance at 358 (SE=5.6), a score substantially below the international centre point of 500 (Reddy, Visser, Winnaar, Arends, Juan, Prinsloo & Isdale, 2016). Despite poor performance, Reddy *et al.* (2016) note encouragingly that South African learner performance has shown the biggest positive change across cycles when drawing cross-country comparisons, with an improvement of 90 points from TIMSS 2003 to TIMSS 2015, having started at a very low level in the 2003 cycle. While the work of Prinsloo *et al.* (2017) and others (see for example Juan & Visser, 2017; Visser, Juan & Feza, 2015) extensively investigates different contextual factors that affect science education performance, of importance and relevance to the current study's analysis is the role language and reading play from as early as Foundation Phase (FP) when learners start their formal schooling careers. Language and its relationship(s) with reading comprehension is now discussed.

3. LANGUAGE IN EDUCATION AND THE DEVELOPMENT OF READING COMPREHENSION

Language in education in South Africa presents a challenge into an already complicated landscape, a common reality for post-colonial, multilingual countries. Despite constitutional and policy revisions, there remains a three-tiered pyramid that positions English on top, followed by Afrikaans, and lastly the African languages³ (Gupta, 1997; Kamwangamalu, 2000). This has influenced the choice of Language of Learning and Teaching (LoLT) in South African schools. Notwithstanding recommendations that the LoLT be the language of the learner,

3 The 11 official languages are English, Afrikaans, Sepedi, Siswati, Sesotho, Setswana, isiXhosa, isiZulu, isiNdebele, Xitsonga and Tshivenda. The latter 9 languages are termed African languages.

several authors have found evidence that in the majority of South African schools, as a norm, learners are not taught in their primary language (for example, see Fleisch, 2002; Setati & Adler, 2000; World Bank, 2008). However, non-equivalence between home- and instructional language can have negative implications for reading development. Oral language exposure and proficiency provide the vocabulary, grammar and semantic knowledge that assist learners in developing the ability to read (Center & Niestepski, 2014). This process is disrupted when a learner's home language is different from the instructional language. With a lack of basic reading skills, there is little hope of significant improvement as these children grow older and progress from one grade to the next. This outcome was evinced in the PIRLS 2006, 2011 and 2016 cycles.

In PIRLS 2006, 2011 and 2016, Grade 4/5 South African learners were tested across all 11 official languages (see Howie, Venter, van Staden, Zimmerman, Long, Scherman & Archer, 2009; Howie, van Staden, Tshele, Dowse & Zimmerman, 2012; Howie, Combrinck, Roux, Tshele, Mokoena & McLeod Palane, 2017). The PIRLS studies provided evidence that South African children from a young age are not necessarily taught in their home language and are often taught in their second or even third language. PIRLS results across cycles found that children cannot read with understanding and rather engage with text at a surface level where only explicitly stated information could be accessed, at best. The most recent PIRLS 2016 cycle found that 78% of South African Grade 4/5 children could not read for meaning in any language (Mullis, Martin, Foy, & Hooper, 2017, Howie *et al.*, 2017). The results from SACMEQ III and SACMEQ IV, which assesses reading competency at sixth grade level, also shows great concern for South African learners' and teachers' reading comprehension as the SACMEQ III results showed that the poorest quarter of South African learners ranked 14th out of 15 countries (Spaull, 2011) and that the teachers performed worse in SACMEQ IV, in 2011, than in SACMEQ III, in 2007 (DBE, 2017). While SACMEQ assesses reading competency, it is worth mentioning that SACMEQ also assesses mathematics competency levels and SACMEQ IV has also shown that teachers performed worse in SACMEQ IV than in SAQMEC III regarding mathematics competency (DBE, 2017), which is also disconcerting. Turning the focus back to reading competency, the transition from *learning to read* in the early grades to *reading to learn* in subsequent grades is thus highly problematic, since many learners progress to Grade 4 without having basic reading skills in place (see Howie *et al.*, 2009, Howie *et al.*, 2012). Additionally, language-specific difficulties have been highlighted by van Staden, Bosker and Bergbauer (2016) whose analyses of pre-PIRLS 2011 data found that testing in African languages predicts significantly lower results compared to their English counterparts. Reading achievement outcomes for Grade 4 learners who wrote pre-PIRLS 2011 are shown in Figure 1, as taken from van Staden *et al.* (2016), and is now briefly discussed.

Learners tested in English outperformed learners tested in any of the African languages. Additionally, learners across all languages performed worse when the language in which they were tested in pre-PIRLS 2011 differed from their home language. In fact, van Staden *et al.* (2016:1) reported that

... testing in African languages predicts significantly lower results as compared to English, but that exponentially worse results by as much as 0.29 points lower of a standard deviation can be expected when the African language of the test did not coincide with the learners' home language.

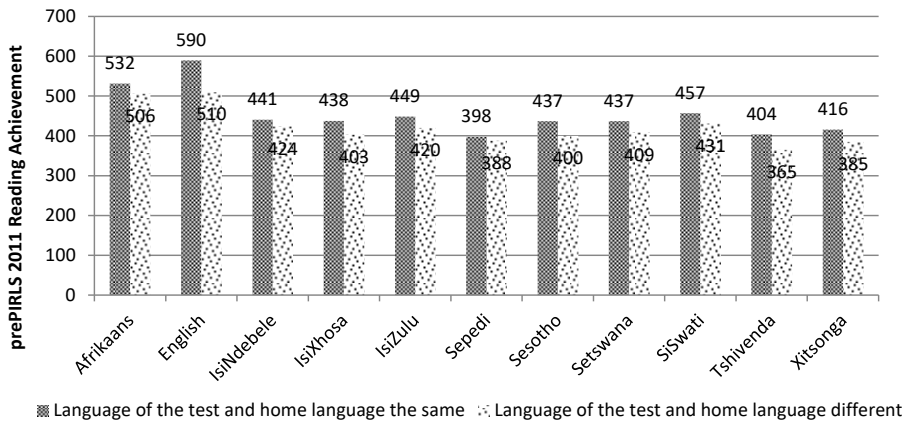


Figure 1: South African Grade 4 learner performance by test language in the same or different language to their home language (van Staden *et al.*, 2016).

Findings from this study provide evidence that African children stand to be disadvantaged the most when a strong home language base has not been developed and when education for children between Grade 1 and 3 is only available through a medium of instruction other than the home language (van Staden *et al.*, 2016).

Language issues are compounded in rural schools where learners have fewer opportunities to engage with science content and there is a lower probability that learners have a solid foundation in English that would allow them access to subject-related vocabulary and terminology. There can be systemic factors that negatively impact literacy development; for example, low family socioeconomic status (see Hemmerechts, Agirdag & Kavadias, 2017; Dowd, D'Sa, Noble, O'Grady, Pisani & Seiden, 2018; Zuilkowski, McCoy, Jonason & Dowd, 2019); poor education quality (see Harley, Woldie & von Gogh, 2019; Harris, Slate, Moore & Lunenburg, 2020) and a lack of available resources including African language use in published and online texts. In a South African context, the latter obstacle is exacerbated by the fact that South Africa has one of the highest linguistic diversities in the world with 11 official languages and many other indigenous languages that are not official, and the fact that a large percentage of rural English Second Language (ESL) learners have been shown to be "non-readers in English". For example, Draper and Spaul (2015) found that 41% of a sample of Grade 5 rural ESL learners were classified as "non-readers in English" after analysing data from the National Education and Evaluation Development Unit (NEEDU) of South Africa.

Since reading forms the foundation of all future learning, the impact of poor literacy development continues to hamper learners as they progress through their academic trajectory. In addition, conceptual gaps progressively worsen as the curriculum increases in difficulty. Within science, learners are positioned as outsiders, not only to subject-specific language and customs, but also to the English language. The next section thus pays particular attention to issues of readability, the role of scientific language in science achievement and some measurement recommendations that have emanated from previous studies based on readability concerns. The section concludes with an overview of selected, previous TIMSS science item readability studies that utilised a wide array of readability measures before presenting the methods that will be used for the current analyses.

4. AN OVERVIEW OF READABILITY ISSUES RELATED TO SCIENTIFIC LANGUAGE AND MEASUREMENT

For purposes of the current study, the definition provided by Oakland and Lane (2004: 244) will be used, namely that readability is “the ease with which a reader can read and understand text”. Linguistic challenges in the South African education context have been recognised as stemming, in part, from the difference between home language and LoLT for learners across grades, but most importantly in the early grades when solid foundations for reading have to be laid. But in uncovering any issues of readability for purposes of the current study, the language of science presents another challenge. The language of science, with its specific genre, often serves as a barrier to the learning of science (Ramnarain, 2012). Some of these barriers include specific vocabulary, terminology, grammar and text structure (Nyström, 2008). It was illustrated in earlier sections that an absence of basic reading skills in the early years predicts nothing good in terms of progress and academic success in later years. Yet, English first-language learners, as well as second-language learners, often struggle to understand specialised terminology, since words often mean something different when used in a scientific context than in an everyday context (Dempster & Reddy 2007). Words such as power, consumer, energy and conduct are cited by Dempster and Reddy (2007) as examples of such everyday constructs that take on a different meaning in a scientific context, with matters being exacerbated by African languages that often use a single term for a concept that is embodied by three or four different terms in English.

Linguistic issues like these complicate the assessment of science considerably, especially where learners’ home language is not compatible with the language of science. Nyström (2008) refers to the work of Schleppegrell (2007), who expounded on the multi-semiotic formation of mathematics, its dense noun phrases and the precise meaning of conjunctions that link elements to one another. Dempster and Reddy (2007) provide examples of logical connectiveness (for example, if, therefore, unless), and prepositions of location (for example in, after and of) as particularly problematic at the interface of English and any of the African languages. Across indigenous South African languages (with the exception of Afrikaans) a dearth of linguistic tools such as those found in English simply means learners stand to lose important information when translating English questions in a test to their home language. In the presence of linguistic dissimilarities, a contextual background of impoverishment and deprivation for most learners makes the analysis of links between language and performance difficult to isolate (Dempster & Reddy, 2007).

In the presence of linguistic challenges (as discussed here in terms of differences between home language and LoLT on the one hand and challenges around the use of scientific language on the other) for a learner population that already lacks basic reading skills from the early grades, it has to be asked how readability can be measured. Oakland and Lane (2004) present the strengths and limitations associated with the use of readability formulas in their work. Readability formulas typically estimate the difficulty of text of two surface-level features (i.e. vocabulary and syntax) of language and reading in paragraph text form. According to Oakland and Lane (2004), such formulas do not consider structure-level features (i.e. story structure) that also affect the difficulty of the text. These authors warn that surface-level and structure-level features may be independent and uncorrelated, therefore surface-level formulas can only speak to surface-level features of the text, not structure-level features and vice versa. Lastly, Oakland and Lane (2004) warn that, given the brevity and density of information contained in single test items, readability formulas are likely to yield unreliable

results, and recommend the use of quantitative and qualitative methods of ascertaining text difficulty where readability is one indicator of such difficulty. To illustrate this point, Hewitt and Homan (2004) developed the Homan-Hewitt Readability Formula and applied it in their study across three grade levels. Their results support the belief that the higher the item unreadability, the greater the chance for learners to respond incorrectly (i.e. greater difficulty). Incorrect responses are therefore ascribed to reading problems, not because of lack of content knowledge. However, these authors do not indicate if issues of readability still occur when the difficulty level of content covered by items is accounted for. Whichever method of readability measurement is considered, the fact remains that it is an issue that speaks to the heart of reliability and validity of measurement.

Mullis *et al.* (2013) tested two hypotheses in their analyses of TIMSS mathematics and science items: firstly, Grade 4 learners with high reading ability would not be impacted by the level of reading demand, and secondly, that learners with lower reading ability would perform relatively better on items that required less reading. Mullis *et al.* (2013) analysed the mathematics and science items separately according to reading demands that included the number of words, vocabulary, symbolic language and visual displays. These indicators of reading difficulty were then used to rate the items into low, medium and high categories according to:

- the number of words (anywhere in the item, including titles of graphics and labels);
- the number of different symbols (e.g., numerals, operators);
- the number of different specialised vocabulary words; and
- the total number of elements (density) in the visual displays (e.g., diagrams, graphs, tables).

This method is used in the current article.

5. RESEARCH HYPOTHESES

The main research question asked by the current study is: What is the relationship between the reading demand of selected released TIMSS 2015 items and learners' ability to respond correctly to the items? Similar to the work of Mullis *et al.* (2013), the current study is guided by the following null and alternative hypotheses:

Ho: There are no statistically significant differences between the categorisations of low, medium and high reading demand in terms of the percentage correctly answered.

Ha: There are statistically significant differences between the categorisations of low, medium and high reading demand in terms of the percentage correctly answered.

In this study, a level of significance of 5% is used. If the p-value is less than 0.05, then the null hypothesis is rejected and there are statistically significant differences between the categorisations. On the other hand, if the p-value is greater than 0.05, then the null hypothesis is not rejected and there are no statistically significant differences between the categorisations.

6. METHOD

The data of South African learners used for this paper was taken from the TIMSS 2015 cycle. This is the most recent data that has been released by the IEA. TIMSS 2015 used a stratified two-stage cluster sampling design. In stage 1, schools were selected (from a sampling frame provided by the country's National Research Coordinator) using a stratified sampling

approach according to key demographic variables. Schools were sampled with probabilities proportional to their size. In stage 2, intact classes were selected with equal probabilities within schools. Although TIMSS 2015 assessed Grade 8 learners, South Africa, along with a few other countries opted to assess their Grade 9 learners instead to reduce bunching at the lower end of the achievement scale, thereby making estimation possible.

6.1. Participants and data collection instruments

In the case of South Africa, a total of 12 514 Grade 9 learners from 292 schools participated in TIMSS 2015. Not all learners answered all of the TIMSS 2015 science items due to the matrix sampling approach: a pool of science items were packaged into different blocks with each learner completing two of the fourteen blocks in their booklet. On average 1 698, 1 691 and 1 600 learners responded to the three cognitive domain items, Knowing, Applying and Reasoning, respectively.

These cognitive domains are related to the variety of cognitive skills learners draw on when being assessed. The *Knowing* domain involves the recall of science facts, information, concepts and tools. The *Applying* domain asks learners to apply their science content knowledge to straightforward situations while *Reasoning* extends both the previous two domains through problem-solving familiar and unfamiliar scenarios (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009). The restricted item list had 44 items for *Knowing*, 50 items for *Applying* and 21 items for *Reasoning*. In the interest of time, only 50% of the restricted items were used in this study. They were obtained using a simple random sample, keeping proportions across domains, rendering 22 *Knowing* items, 25 *Applying* items and 11 *Reasoning* items.

6.2. Data analysis

A discriminant function analysis was performed to validate the holistic categorisation of items. Following this, the data were tested for normality and, failing the test for normality (since the p-value for the Kolmogorov-Smirnov test was less than 0.05), nonparametric methods were used for all statistical analyses. The Mann-Whitney test was used for the comparison between two groups, since the Mann-Whitney test is the nonparametric counterpart to the well-known independent samples t-test (Field, 2014:217), and the Kruskal-Wallis test was used for all comparisons of three groups or more, since the Kruskal-Wallis test is the nonparametric counterpart to the well-known ANOVA F-test (Field, 2014:236).

7. CATEGORISING THE TIMSS GRADE 8 SCIENCE ITEMS ACCORDING TO READING DEMANDS

Following Mullis *et al.* (2013), the number of non-subject-specific words, the number of symbols, the number of subject-specific terminology and the number of visual displays were taken into account in order to categorise the reading demand into either a low, medium or high density. Although Mullis *et al.* (2013) used the actual count for the number of words, in this study, a cluster of 10 non-subject-specific words were counted as one element. An example of how and why this was done is given below using one of the restricted items' images from TIMSS 2015 (see Figure 2 for item S01_11, S042195).

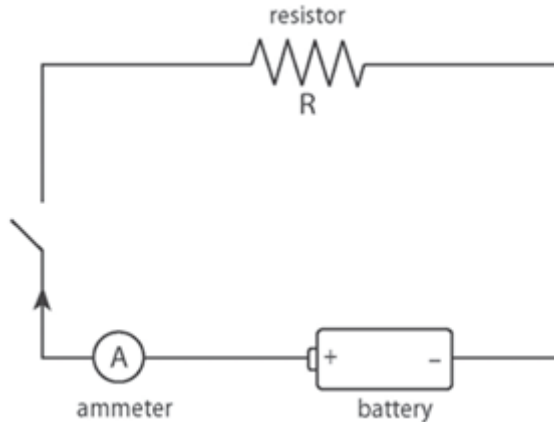


Figure 2: Example of a figure in TIMSS 2015 science

The words *resistor*, *battery* and *ammeter* are not counted as part of the density, following the coding guide by Mullis *et al.* (2013) (see Technical Appendix A: Quantifying the Reading Demands of the TIMSS 2011 Fourth Grade Mathematics and science items) where it is stated that a label should be counted with its object, not separately. The calculation for Figure 2 is, therefore, **six** visual displays (1 resistor, 1 battery, 1 ammeter, 1 arrow, 1 wire, 1 opening) plus **four** symbols (A, R, +, -) to give a total density of **ten**. We believe that it takes longer, and is more difficult in terms of reading demand, to process this total of **ten** against simply reading the first **ten** non-subject-specific words of an item such as: “*For each characteristic in the list below, fill in a...*” Following this reasoning, ten non-subject-specific words were grouped or clustered and counted as **one** element in this study. The fact that pictorials, tables, figures, etc. have a higher reading demand than non-subject-specific words must be addressed when categorising the TIMSS 2015 science items in categories of low, medium or high reading demand.

For the rest of the categorisations, the coding guide by Mullis *et al.* (2013) was strictly followed. For example, for the indicator of subject-specific terminology, words such as *pupa* and *larva* were counted, but not *puppy* as the latter is familiar to most eighth or ninth grade learners in everyday life. In summation, the holistic categorisations in this paper were allocated using the following indicators:

- the number of clusters of 10 non-subject-specific words;
- the number of symbols;
- the number of subject-specific vocabulary; and
- the density of the visual displays.

In order to validate the holistic categorisations of items, a discriminant function analysis (DFA) was performed (Field, 2014:654). According to Slate and Rojas-LeBouef (2011), DFA is appropriate when determining variables that predict group membership. The results show that the indicator *number of clusters of 10 non-subject-specific words* loaded the most heavily on this function (Table 1).

Table 1: Discriminant function analysis: Loading of indicators on discriminant functions

Indicators	Function 1	Function 2
Number of clusters of 10 non-subject-specific words	0.405	0.211
Number of subject-specific vocabulary	0.334	0.125
Density of visual displays	0.302	-0.043
Number of symbols	0.282	-0.262

The DFA classification results are given in Table 2. In order to make sense of the percentages, it is important to note that 20 items we classified as “low”, 20 items as “medium” and 18 items as “high” originally. From Table 2 it can be seen that 85% (17 out of 20) of the low categorisations matched the holistic categorisations, 95% (19 out of 20) of the medium categorisations matched the holistic categorisations and 94.4% (17 out of 18) of the high categorisations matched the holistic categorisations. In total 91.4% of the classifications were performed correctly.

Table 2: Discriminant function analysis: Classification results

Classification results		Readability density	Predicted group		
			Low	Medium	High
Original group	Count	Low	17	3	0
		Medium	1	19	0
		High	0	1	17
	%	Low	85.0	15.0	0.0
		Medium	5.0	95.0	0.0
		High	0.0	5.6	94.4

8. STATISTICAL ANALYSIS

The mean percentages of correctly answered items are given in Table 3 for each cognitive domain along with the minimum and maximum values of the correctly answered items per domain. The highest percentage of correctly answered questions is in the *Knowing* domain (38.1%), followed by the *Applying* domain (17.5%) and the lowest percentage is found in the *Reasoning* domain (12.7%). The minimum and maximum values are also provided and can be interpreted as follows: For the *Knowing* domain, it can be seen that the learner(s) that performed the worst answered 15% of the items correctly and the learner(s) that performed the best answered 66% of the items correctly. Thus, the range for the percentage of correct answers for the *Knowing* domain is 66% – 15% = 51%. This is quite wide, when compared to, say, the *Reasoning* domain where the learner(s) that performed the worst answered only 2% of the items correctly and the learner(s) that performed the best answered only 26% of the items correctly. Thus, the range for the percentage of correct answers for the *Reasoning* domain is 26% – 2% = 24% which is quite narrow.

Table 3: Minimum, maximum and mean percentage of correctly answered items per domain

Cognitive domain	Statistic			
	Number	Minimum % items correctly answered	Maximum % items correctly answered	Mean % items correctly answered
Knowing	22	15.0	66.0	38.1
Applying	25	1.0	57.0	17.5
Reasoning	11	2.0	26.0	12.7

In Table 4, each cognitive domain is further investigated by exploring the frequency and percentage by category of reading demand. This table shows that the majority (63.6%) of the *Knowing* domain was categorised as low, the *Applying* domain were almost equally categorised over medium (36.0%) and high (40.0%) and the *Reasoning* domain was mostly categorised as high (72.7%). It is worth noting that none of the items in the *Knowing* domain has been categorised as having a high reading demand and none of the items in the *Reasoning* domain has been classified as having a low reading demand. This limits the analysis of the impact of reading categories within each cognitive domain, see Table 4 and related analyses.

Table 4: Frequencies and percentages of per domain by category of reading demands

Cognitive domain	Readability category	Frequency	Percentage
Knowing	Low	14	63.6
	Medium	8	36.4
	High	0	0
Applying	Low	6	24.0
	Medium	9	36.0
	High	10	40.0
Reasoning	Low	0	0
	Medium	3	27.3
	High	8	72.7

Turning to reading demand categories, Table 5 indicates the statistics (number, minimum, maximum, mean and standard deviation) per category. For a visual representation, the mean percentage correct for each category of reading demand is plotted in Figure 3 and it is clear that there are differences between the groups. The minimum and maximum values are also provided in Table 5 and can be interpreted as follows: For the *low* category, it can be seen that the learner(s) that performed the worst answered 9% of the items correctly and the learner(s) that performed the best answered 66% of the items correctly. Thus, the range for the percentage of correct answers for the *low* category is 66% – 9% = 57%. This is quite wide, when compared to, say, the *high* category where the learner(s) that performed the worst answered only 2% of the items correctly and the learner(s) that performed the best answered only 26% of the items correctly. Thus, the range for the percentage of correct answers for the *high* category is 26% – 2% = 24% which is quite narrow.

Table 5: Statistics for the percentage correct per category of reading demands

Readability category	Statistics				
	Number	Minimum % items correctly answered	Maximum % items correctly answered	Mean % items correctly answered	Standard deviation
Low	20	9.0	66.0	32.5	18.936
Medium	20	1.0	56.0	26.3	17.885
High	18	2.0	26.0	13.4	8.304

Figure 3: The mean percentage correct per category of reading demands

The question arises whether these differences are statistically significant. Normality could not be assumed [$D(58) = 0.132, p = .013$]. From the histogram (Figure 4) it can be seen that the data is skewed to the right showing not only that the data is non-symmetric, but also emphasising the poor performance of South African learners.

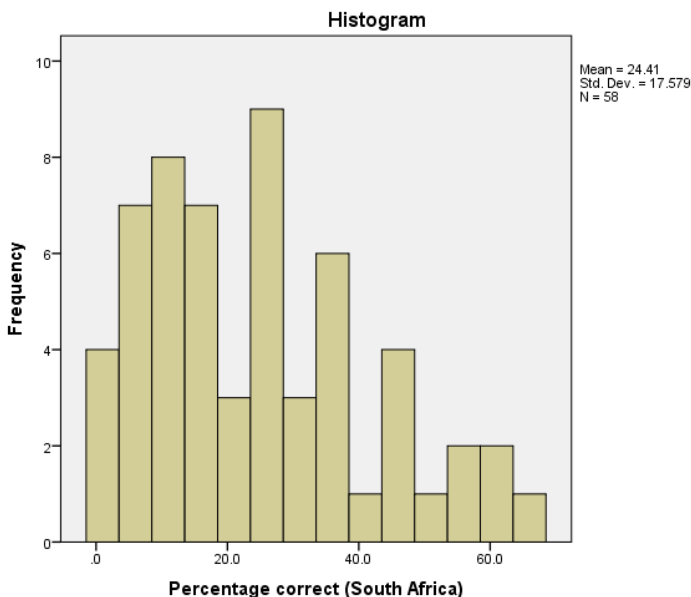


Figure 4: Histogram for percentage correct

The Kruskal-Wallis test, along with the post-hoc Mann-Whitney tests (Field, 2014: 217), was therefore used which indicated significant differences between the categorisations of low, medium and high reading demand in terms of the percentage correctly answered [$H(2) = 11.849, p = .003$]. Post-hoc Mann-Whitney tests showed significant differences in the percentage correctly answered between the low and high categories of reading demands and between the medium and high categories of reading demands (Table 6). The results thus far support the belief that the higher the reading demand, the greater the chance for learners to answer incorrectly. However, reading demand must be separated from content difficulty.

Table 6: Mann-Whitney summarised results

Reading demand categories compared	p-value
Low (mean = 32.5%) and medium (mean = 26.3%)	0.273
Low (mean = 32.5%) and high (mean = 13.4%)	0.001*
Medium (mean = 26.3%) and high (mean = 13.4%)	0.029*

* The p-value is less than 0.05, indicating that the null hypothesis is rejected and that there is a statistically significant difference.

Further analyses were performed in order to test differences between low, medium, and high reading demand categories within each cognitive domain. The results summarised in Table 7 did not show any significant differences since the null hypothesis is not rejected (all p-values are greater than 0.05).

Table 7: Mann-Whitney summarised results

Cognitive domain	Reading demand categories compared	p-value
Knowing	Low (mean = 37.1%) and medium (mean = 40.0%)	0.539
Applying	Low (mean = 21.7%) and medium (mean = 18.8%)	0.555
	Low (mean = 21.7%) and high (mean = 13.8%)	0.368
	Medium (mean = 18.8%) and high (mean = 13.8%)	0.604
Reasoning	Medium (mean = 12.3) and high (mean = 12.9%)	0.919

9. DISCUSSION

Curricular implementation at classroom level has raised the issue of language complexity. South Africa is a recognised multilingual country but it cannot be assumed that the majority of learners have the advantage of home language education when starting their school careers. Language issues are especially compounded in rural schools where learners have fewer opportunities to engage with science content with a solid foundation in English that would allow them access to subject-related vocabulary and terminology. One recommendation is to possibly bridge the gap for learners who have a different home language than that of the LoLT, if for stakeholders, such as provincial and educational officials, the school governing body, parents and other community leaders, are made aware of the study and understand the diverse language situations of communities in order to find ways to allow for some alignment between home language and the LoLT context. A second recommendation is to appoint teachers that are able to speak the learners’ home language in order to provide support to learners whose home language differs from the language of instruction. A last recommendation stems from the fact that, according to the current policy of CAPS, Grade 1 to 3 learners are taught in their mother tongue, whereas from Grade 4 onwards, English is the LoLT (DBE, 2011); if more effective English LoLT teaching and learning support is not available to learners in the early grades, improved achievement in mathematics and science in later grades will remain disappointingly low.

Achievement trends and previous research provided the rationale for investigating the role that reading demands play in South African Grade 9 learners' ability to demonstrate understanding of, and engagement with, released science items from the TIMSS 2015 cycle. It was hypothesised that learners who were able to respond adequately to the randomly selected sample science items were less impacted by reading demand.

At this point, two aspects of readability and content are discussed as possible limitations. While the current results show that language difficulty is associated with performance, there are factors that were not controlled for or investigated in the current study. Firstly, that simple (different than complex, at the other end) topics and work will correlate with and require simple language to discuss and assess, while complex items will inevitably involve more complex language. Secondly, it may be logical and intuitive that better item reading ability and mastery of the subject will lead to better marks, not just the readability of the item. The implication is, therefore, that high-proficiency learners would do well because (1) they understand the work *and* (2) the language of the item. In taking these issues into account, results from the current study could be interpreted not only in terms of item readability, but also readability coupled with the complexity of topics and mastery of the subject as added predictors to the current findings.

As illustrated using the curriculum intentions, implementation and attainment framework, learning science content may be so undermined during classroom teaching by a lack of learner reading proficiency as to be unattainable. The preliminary results of the current study showed significant differences between the percentage correctly answered between the low and high categories of reading demands as well as between the medium and high categories of reading demands. However, further analyses showed that there were no significant differences between the categories of reading demand within each cognitive domain. There are three possible interpretations of these results: 1) reading demand does not have an effect when the same cognitive skills are being tested, 2) reading demand impact is so severe that its effect cannot be separated from content knowledge or, 3) the initial significant observations have now disappeared due to the samples of item-combinations being too small. The first is unlikely given the plethora of research showing an impact of poor literacy on South African learner achievement and in other countries. The authors, therefore, propose that the full impact of reading demand on learner performance cannot be fully evaluated without correlating English proficiency scores but that it remains a plausible factor in the ability of learners to understand questions and present their answers.

As seen by the poor achievement of South African learners in TIMSS 2015, although this has improved over the course of repeated TIMSS assessments, addressing factors that impact achievement such as reading proficiency is crucial. If learners are unable to engage with the content and/or their teachers, they are understandably unable to grasp what is being taught and will do poorly on assessments. However, reading proficiency has far-reaching effects beyond subject-specific achievement. It is expected of the teaching and learning of natural sciences to develop in learners a range of process skills that may be used in everyday life, in the community and in the future workplace. The development of these skills speaks to the development of scientific citizenship, which, for example, means that every learner should develop the ability to read and write proficiently in a clear and 'scientific' manner even if a STEM career is not pursued. A lack of reading and writing proficiency may also impact other factors in academic achievement and scientific citizenship, such as self-efficacy beliefs, self-esteem or engagement during classes.

The ubiquitous effect of reading demand does not only have implications for a subject such as science with its own language and citizenship. As reading forms the foundation of all future learning, the impact of poor literacy development continues to hamper learners as they progress through their academic trajectory. This means that if the basic skill of reading is not adequately developed in the early grades of a learner's school career, coupled with conceptual subject-related gaps and an inability to communicate academically in a proficient manner, it spells continued underperformance and bleak future prospects for a majority of South African learners who come from contextually varied and challenging circumstances.

Linguistic issues like these thus complicate the teaching and assessment of science in later grades considerably. The following recommendations can be made for further analyses. Firstly, there is the need to disaggregate learner achievement by language according to reading demand in further analyses. According to Martin *et al.* (2016), only 12% (SE=2.3) of South African sampled schools reported that more than 90% of their learners responded to the TIMSS 2015 science items in their home language. As much as 80% (SE=2.7) of learners, therefore, wrote the test in English when, in fact, it was not their home language. Expected achievement results for this group of learners can be as low as 342 points (SE=6.7), as opposed to expected results for those learners for whom the language of the test and home language coincided (423 points, SE=17.6). In their work, Dempster and Reddy (2007) found that the maximum readability and comprehensibility were not met in the TIMSS 2003 items, rendering the results invalid for learners with limited English proficiency. Further analyses of the current TIMSS 2015 data by language and reading demand could make findings like those of Dempster and Reddy (2007) more nuanced and indicative of learner abilities across cognitive domains when applying indicators of reading demand and language disaggregation. A second recommendation refers to the use of Rasch analysis in possible further analyses to add scientific and methodological rigour to item analyses, as was done by Glynn (2012). By evaluating science items psychometrically, it could be determined if the science items from TIMSS 2015, in fact, reduced the reading load for learners, thereby making them developmentally appropriate.

The article's findings are considered important because it speaks to the complex language situation we encounter in a multilingual country such as South Africa and the impact that literacy comprehension has on other subjects such as mathematics, and more specifically, science, using reputable international data.

10. DECLARATION

The authors declare that the calculations and the interpretation of the statistics are correct.

REFERENCES

- Center, C. & Niestepski, M. 2014. Hey, did you get that?: L2 student reading across the curriculum. In T. Zawacki & M. Cox (Eds.). *WAC and second language writes: Research towards linguistically and culturally inclusive programs and practices* (pp. 93–111). Unites States of America: The WAC Clearinghouse. <https://doi.org/10.37514/PER-B.2014.0551.2.03>.
- Dempster, E.R. & Reddy, V. 2007. Item readability and science achievement in TIMSS 2003 in South Africa. *Science Education*, 91(6): 906–925. <https://doi.org/10.1002/sce.20225>.

Department of Basic Education (DBE). 2011. *English first additional language (EFAL). Curriculum and assessment policy statement (CAPS)*. Pretoria: Department of Basic Education.

Department of Basic Education (DBE). 2017. *The SACMEQ IV project in South Africa: A study of the conditions of schooling and the quality of education*. Available at: http://www.sacmeq.org/sites/default/files/sacmeq/reports/sacmeq-iv/national-reports/sacmeq_iv_project_in_south_africa_report.pdf [Accessed: 17 May 2020].

Dowd, A.J., D'Sa, N., Noble, F., O'Grady, M., Pisani, L. & Seiden, J. 2018. *Beyond access: Exploring equity in early childhood learning and development*, Washington, DC: Save the Children.

Draper, K. & Spaul, N. 2015. Examining oral reading fluency among rural Grade 5 English Second Language (ESL) learners in South Africa: An analysis of NEEDU 2013. *South African Journal of Childhood Education*, 5(2): 44–77. <https://doi.org/10.4102/sajce.v5i2.382>.

Field, A. 2014. *Discovering statistics using IBM SPSS Statistics*, 4th ed. Thousand Oaks: Sage Publications.

Fleisch, B. 2002. *Managing educational change: The state and school reform in South Africa*. Heinemann, Sandown.

Glynn, S.M. 2012. International assessment: A Rasch model and teachers' evaluation of TIMSS science achievement items. *Journal of Research in Science Teaching*, 49(10): 1321–1344. <https://doi.org/10.1002/tea.21059>.

Gupta, A.F. 1997. When mother-tongue education is not preferred. *Journal of Multilingual and Multicultural Development*, 18(6): 496–506. <https://doi.org/10.1080/01434639708666337>.

Harley, K., Woldie, A. & von Gogh, K. 2019. *Putting communities into the early literacy frame: Learning from A children's storybook writing project in ethiopian community libraries*. Available at <http://hdl.handle.net/11599/3292> [Accessed: 17 May 2020].

Harris, L., Slate, J.R., Moore, G. & Lunenburg, F.C. 2020. Differences in the reading performance of texas grade 4 students as a function of their economic status: A multiyear, statewide analysis. *Asian Journal of Interdisciplinary Research*, 3(1): 83–102. <https://doi.org/10.34256/ajir2016>.

Hemmerechts, K., Agirdag, O. & Kavadias, D. 2017. The relationship between parental literacy involvement, socio-economic status and reading literacy. *Educational Review*, 69(1): 85–101. <https://doi.org/10.1080/00131911.2016.1164667>.

Hewitt, M.A. & Homan, S.P. 2003. Readability level of standardized test items and student performance: The forgotten validity variable. *Literacy Research and Instruction*, 43(2): 1–16. <https://doi.org/10.1080/19388070409558403>.

Howie, S.J., Venter, E., van Staden, S., Zimmerman, L., Long, C., Scherman, V. & Archer, E. 2009. *PIRLS 2006 summary report: South African children's reading literacy achievement*. Pretoria, South Africa: Centre for Evaluation and Assessment.

Howie, S.J., van Staden, S., Tshele, M., Dowse, C. & Zimmerman, L. 2012. *South African children's reading literacy achievement summary report*. Pretoria, South Africa: Centre for Evaluation and Assessment.

Howie, S.J., Combrinck, C., Roux, K., Tshele, M., Mokoena, G.M. & McLeod Palane, N. 2017. *PIRLS literacy 2016 progress in international reading literacy study 2016: South African children's reading literacy achievement*. Pretoria: Centre for Evaluation and Assessment.

- Juan, A. & Visser, M. 2017. Home and school environmental determinants of science achievement of South African learners. *South African Journal of Education*, 37(1): 1–10. <https://doi.org/10.15700/saje.v37n1a1292>.
- Kamwangamalu, N.M. 2000. A new language policy, old language practices: Status planning for African languages in a multilingual South Africa. *South African Journal of African Languages*, 20(1): 50–60. <https://doi.org/10.1080/02572117.2000.10587412>.
- Martin, M.O., Mullis, I.V.S., Foy, P. & Hooper, M. 2016. *TIMSS 2015 international results in science*. Available at <http://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/full%20pdfs/T15-International-Results-in-Science-Grade-8.pdf> [Accessed: 1 November 2015].
- Martin, M.O., Mullis, I.V.S., Foy, P. & Stanco, G.M. 2012. TIMSS 2011 international results in science. Available at <https://timssandpirls.bc.edu/timss2011/international-results-science.html> [Accessed: 10 June 2018].
- Mullis, I.V.S., Martin, M.O., Foy, P. & Arora, A. 2012. *TIMSS 2011 international results in mathematics*. Available at https://timssandpirls.bc.edu/timss2011/downloads/T11_IR_Mathematics_FullBook.pdf [Accessed 17 May 2020].
- Mullis, I.V.S., Martin M.O., Foy, P. & Drucker, K.T. 2012. *PIRLS 2011 international results in reading* Available at <https://timssandpirls.bc.edu/pirls2011/international-results-pirls.html> [Accessed: 17 May 2020].
- Mullis, I.V.S., Martin, M.O., Foy, P. & Hooper, M. 2016. *TIMSS 2015 international results in mathematics*. Available at <http://timssandpirls.bc.edu/timss2015/international-results/> [Accessed 17 May 2020].
- Mullis, I.V.S., Martin, M.O., Foy, P. & Hooper, M. 2017. *PIRLS 2016 international results in reading*. Available at <http://timssandpirls.bc.edu/pirls2016/international-results/> [Accessed 17 May 2020].
- Mullis, I.V.S., Martin, M.O. & Foy, P. 2013. The impact of reading ability on TIMSS mathematics and science achievement at the fourth grade: An analysis by item reading demands. In I.V.S. Mullis & M.O. Martin (Eds.). *Relationships among reading, mathematics, and science achievement at the fourth grade – implications for early learning* (pp. 67–108). Boston College, MA: TIMSS and PIRLS International Study Centre.
- Mullis, I.V.S., Martin, M.O., Goh, S. & Cotter, K. (Eds.). 2016. *TIMSS 2015 encyclopedia: Education policy and curriculum in mathematics and science*. Available at <http://timssandpirls.bc.edu/timss2015/encyclopedia/> [Accessed: 17 May 2020].
- Mullis, I.V.S., Martin, M.O., Kennedy, A.M. & Foy, P. 2007. *PIRLS 2006 International report*. Available at https://timssandpirls.bc.edu/pirls2006/intl_rpt.html [Accessed: 17 May 2020].
- Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y. & Preuschoff, C. 2009. *TIMSS 2011 assessment frameworks*. Boston College, MA: TIMSS and PIRLS International Study Centre.
- Nyström, P. 2008. Identification and analysis of text-structure and wording in TIMSS-items. In *The 3rd IEA International Research Conference*, vol. 12.
- Oakland, T. & Lane, H.B. 2004. Language, reading, and readability formulas: Implications for developing and adapting tests', *International Journal of Testing*, 4(3): 239–252. https://doi.org/10.1207/s15327574ijt0403_3.

- Prinsloo, C.H., Harvey, J., Mosimege, M., Beku, U., Juan, A., Hannan, S. & Zulu, N. 2017. *Trends in international mathematics and science study 2015 diagnostic report: South Africa grade 9 science*. Pretoria, South Africa: Department of Basic Education (DBE).
- Ramnarain, U. 2012. The readability of a high stakes physics examination paper. *Acta Academica*, 44(2): 110–129. <https://doi.org/10.1080/00268976.2011.646415>.
- Reddy, V., Visser, M., Winnaar, L., Arends, F., Juan, A., Prinsloo, C. & Isdale, K. 2016. *TIMSS 2015 highlights of mathematics and science achievement of grade 9 South African learners*. Pretoria, South Africa: Human Sciences Research Council (HSRC).
- Reddy, V., Arends, F., Juan, A. & Prinsloo, C. 2016. South Africa. In I.V.S. Mullis, M.O. Martin, S. Goh & K. Cotter. *TIMSS 2015 Encyclopedia* (pp. 85–99). Chestnut Hill: Boston College.
- Reinikainen, P. 2007. *Sequential explanatory study of factors connected with Science achievement in six countries: Finland, England, Hungary, Japan, Latvia and Russia, study based on TIMSS 1999*. Jyväskylä: University Printing House.
- Setati, M. & Adler, J. 2000. Between languages and discourses: Language practices in primary multilingual mathematics classrooms in South Africa. *Educational Studies in Mathematics*, 43(3): 243–269. <https://doi.org/10.1023/A:1011996002062>.
- Schleppegrell, M.J. 2007. The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly*, 23(2): 139–159. <https://doi.org/10.1080/10573560601158461>.
- Shorrocks-Taylor, D. & Jenkins, E.W. 2001. *Learning from others*. Kluwer, The Netherlands. <https://doi.org/10.1007/0-306-47219-8>.
- Slate, J.R. & Rojas-LeBouef, A. 2011, Calculating basic statistical procedures in SPSS: A self-help and practical guide to preparing theses, dissertations, and manuscripts. Available at <https://archive.cnx.org/contents/52e34675-4ad8-42dc-ad85-8db16e745821@2/discriminant-analysis-assumptions> [Accessed: 16 November 2017].
- Spaull, N. 2011. *A preliminary analysis of SACMEQ III South Africa. Stellenbosch Economic Working Papers: 11/11, A Working Paper of the Department of Economics and the Bureau for Economic Research, November 2011*. Stellenbosch, South Africa: University of Stellenbosch.
- Statistics South Africa 2016, *Community survey 2016 in brief. Report 03-01-06*. Pretoria. RSA.
- van Staden, S., Bosker, R. & Bergbauer, A. 2016. Differences in achievement between home language and language of learning in South Africa: Evidence from prePIRLS 2011. *South African Journal of Childhood Education*, 6(1): 1–10. <https://doi.org/10.4102/sajce.v6i1.441>
- Visser, M., Juan, A. & Feza, N. 2015. Home and school resources as predictors of mathematics performance in South Africa. *South African Journal of Education*, 35(1): 1–10. <https://doi.org/10.15700/201503062354>.
- World Bank 2008. *World development indicators*. Available at <http://www.google.com/publicdata> [Accessed: 14 June 2018].
- Zuilkowski, S.S., McCoy, D.C., Jonason, C. & Dowd, A.J. 2019. Relationships among home literacy behaviors, materials, socioeconomic status, and early literacy outcomes across 14 low- and middle-income countries. *Journal of Cross-Cultural Psychology*, 50(4): 539–555. <https://doi.org/10.1177/0022022119837363>.