What lies behind South Africa's improvements in PIRLS? An Oaxaca-Blinder analysis of the 2011 and 2016 data

MARTIN GUSTAFSSON
STEPHEN TAYLOR

Stellenbosch Economic Working Papers: WP02/2022

www.ekon.sun.ac.za/wpapers/2022/wp022022

June 2022

KEYWORDS: PIRLS, educational improvement, Oaxaca-Blinder, South Africa
JEL: B50, C21, I21

ReSEP (Research on Socio-Economic Policy)
https://resep.sun.ac.za

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH

SOUTH AFRICA

A WORKING PAPER OF THE DEPARTMENT OF ECONOMICS AND THE
BUREAU FOR ECONOMIC RESEARCH AT THE UNIVERSITY OF STELLENBOSCH

www.ekon.sun.ac.za/wpapers

What lies behind South Africa's improvements in PIRLS?

An Oaxaca-Blinder analysis of the 2011 and 2016 data

MARTIN GUSTAFSSON AND STEPHEN TAYLOR

FEBRUARY 2022

ABSTRACT

Three international testing programmes, including PIRLS, point to educational quality improvements in South Africa during the period 2002 to 2019. The gains were substantial, relative to the steepness of improvements seen in other countries. What lay behind these trends? National education quality trends are not easy to explain, and this is seldom attempted in a systematic manner. This is in part because there is little guidance on the optimal methods to follow, methods which must inevitably employ a mix of statistical and non-statistical approaches. This paper offers a brief historical account of major policy and implementation shifts in South Africa's schooling sector, with a focus on the primary level, the level tested by PIRLS. A statistical analysis then employs an Oaxaca-Blinder decomposition, using the PIRLS 2011 and 2016 data in an attempt to identify factors that explain the improvement in reading scores between the two surveys. While this technique adds value to the analysis, there are serious limitations relating to missing values in the background questionnaire data and the fact that these questionnaires are international, and thus do not capture many local policy specificities. When viewed jointly, the historical account and statistical analysis point to improvements in the home background circumstances of learners, including more educated adults and increased access to digital technologies, playing an important role. Certain policy interventions are likely to have played an important role: a large expansion of participation in pre-school education; an increased focus on learning outcomes prompted in part by standardised national assessments; improved initial teacher education; increased provision and use of books in classrooms; and curriculum reforms.

Martin Gustafsson
Department of Economics
University of Stellenbosch
Private Bag X1, 7602
Matieland, South Africa
E-mail: mgustafsson@sun.ac.za

Stephen Taylor
Directorate: Research Co-ordination,
Monitoring and Evaluation
Department of Basic Education
222 Struben Street
Pretoria, South Africa
E-mail: Taylor.S@dbe.gov.za

RESEP
Research on Socio-Economic Policy

A revised version of this paper is expected to be published as a chapter in a book titled *Ten years of South African PIRLS data.*

1 Introduction

While learning outcomes in South Africa's schools remained low in 2019, the last year for which international test data were available (in 2021), improvements have occurred, according to multiple sources, since as far back as 2002. Improvements at the primary level appear to have begun later than at the secondary level, and appear to have stalled in recent years. Specifically, the years 2006 to 2016 saw clear gains at the primary level, the most reliable source for this being PIRLS. This paper focusses on the primary level, using PIRLS 2011 and 2016 data to attempt to explain what lay behind the improvements.

Understanding why learning outcomes within countries improve is important, yet surprisingly little analysis of this has occurred. The methodological challenges illustrated in this paper explain in part why this is the case. But the paper also illustrates, we hope, how new knowledge can be extracted from the available data.

Why learning improved in South Africa can be considered one of the most important questions on the recent history of the country's education system. Understanding history has intrinsic value, while from a pragmatic policy perspective, understanding past improvements should inform planning for further improvements.

South Africa's Department of Basic Education (DBE), the government entity responsible for schools at the national level, has speculated on why learning gains occurred in the country. A plan released in 2020 points to five likely factors: new curriculum implementation guides given to teachers; access to high-quality books; better assessment practices; improved subject knowledge among teachers; and greater pre-school attendance (Department of Basic Education, 2020). The current paper is in part about investigating the extent to which PIRLS can, or cannot, confirm this.

The paper is structured as follows. Section 2 provides background on the magnitude and equity of South Africa's improvements. Section 3 discusses some of the literature of relevance to this paper and justifies the use of the Oaxaca-Blinder decomposition in the statistical analysis. Section 4 discusses policy and contextual changes which occurred in South Africa and which might lie behind the improvements. These changes should inform the way the PIRLS data are analysed. Section 5 discusses the PIRLS datasets used, and how the data were prepared for the analysis. Section 6 presents the statistical analysis. Section 7 attempts to explain the improvements, given what has been presented. Finally, section 8 concludes.

2 The magnitude and nature of the post-2002 improvements

How three international testing programmes, PIRLS, TIMSS[1] and SACMEQ[2], all point to improvements of a roughly similar magnitude within the period 2002 to 2016, has been documented in Department of Basic Education (2020). Subsequent to the release of that document, TIMSS 2019 results for grades 5 and 9 were released. While the Grade 9 results pointed to a continued improvement, the trend for TIMSS Grade 5 mathematics between 2015 and 2019 was flat, suggesting that at least at the primary level, progress was stalling (Reddy *et al*, 2020). This has added urgency to the question of what drives change across South African primary schools.

PIRLS is the only international programme providing very clear measures of national learning outcomes at the primary level at three points in time, namely 2006, 2011 and 2016. SACMEQ, though measuring Grade 6 reading and mathematics outcomes across three years, namely 2000, 2007 and 2013, is limited with respect to the availability of the microdata. Furthermore, a lack

[1] Trends in International Mathematics and Science Study.
[2] Southern and Eastern Africa Consortium for Monitoring Educational Quality.

of technical documentation makes it difficult to interpret certain patterns in the test scores (Crouch and Gustafsson, 2018).

PIRLS is not without problems, however. While the 2006 Grade 5 microdata for the country are publicly available, the 2006 Grade 4 data are not. Most notably, an initial finding, published in the 2016 international PIRLS reports, that South Africa saw no improvement in Grade 4 reading between 2011 and 2016, was subsequently removed from the international reports following a re-analysis of the data initiated by the Department of Basic Education. The problem lay not with the 2016 average score, but with the way the 2011 results had been rescaled in order to make them comparable to the 2016 data. Gustafsson (2020) concluded that of the 43 countries with 2011 to 2016 trends in PIRLS, South Africa displayed the third-steepest improvement, after Morocco and Oman, and an improvement which was comparable to the steepest seen in the international testing systems in general. An unpublished re-analysis by the international PIRLS team produced a steeper 2011 to 2016 improvement for South Africa, which would make South Africa the second-steepest PIRLS improver in this period, after Morocco (Boston College, 2021). Importantly, given South Africa's low point of departure, the country was still under-performing relative to other middle income countries in 2016. Pritchett (2019) provides an important analysis illustrating how South Africa is an exceptional under-performing outlier an international comparison. South Africa needs another decade or so of consistent improvement before the country ceases to be an outlier.

For the reasons given above, the PIRLS 2011 and 2016 data for Grade 4 seem to offer the best data source for examining why South African primary schools improve.

Gustafsson (2020) points to the 2011 to 2016 PIRLS improvement having been largest among the socio-economically most disadvantaged, which in turn indicates a reduction in inequality. Though PIRLS is primarily designed to gauge national trends, to a limited extent it can be used to gauge provincial trends, especially if changes at this level are large. There were statistically significant improvements in five provinces: Free State, Limpopo, Mpumalanga, Northern Cape and North West. In no province was there a clear decline in Grade 4 reading outcomes (Department of Basic Education, 2020: 72).

Table 1 below presents key PIRLS statistics for South Africa. The revised 2011 national figure calculated by the international PIRLS analysts is 278, making the 2011 to 2016 gain 42. The 278 aggregate would make the 2011 to 2016 gain larger than the 2006 to 2011 gain, while the 295 estimate reflected in Table 1, and taken from Gustafsson (2020), results in a larger gain in the earlier period, relative to the later period.

Table 1: Key PIRLS aggregates 2006 to 2016

	2006	2011	2016	2011-2016 gain
Grade 4 comparable means[3]	253	295	320	25
Above without top 20% of learner-weighted schools		262	298	36

The degree to which inequalities with respect to learning outcomes are concentrated *within* schools, as opposed to *between* schools, is important for identifying policy solutions and selecting appropriate analytical approaches. Table 2 presents measures of the proportion of total inequality existing between schools, as opposed to within schools, with respect to PIRLS and TIMSS scores at the primary level. Developing countries tend to have more between-school inequality than developed countries. Notably, in the PIRLS 2016 data the share of overall inequality attributable to between-school differences in South Africa emerges as approximately normal for a developing country. However, the same cannot be said for Grade 5 mathematics,

[3] The 2006 value is discussed in Gustafsson (2020).

where between-school inequality is high. South Africa's between-school inequality with respect to reading, and to a limited extent mathematics, declined between 2011 and 2016.

Table 2: Degree of between-school inequality at the primary level

	All schools		Just bottom four learner-weighted quintiles of schools	
PIRLS	2011	2016	2011	2016
South Africa	0.49	0.33	0.28	0.17
Iran	0.39	0.34		0.25
Morocco	0.35	0.35		0.23
Egypt		0.36		0.21
United States	0.21	0.24		
Finland	0.08	0.09		
TIMSS	2015	2019	2015	2019
South Africa	0.51	0.48	0.25	0.21
Iran	0.34	0.31		0.20

Note: Statistics are intraclass correlation coefficients calculated from the international PIRLS and TIMSS datasets. Blank cells mean mostly that the data existed, but the statistic was not calculated.

In terms of the analysis presented below, the key point about Table 2 is that between-school differences have been large in South Africa, though they might have declined, meaning school-level variables describing what occurs in schools deserve special attention. If one excludes the top one-fifth of schools, in terms of performance, South Africa's proportion of between-school inequality emerges as somewhat low for a developing country. This underscores the fact that South Africa's high *overall* inequality is driven largely by the large socio-economic inequalities between, roughly, the top quintile and the bottom four quintiles of the country. It appears beneficial to devote a part of the analysis of the data to just the bottom four quintiles, as in many ways the challenge is to understand how historically disadvantaged schools improved. This might be best done by excluding the data from the top quintile in the modelling.

It should be pointed out that participation in schooling in South Africa for children aged 7 to 14 has been close to 100% for a couple of decades. This is especially so for the period since 2011, meaning that explanations for qualitative improvements can be sought in the data of the enrolled. Changes in who is enrolled and who is not enrolled have played virtually no role.

3 Methods for explaining national gains in learning outcomes

Pol (2013) provides a useful summary of five approaches used by economists, but not only economists, to identify cause and effect. Explaining national gains in learning outcomes typically draws from at least three of these approaches: the experimentalist, comparative statics and narrative approaches.

The experimentalist approach involves examining statistically how the presence or absence of a factor, affecting individuals or institutions more or less randomly, results in different changes over time. Where the differences are statistically significant, it is concluded that the factor causes the change. This approach has become popular in the last three decades, given the relatively high level of certainty it provides regarding cause and effect. The approach can be used in a controlled experiment, where certain individuals or institutions receive a 'treatment', such as a certain kind of teacher training, while others do not – see for instance the reports of South Africa's Early Grade Reading Study (Taylor *et al*, 2013). A related set of 'quasi-experimental' methods can be used, where the argument is made that a situation similar to an experiment allows for valid causal inferences to be made. For example, historical circumstances sometimes create a natural experiment. One example of the this in the South African context would be Gustafsson and Taylor's (2018) analysis of Grade 12 examination results during a period when provincial boundaries shifted slightly, placing some schools under new administrations. This allowed for an evaluation of the impacts of specific administrations on results, but also more broadly the ability of any administration to make a difference. The experimentalist approach, while valuable in terms of its rigour, focusses mostly on one potential causal factor per study. Moreover, interesting natural experiments with good data are rare. The approach can inform investigations into national gains, but is in itself not sufficient.

The background questionnaires in assessment programmes such as PIRLS are intended to assist in producing 'policy-relevant information about how to improve teaching and learning' (Mullis *et al*, 2017: 4). In part, the idea is that viewing trends in learning outcomes against the background information relating to the tested learners will illuminate why, for instance, learning in certain countries improved. However, how to analyse the data in a manner that moves beyond just descriptive statistics, to cause and effect, is neither easy nor often explained.

Cordero *et al* (2018) provide what is arguably the clearest recent account of how quasi-experimental methods can be applied to international assessment data to explore cause and effect. Only 66 studies from the period 2004 to 2016 that used data from the PIRLS, TIMSS and PISA programmes were considered sufficiently rigorous. These studies, like experimentalist studies in general, tend to focus on one causal factor each, such as class size or degree of school autonomy. A key difficulty, and a reason why there is not more experimentalist use of PIRLS-like data to gauge cause and effect, is that assumptions around the randomness of who receives the treatment must be sufficiently plausible. To illustrate using a South African example, Wills (2020), in analysing SACMEQ data to explore the effects of a protracted teacher strike on learning, acknowledges that a key assumption is that within the same school teachers who participated in the strike are not on average weaker in their teaching abilities than colleagues who did not participate in the strike. Such assumptions are obviously open to question. Only five of Cordero *et al*'s 66 studies attempted to understand learning trends over time in one country. All five dealt with developed countries. On the whole, the remainder of the studies either attempt to explain why some learners in one country perform better than others in one point in time, or examine global patterns with the aim of arriving at fairly generic policy conclusions.

One statistical method which has been employed to understand trends over time, but does not appear within Cordero *et al*'s 66 studies, is the Oaxaca-Blinder (OB) decomposition. This has been used to examine trends in Indonesia (Barrera-Osorio *et al*, 2011), Thailand (World Bank, 2012) and Iran (Network on Education Quality Monitoring in the Asia-Pacific, 2015). The first two make use of PISA data, the third both TIMSS and PIRLS data. The OB decomposition is not designed to identify cause and effect, and would therefore not qualify for Cordero *et al*'s list. Yet its use for the three countries reflects its value specifically for understanding factors behind improvements using PIRLS-like data. The current paper makes use of this method. OB focusses not on one factor, but several. As will be explained further in section 6, OB as applied in this education context draws from two simple multivariate regressions run in two points in time, and decomposes change into two parts: change associated with the *efficiency* of specific inputs, and change associated with the *amount* of each input present. While this is not the same as identifying causes, the OB analysis provides valuable statistics which can inform the search for causal factors. OB fits best into the comparative statics approach listed by Pol (2013): two comparable sets of statistics are used as a basis for theorising what drove the changes in the outputs and inputs over time.

While statistical analysis can greatly enhance the exploration of national gains, the task must inevitably draw from methods used widely by historians, or what Pol refers to as the narrative approach. National gains in learning outcomes, like changes in poverty levels or fertility, or the beginning and end of a war, are too complex to be captured in a single model.

It is striking that there is not more work available exploring, firstly, methods to explain national learning gains and, secondly, the causes behind the gains seen in specific countries. Apart from the three country-specific OB analyses referred to above, and the five developed country analyses covered in Cordero *et al*, one of the few other studies that stands out is that of Bruns (2012), which explores the causes behind large gains in Brazil, using a mix of statistical and historical analysis.

Why is more guidance not available, although participation in international assessment programmes has increased, especially among developing countries? Why is there not more work published in peer-reviewed journals? None of the three OB analyses referred to above were published in such journals. Answering these questions is instructive.

One explanation lies in the fact that academic journals with a strong emphasis on identifying cause and effect have, in an attempt to bring about more rigour, increasingly favoured the experimentalist approach. This could be because it is relatively straightforward to distinguish more rigorous analyses from less rigorous ones within this approach. Distinguishing, for instance, a more rigorous historical narrative from a less rigorous one is arguably a far more time-consuming exercise.

Expectations among policymakers and their critics are high: improvements which would they would consider good tend to be far greater than what history suggests is possible (UIS, 2019). Put differently, there is an insufficient appreciation how national learning gains are more a question of a steady slog, and less a question of highly visible spurts. Important smaller improvements are thus not considered worthy of study, when such improvements, or their absence, should be a critical matter in national development debates. Scepticism about the importance of small annual annual gains would have been exacerbated by a few cases where reported trends in the international testing programmes have been found to be substantially inaccurate. Where South Africa's PIRLS trend had to be adjusted from no gain to a considerable gain, as explained in section 2, PISA gains in Brazil appear to have been over-estimated, though they remain substantial after the necessary adjustments are applied (Carnoy *et al*, 2016). A further factor is perhaps that the limited analytical capacity in the education arena tends to become concentrated around specific techniques and topics, and here the experimentalist work referred to above has been dominant for a few decades. Exploring national learning gains is inevitably multi-disciplinary, involving topics such as the political economy, children's health, poverty, the broader impacts of new technologies as well as what occurs in the classroom. Multi-disciplinary research tends to be especially difficult to bring about successfully.

4 Policy and contextual changes

Figure 1 illustrates key policy and socio-economic developments that could conceivably have had an impact on learning in South Africa's schools in the last quarter century. These developments have been documented in detail in various places, including Gustafsson (2019) and Van der Berg and Gustafsson (2017).

The initial years following South Africa's first democratic elections in 1994 were characterised by intensive work focussing on dismantling the apartheid policy infrastructure. Fifteen education administrations defined along ethnic lines were replaced by nine provincial and one national administration. National policies aimed at equitable teacher salary scales, teacher allocation and school funding replaced the unequal race-based policies of apartheid.

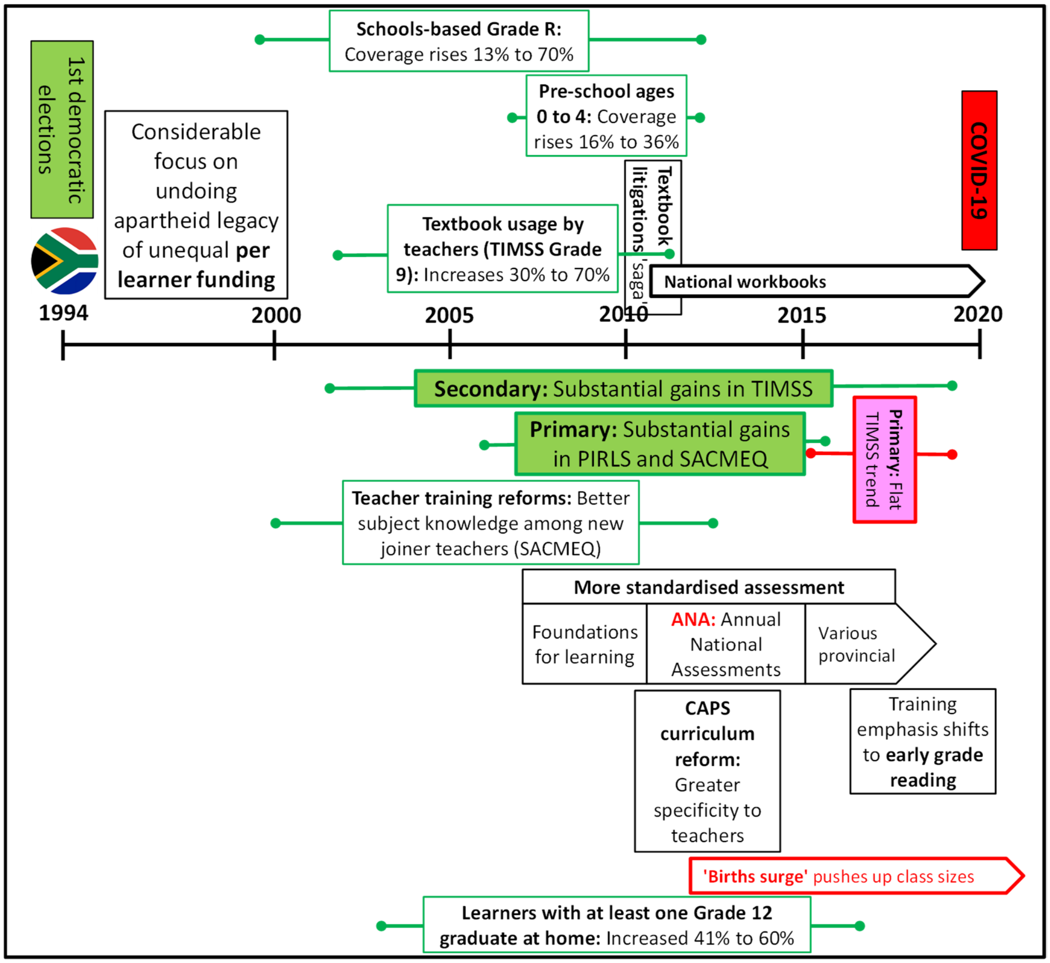
In an effort to professionalise teaching, all teacher training was shifted to universities. While this was criticised as a measure that would slow down teacher supply, teacher test scores from the SACMEQ programme indicate that teachers emerging from exclusively university-based training perform better in terms of subject knowledge.

Given South Africa's relatively late age of admission for Grade 1, pressure grew to accommodate younger children in a grade below Grade 1, known as Grade R. Grade R participation in schools as a percentage of one age cohort rose from 13% to 70% between 1999 and 2012. This would have contributed, for instance, to a longer experience of the school environment for the average Grade 4 learner over the period 2003 to 2016.

Outside of schools, participation in some form of pre-school increased too, particularly in the 2007 to 2012 period, from 16% of zero to four-year-olds to 36% according to household data. While the quality of education and care in these pre-schools has often been questioned, more pre-schooling conceivably improved the readiness of children for school.

The new post-apartheid school curriculum initially de-emphasised textbooks, and favoured resources developed by teachers. A realisation that this was impractical in most school environments led to a renewed prioritisation of textbooks at both the primary and secondary levels. TIMSS data point to the percentage of mathematics teachers in Grade 9 using a textbook as their principal teaching guide increasing from 30% to 70% between 2002 and 2012. This was indicative of similar changes occurring across other grades. Much of the change in learner access to textbooks occurred around 2011, following litigation initiated by civil society groups in response to insufficient state spending on books and weak distribution systems. The so-called national workbooks became increasingly accessible to primary learners. These voluminous and full-colour workbooks, which became the property of learners, were designed to facilitate pacing through the school curriculum. South Africa's national workbooks programme has been considered exemplary by, for instance, UNESCO.

Figure 1: Policy and other changes 1994-2000



Standardised assessment of primary-level children went through various phases. Initially, there was little such assessment, apart from sample-based testing. The release of the 2000 SACMEQ results in 2003, which indicated that South African primary learners under-performed substantially relative to learners in less developed countries spending less per learner, such as Kenya, led to a much stronger interest in assessing learners, in part to identify where in the system teaching and learning was especially weak. The Foundations for Learning initiative, launched in 2008, involved, in part, distributing better learning assessment tools to teachers and emphasising the use of test score data in planning for improvements at the school level. This initiative grew into the very ambitious Annual National Assessments (ANA) programme, involving the testing of learners in grades 1 to 9 and the release of aggregate results in national reports following the style of the Grade 12 examinations report. ANA was run from 2011 to 2015, at which point opposition from teacher unions and some academics led to the halting of the programme. ANA arguably attempted to gather too much information at once, without a sufficient policy basis and without sufficient standardisation of results. Beyond 2015, emphasis on assessing learners continued, though this was largely driven by provinces. The key point is that from around 2008 understanding learning outcomes, as opposed to just the process of teaching and learning, became more important at the primary level than it had previously been. Though the new systems were not ideal, they helped teachers and managers understand the centrality of learning outcomes.

Apart from the national workbooks, two other initiatives shifted teaching practices. Firstly, during the period 2010 to 2014 new curriculum guides, known collectively as the Curriculum and Assessment Policy Statement (CAPS), were issued which provided far more detail on what had to be taught than had previously been the case. Secondly, since around 2016 there has been a growing interest among researchers and teacher trainers in better ways of teaching reading in the early grades. This has been in part been driven by global trends, and the strong emphasis on reading, and measuring reading, in the Sustainable Development Goals of the United Nations.

Two important developments outside the education policy ambit warrant mention. An unexpected and large increase in fertility in around 2004 saw the annual intake into schools rise by over 10% starting in 2011. This resulted in class sizes growing. To illustrate, PIRLS data pointed to the average class size in Grade 4 increasing from 40 in 2011 to 45 in 2016. Classes were large to begin with, and the trend could be expected to affect the teaching process negatively.

Finally, there has been an ongoing improvement in the years of education achieved by adults in the households of learners. For instance, learners in school with at least one adult with a Grade 12 qualification at home increased from 41% to 60% between 2003 and 2017. This conceivably improved the support learners could receive in the home with their schoolwork.

To conclude, there are several possible explanatory factors behind the 2011 to 2016 improvements in Grade 4 reading evident in this brief historical account. Having more pre-schooling is one such factor. The national workbooks may have played a critical role. ANA may have facilitated a stronger focus in classrooms on learning outcomes, but may also have made learners more adept at participating in formal tests. The CAPS guides were rolled out in Grade 4 in 2013, providing more detailed guidance to teachers. Evidence from around the world suggests that these kinds of policy interventions all have the potential to improve learning (UNESCO, 2014). Moreover, and outside the typical menu of recommended policy interventions, improvements in the level of education of parents could have facilitated the learning gains.

5 Data completeness and the selection and transformation of variables

The 2011 PIRLS national sample data covered 15,744 learners in 341 schools. The 2016 data covered 12,810 learners in 293 schools. The data used were those available through the international TIMSS and PIRLS website of Boston College in the United States[4].

The 2011 data included 420 variables reflecting responses to background questionnaires completed by the school principal, the learner's teacher, the learner's parent and the learner. Variables per questionnaire ranged from 159 for the teacher questionnaire to 71 for the learner questionnaire. In 2016, the number of background variables was lower, at 390 in total.

The first step was to identify variables which seemed possible predictors of the 2011 to 2016 improvement, and which also appeared in both years, with identical or sufficiently similar questions asked. A few interesting background questions did appear in one year, but not the other. For example, in 2011 the school principal was asked whether teachers are evaluated by people external to the school. This question was not repeated in 2016. The general principle followed was to change and recode variables as little as possible. Variables were not combined to create composite indicators. Ordinal values, where for instance 1 represented the learner used the school library 'At least once a week' and 4 represented 'Never or almost never', were mostly treated as if they were cardinal values. In this example, the series 1 to 4 would be reversed so that 4 represented the assumed best situation. While not combining variables and changing coding as little as possible might reduce the predictive power of the background variables, this comes with the advantage of a more transparent analysis and a simpler discussion of the results.

[4] https://timssandpirls.bc.edu

An exception was the generation of a learner socio-economic background variable, using a principal components analysis. This variable drew from four binary variables from the learner questionnaire describing access to four household items: a computer; one's own study desk; one's own books; and one's own room. The same parameters were used for both 2011 and 2016 in generating the index, making values comparable across time.

The first step of selecting interesting and comparable variables resulted in the selection of 66 background variables. The 66 includes the learner assets index, and one of the asset variables as a separate variable, namely the private computer variable. It is worth noting that one clearly interesting variable was dropped due to strange values. This was the teacher's response to the question 'Do the … students in the … class have computer(s) available to use during their reading lessons?'. The percentage of learners with this facility declined from 22% to 9%. This decline is unlikely to be real. The most plausible explanation seems to be that in 2016, after the word 'computer' the words 'including tablets' were included in brackets. Responding teachers may have been distracted by this and thus answered the question differently.

While province was not included in the data downloadable from the international TIMSS-PIRLS website, the province of each school was obtained from the University of Pretoria, whose Centre for Evaluation and Assessment is the local implementer for PIRLS in South Africa. Province was then treated as a background variable. Moreover, the number of tested Grade 4 learners in the class, and in the school, were treated as two additional background variables. PIRLS tests whole classes. In 2011, 79% of sampled schools had just one class tested, and in the remainder of cases it was nearly always two classes. In 2016, 91% of schools had just one class tested. The addition of nine provinces and two learner count variables raised the number of background variables from 66 to 77.

The revision of the 2011 PIRLS national average for South Africa, discussed above, was not accompanied by a revision to the five plausible values reflecting each learner's reading competencies in 2011 within the publicly available data (at least this had not been done when we conducted our analysis). For the purposes of the current analysis, values in the data were adjusted using a basic standardisation procedure which forced the overall weighted mean for South Africa, for each plausible value, to equal 295 (see Table 1), and the standard deviation to equal 101 – this standard deviation is described in Gustafsson (2020).

The second step involved rejecting background variables for which there was too much missing data. Examination of the data suggested that a cut-off of 0.84 would strike a balance between the number of variables and the number of learner observations. This cut-off meant that each included variable had to display non-missing information in at least 84% of unweighted observations in both 2011 and 2016. The cut-off resulted in the number of available variables declining from 77 to 42. A key factor determining the threshold was missing school principal responses in the 2016 data. To illustrate, raising the threshold to 0.86 resulted in no school principal responses being included, while 0.84 permitted 12 school principal variables. In the case of the school principal questionnaire, a completely blank returned questionnaire was fairly common. In 2011, 8% of schools returned such a questionnaire. In 2016, this figure was worse, at 12%.

The problem of missing data was thus substantial, and its possible impacts are discussed in section 6.3. The problem is more serious than that encountered by Barrera-Osorio *et al* (2011) with respect to the PISA data. Moreover, as the next table shows, South Africa fares poorly when it comes to ensuring that background responses are collected. Improving this situation in future runs of PIRLS will greatly enhance the research potential of the data.

Table 3: 2016 percentage non-missing background questionnaire responses

	School principal questionnaire	Teacher questionnaire
United States	92	93
Germany	89	88
Italy	91	82
Morocco	93	87
Iran	97	94
Trinidad and Tobago	85	88
Egypt	96	90
South Africa	79	76

Note: Comparator countries are selected not on the basis the completeness of their data. Instead, almost all developing countries were selected, plus three well-known developed countries. For each statistic in the table, the percentage non-missing for each of 82 school principal variables and 137 teacher variables were calculated, Thereafter, the mean across the variables was calculated. Only truly missing values were counted as such. A response such as 'Not applicable' would be considered non-missing.

The South Africa 2019 TIMSS Grade 5 data suggest that there is considerable scope for PIRLS data completeness to improve in future. In the TIMSS data, the school principal questionnaire value for Table 3 would be 94%, with only 2% of returned questionnaires being blank.

Means and standard deviations for 33 of the 42 selected variables, being the 42 minus the nine provinces, are provided in Table 8 in the annexure.

In a third step, learner-level observations were collapsed to the level of the school, with means across learners used, and with learner weights being totalled per school. In the case of the 21% of schools in 2011, and 9% in 2016, with more than one teacher linked to the tested learners, mean teacher responses for the school were used.

Though information is lost by collapsing data to the school level, there are advantages with this route. As discussed previously, the key policy concern in South Africa is why schools perform so differently from each other, meaning a school-level analysis lends itself to the necessary discussion. Moreover, the percentage of observations lost due to missing data is lower for a school-level analysis than for a learner-level analysis. To illustrate, if the nine background variables selected for Table 4 below are used, a learner-level dataset would have to drop around 32% of observations due to missing data, while if the same dataset is collapsed to the level of the school, around 18% of school observations would be dropped in each year. This is largely because collapsing to the level of the school masks the fact that some learner responses may be missing. As long as some such responses exist, they are used to represent all the learners in the school.

6 An Oaxaca-Blinder decomposition

6.1 Basic exposition of a model

The Oaxaca-Blinder decomposition is typically used to uncover unfair discrimination in the labour market, where its logic has been described as follows (Jann, 2008: 453):

> An often used methodology to study labor-market outcomes by groups (sex, race, and so on) is to decompose mean differences in … wages based on linear regression models … [The model] divides the wage differential between two groups into a part that is "explained" by group differences in productivity characteristics, such as education or work experience, and a residual part that cannot be accounted for by such differences in wage determinants. This "unexplained" part is often used as a measure for discrimination …

In the education context, this can be illustrated with a simple example. If the number of books is strongly associated with better test scores, in a statistical sense, and test scores improve over time, then one might expect two things. On the one hand, the number of books could have increased. This would be the ‘explained’ part of the improvement. On the other hand, books may have been better utilised, meaning even with the same number of books, improvements would have occurred. Put differently, the difference between the scores of learners with and without books would have widened. The data are unlikely to reveal how books were better utilised, meaning this is the ‘unexplained’ portion. The Oaxaca-Blinder decomposition provides a sense of how much of the change in the outcome is due to, say, the presence of more books and how much is due to better utilisation of the books.

The last two columns of Table 4 provide outputs, specifically coefficients, for two multivariate regression analyses following an informal selection of nine explanatory variables, from the aforementioned set of 42. The intention here is to demonstrate the Oaxaca-Blinder logic. A formal approach that deals with the important matter of the influence of variable selection is presented in section 6.2 below. The meanings of variables will be discussed further at a later point. Schools included in the regressions because these schools have no missing values in the nine variables, are 276 for 2011 and 243 for 2016 (the totals in the data being 341 and 293). The mean values in the first two columns are means for just the schools included in the regressions.

Table 4: Descriptive statistics and regression outputs (‘demo’ model)

	2011 mean	2016 mean	Diff.	2011 coef.	2016 coef.
Score/Constant	295	323	29	366.97***	82.04
L is girl	0.48	0.48	0.00	74.55	143.50***
L assets	-0.05	0.31	0.36	61.45***	47.44***
L internet	0.26	0.35	0.09	96.93**	33.13
L home comp.	2.83	2.52	-0.31	-39.92***	4.22
T collaborate	2.71	1.83	-0.87	2.37	15.26***
T is female	0.86	0.78	-0.08	10.15	46.01***
T L behind	1.80	1.96	0.16	-11.30*	25.77**
T remedial	0.64	0.91	0.27	5.01	-7.06
T ext. tests	2.22	2.51	0.29	-6.78	11.10
			N	276	243
			R squared	0.536	0.511

*Note: *** indicates that the estimate is significant at the 1% level of significance, ** at the 5% level, and * at the 10% level (also applies to next table).*

Table 5 below provides the outputs of an Oaxaca-Blinder decomposition, using the ‘oaxaca’ command in Stata 16. All the values, though not the significance levels, can be calculated using Table 4. For instance, the change linked to learner asset endowments is the difference in the two means, multiplied by the 2011 coefficient[5]:

$$22.22 = \big(0.31 - (-0.05)\big) \times 61.45 \tag{1}$$

Thus, the gain in the average asset level of learners can be said to contribute around 22 additional PIRLS points to the national mean.

[5] Rounding in the equation means an almost exact result is produced.

Table 5: Oaxaca-Blinder decomposition outputs ('demo' model)

	Endowments (explained)	Coefficients (unexplained)	Interaction	Overall
L is girl	-0.26	33.11	-0.24	32.6
L assets	22.22***	0.67	-5.06	17.8
L internet	8.95**	-16.33	-5.89	-13.3
L home comp.	12.33***	124.86***	-13.63***	123.6
T collaborate	-2.07	34.91*	-11.27*	21.6
T is female	-0.78	30.70**	-2.74	27.2
T L behind	-1.83	66.80***	6.00**	71.0
T remedial	1.34	-7.71	-3.24	-9.6
T ext. tests	-1.98	39.77	5.21	43.0
Constant		-284.92		-284.9
Overall	37.92***	21.87**	-30.87***	28.93***

Turning to the coefficients, or 'unexplained', column of Table 5, the large positive value for the teacher being female is the difference between the coefficients, multiplied by the 2011 mean:

$$30.70 = (46.01 - 10.15) \times 0.86 \qquad (2)$$

This suggests that the impact on learning of having a female changed from almost no statistically significant impact in 2011, to a relatively large and significant impact in 2016. This was not due to there being more female teachers in 2016: the percentage of such teachers declined from 86% to 78%. Instead, the average female teacher had an effect in 2016 which was greater than that in 2011. The words 'impact' and 'effect' are used loosely here. The Oaxaca-Blinder results do not produce firm evidence of cause and effect, as discussed in section 3. In particular, female teachers could be masking some other factor not captured in the data, which is what is actually bringing about the improvement.

The interaction values of the third column of Table 5 represent contributions to the score change which are a result of the interaction between the explained and unexplained. For instance, the interaction for the variable describing the teacher's view on the grade-readiness of learners, *T L behind*, is calculated as follows:

$$6.00 = (1.96 - 1.80) \times \big(25.77 - (-11.30)\big) \qquad (3)$$

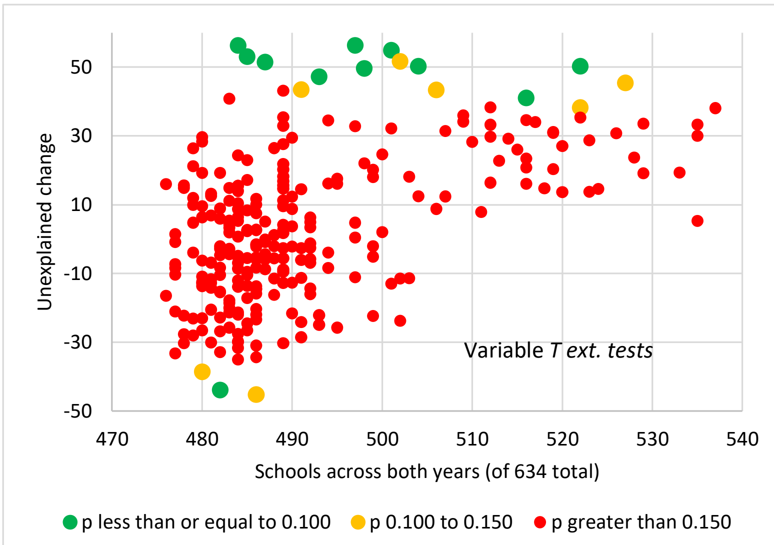
The value for the constant, -284.92, in Table 5 is the difference between the two constant values in Table 4 (last two columns). Column totals in Table 5 provide the overall influence of changes attributable to endowments and the coefficients. Row totals provide the overall influence of each explanatory variable (here there are no significance asterisks as Stata does not explicitly calculate these values). Column totals, and row totals, each produce the same 28.93 value, which is the change in the country's PIRLS score, using just the schools available for the analysis (hence the 28.93 differs from the 25 of Table 1). Clearly, changes in the endowments of the various inputs account for most of the change within this demo model.

6.2 Taking into account variable selection

The Oaxaca-Blinder analysis described above made use of only around 80% of schools in each year. It also made use of nine informally selected explanatory variables. The more variables selected, the smaller the number of schools would become, because different variables have missing data in different observations.

Different combinations of explanatory variables would produce somewhat different results. To deal with this, the Oaxaca-Blinder analysis was run 10,000 times. Each time, just ten explanatory variables were randomly selected, from a set of 41 variables. The aforementioned set of 42 was reduced to 41 after one province was chosen as a reference province and removed. This province was North West. The 0-1 dummy variable for this province seldom explained anything in repeated models. Of the 10,000 analyses run, 2,845 used at least 75% of school observations from each of the two years. To illustrate, among these 2,845 analyses, the variable *T ext. tests*, representing the teacher's response to the question of the emphasis placed on 'national or regional achievement tests', appeared 269 times. These 269 occurrences are represented by 269 points in Figure 2. A focus on the coefficient for *T ext. tests*, which in the 'demo' model of Table 5 emerged as statistically insignificant, reveals that only in 13 of the 269 analyses containing this variable, did the change linked to the coefficient emerge as statistically significant, in the sense that *p* was less than or equal to 0.100 (here the reference is to the significance levels which would appear in the 'Coefficients' column of Table 5).

Figure 2: Distribution of an Oaxaca-Blinder statistic over repeated analyses



It thus appears that changes in the coefficient for this variable do not explain the change in the PIRLS outcomes consistently.

The next table identifies those explanatory variables which do explain the change, either through the endowment or coefficient portion of the Oaxaca-Blinder analysis. The left-hand panel is based on an analysis of all schools. The right-hand panel excludes the top performing 20% of learner-weighted schools, as a key question is what may lie behind improvements seen in the more disadvantaged part of the schooling system. The left- and right-hand panels draw from completely separate analyses, each involving 10,000 runs of the Oaxaca-Blinder model. For an endowment or coefficient value to appear for a variable, two conditions had to be met: (a) the model had to include at least 75% of schools from each of the two years; (b) using just such models, the value had to carry a *p* statistic of no more than 0.150 in at least half of the models. Values in the table are the mean across the various Oaxaca-Blinder values which met the two conditions. Variables not listed in Table 6 are variables for which no value in any of the four columns passed the two conditions.

Table 6: Coefficient and endowment statistics from repeated analyses

	All schools		Top quintile excluded	
	Endowments (explained)	Coefficients (unexplained)	Endowments (explained)	Coefficients (unexplained)
L assets	23.08		14.38	7.46
L home comp.	7.36	111.37	6.74	76.39
L internet	18.28	-25.80		
P T success	4.93			-37.45
T collaborate		43.61		
T groups		-64.86		-41.02
T L behind		50.65	-3.38	61.63
T L disruptive		44.24		37.12
T writing	-5.83		-3.13	
Is FS		2.35		1.96

Running the 10,000 Oaxaca-Blinder analyses a second time produced the same pattern of variables and the same presence of values. Actual values differed, as one would expect given the randomness of the variable selection process, but differences were tiny.

Background variables from the learner and teacher questionnaires – those beginning with 'L' and 'T' – clearly predominate in Table 6. Some statistics display an unexpected sign, for instance those for 'T writing'. Teachers appeared more inclined to have learners write in 2016 than in 2011, yet this explains a *worsening* of PIRLS scores, when one might expect an improvement, hence the negative values in Table 6. An examination of the underlying regression analyses reveals that in both 2011 and 2016 the coefficients on *T writing* are negative, meaning more writing was associated with worse PIRLS results within each year. Similarly, a greater use of same-ability groups by teachers is associated a worsening of results – see *T groups*. How to interpret these patterns is discussed in section 7.

Variables representing judgements of other people, such as the principal's assessment of the success of teachers in *P T success*, and variables representing ratings of one's own professional behaviour, such as *T writing*, were removed from the dataset before re-running the 10,000 analyses for the right-hand panel of Table 6, in other words for quintiles 1 to 4. The result was like that in Table 6, but with the removed variables now absent. No other variables emerged to take their place. The judgement and self-rating variables are thus not 'crowding out' other variables, such as class size, which are more directly amenable to policy interventions.

Table 7: Three-way breakdown of overall change

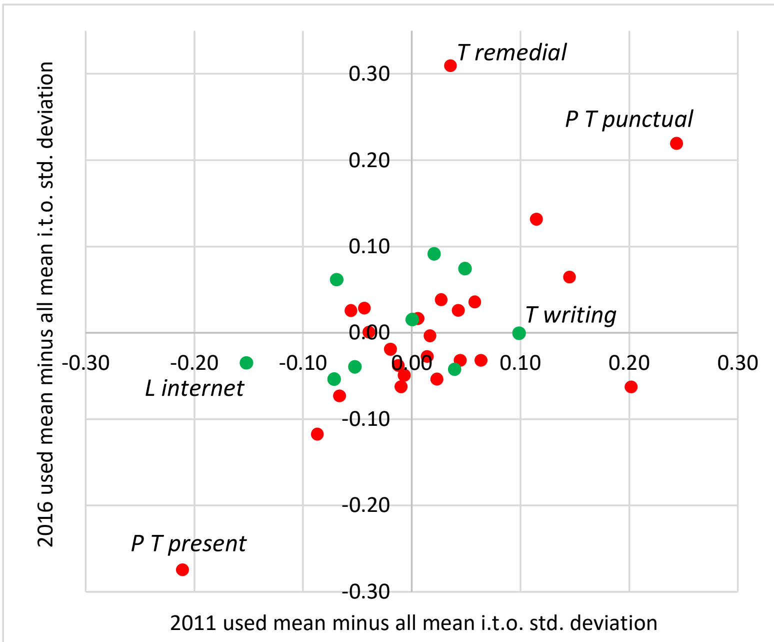
	All schools	Top quintile excluded
Explained	9.48	3.17
Unexplained	28.23	29.96
Interaction	-7.55	-4.47
Total	30.15	28.65

Table 7 above uses data only from models covering 75% of the available school observations. Each value in the table is the mean across the eligible models with respect to the three statistics seen in the bottom row of earlier Table 5. As in the case of the Indonesia and Iran studies mentioned in section 3, most of the change is unexplained in the sense that changes in the regression coefficients could be driven by various unknown factors. However, it is possible to know from Table 6 how the unexplained change is distributed across several explanatory variables.

6.3 Taking into account the non-randomness of missing data

One way of assessing the risk that missing data resulted in biases in the Oaxaca-Blinder analysis, is to compare mean values derived from variables when considered separately, to mean values with only observations used in the analysis included. Each marker in Figure 3 represents one of the 33 non-province variables referred to earlier. Figure 3 focusses on all available schools, with the top quintile included. The closer each marker is to the intersection of the two axes, the smaller the difference between the mean using all the data, and the mean using just observations from the analysis. The ‘used mean’ is, for each variable and year, the average across the means found in the aforementioned 2,845 analyses (each variable would be selected only in some of the 2,845).

Figure 3: Effect of the loss of observations on means



Note: Green markers reflect variables listed in Table 6.

The variable *T writing* can be used to explain the analysis of Figure 3. For 2016, there is virtually no difference between the mean using all the data and the mean derived from the data used in the models. However, for 2011 the mean for the used data is greater than the mean obtained from all the data, the difference being around 0.1 2011 standard deviations. The 2011 standard deviation is 0.77, meaning that instead of the ideal mean of 2.95 (see Table 8), the mean from the used data was 3.03. This is in a context where the *T writing* value increased from 2.95 in 2011 to 3.18 in 2016. Such discrepancies are not worryingly large.

7 Explaining the improvements

The Oaxaca-Blinder results presented in section 6 must be viewed in combination with the historical account provided in section 4. Arguably the most important finding from the Oaxaca-Blinder analysis is that home background factors account for by far most of the improvement in the country’s PIRLS scores, whether one considers all schools, or just performance quintiles 1 to 4. For each model behind Table 6 where the 75% school coverage threshold was reached, the sum across the final ‘contributions’ of the three home background variables, dealing with home assets, a computer at home, and the internet at home, were taken. By ‘final contribution’ we mean what is presented in the last column of Table 5, meaning the sum of the endowments, unexplained and interaction components. In some models not all three of the three home background variables appeared, in which case whatever values were found were used. The mean across the 2,800 or so overall contributions was a little over 30 PIRLS points, whether all schools or just quintiles 1 to 4 were considered. The three home background variables thus easily explain most of the overall gain between 2011 and 2016 reflected in Table 1 and the bottom row of Table 7.

The negative unexplained value for internet access in the home in Table 6 is not easy to interpret, given its close relationship to the possession of a computer. For the purposes of the current discussion, what matters is that taken together, the three home background variables explain so much of the 2011 to 2016 gain.

Do other data support the pattern observed in the PIRLS data that home background circumstances improved? This is a pertinent question given that Statistics South Africa (2017: 15) has found a worsening in household income between 2009 and 2015, while also finding a slight improvement between 2011 and 2015. Other data do corroborate the PIRLS pattern. As discussed in section 4, the education level of the most educated person in the average learner's household improved. This is likely to impact on the support learners can obtain from home. According to General Household Survey (GHS) data, the average learner saw the presence of a computer in the household improve from 17.5% to 19.4% between 2012 and 2016, while the ratio of household members to bedrooms declined (improved) from 2.74 to 2.60 (own analysis of the GHS microdata). The variable *L assets* depends in part on whether the learner has his or her own bedroom. GHS 2012 instead of 2011 data were used as the 2011 GHS data did not cover a computer in the home.

Importantly, home background is associated with the gain in PIRLS performance both through an endowments effect, and an unexplained coefficients effect. Put differently, the national mean score improved both because learners experienced a more enabling home background, and because even in the absence of a change in observed home background factors, those factors were better in 2016 than in 2011 at supporting learners. Some of the unobserved effect could be the improved educational level of adults in the household. While the PIRLS parent questionnaire asks about the highest level of education, the response rate for this question was too low for the relevant variables to pass the criteria used in the analysis.

Being in the Free State (FS) is associated with a small but significant unexplained gain in performance. This province has been found to be relatively well managed. For instance, it is relatively good at preventing overcrowded classes at the primary level (Department of Basic Education, 2020: 106).

All the other variables appearing in Table 6 are based on highly subjective responses. How people respond may change over time, depending for instance on changes in how schools are managed. To illustrate, an increasing emphasis on having learners write in the class could make teachers more inclined to say they encourage writing. This effect may be stronger among teachers who feel less secure about their teaching methods, who perhaps provide responses they believe they ought to provide. This could lie behind the apparent negative impact of more writing seen in Table 6.

If the six variables from Table 6 derived from relatively subjective questions, being the ten variables minus the three home background ones and the Free State indicator, are subjected to the same analysis as that done for the three home background variables, the finding is that jointly the six variables make almost no difference to the gain. Where jointly the three home background variables produced a gain of just over 30, the six subjective variables produce a gain of just 2.9. Put differently, they cancel each other out.

The fact that virtually no factors driven by education policy emerge as important drivers of change is perhaps not surprising. Policy-driven change tends to be country-specific and is thus not easily monitored through international programmes. For instance, it is very likely that learner access to the newly introduced national workbooks, and the way teachers used them, influenced the teaching of reading. Such change would not be captured by the PIRLS questionnaires. PIRLS has no questions that would allow for a measure of the degree of access to books, such as a learner-textbook ratio, to be calculated.

8 Conclusion

What has been offered by this paper is by no means an exhaustive response to the important question of why learning outcomes improved in South Africa's primary schools, at least up to 2016. Analysis of the PIRLS data beyond what is presented here may lead to additional insights. Moreover, it seems important to subject the TIMSS Grade 5 data to a similar analysis to understand the 2015 to 2019 period, when *no* progress in learning outcomes occurred.

It is hoped, however, that the paper will inform future data collection and analysis in ways that will assist in the understanding of national trends. Better response rates for the background questionnaires than what was seen in PIRLS 2011 and 2016 would greatly facilitate future analyses. South Africa's response rates in PIRLS are relatively low. The analysis confirms that the effects of country-specific policy change are not easily captured in large international programmes such as PIRLS. This underscores the importance of also having national sample-based assessment programmes, which follow the methods of, for instance, PIRLS but are more geared towards the national context. In this regard, the planned re-introduction of the Systemic Evaluation programme in South Africa would be a step forward.

The key insight presented in the current paper is that home background factors appear to have played an important role in bringing about the improvements in South Africa's Grade 4 reading results. Even with better data on South Africa's policy interventions, it seems unlikely that this finding would change substantially. That home background improvements have so clearly influenced learning outcomes serves as a reminder that educational improvement needs to be viewed holistically. Achieving this improvement is a responsibility that goes beyond the actors in the education sector, and encompasses planners in other spheres of government, and parents.

References

Barrera-Osorio, F., Garcia-Moreno, C., Patrinos, H.A. & Porta, E. (2011). *Using the Oaxaca-Blinder decomposition technique to analyze learning outcomes changes over time: An application to Indonesia's results in PISA mathematics.* Washington: World Bank.

Boston College (2021). *Scaling the prePIRLS 2011 South Africa data: August 2021.* Chestnut Hill. [Unpublished.]

Bruns, B. (2010). *Achieving world class education in Brazil: The next agenda*. Washington: World Bank.

Carnoy, M., Khavenson, T., Loyalka, P., Schmidt, W.H. & Zakharov, A. (2016). Revisiting the relationship between international assessment outcomes and educational production: Evidence from a longitudinal PISA-TIMSS sample. *American Educational Research Journal*, 53(4): 1054-85.

Cordero, J.M., Cristóbal, V. & Santín, D. (2018). Causal inference on education policies: A survey of empirical studies using PISA, TIMSS and PIRLS. *Journal of Economic Surveys,* 32(3): 878–915.

Crouch, L. & Gustafsson, M. (2018). *Worldwide inequality and poverty in cognitive results: Cross-sectional evidence and time-based trends*. Oxford: Research on Improving Systems of Education.

Department of Basic Education (2020). *Action Plan to 2024: Towards the realisation of Schooling 2030*. Pretoria.

Gustafsson, M. & Taylor, S. (2018). Treating schools to a new administration: Evidence of the impact of better practices in the system-level administration of schools. *Journal of African Economies,* 27(5): 515-537.

Gustafsson, M. (2019). Pursuing equity through policy in the schooling sector 2007-2017. In Spaull, N. & Jansen, J. (eds.), *South African schooling: The enigma of inequality*. Cham: Springer: 47-66.

Jann, B. (2008). The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4): 453-479.

Mullis, I.V.S., Martin, M.O., Foy, P. & Hooper, M. (2017). *PIRLS 2016 international results in reading*. Chestnut Hill: Boston College.

Network on Education Quality Monitoring in the Asia-Pacific (2015). *Factors influencing changes in learning achievement: Iran's performance in TIMSS and PIRLS (2003-2011)*. Bangkok.

Pol, E. (2013). Causality in economics: A menu of approaches. *Journal of Reviews on Global Economics*, 2: 356-374.

Pritchett, L. (2019). *A review essay: The politics and governance of basic education: A tale of two South African provinces.* Oxford: Research on Improving Systems of Education.

Reddy, V., Winnaar, L., Juan, A., Arends, F. *et al* (2020). *TIMSS 2019: Highlights of South African Grade 5 results in mathematics and science.* Pretoria: HSRC.

Taylor, S., Cilliers, J., Prinsloo, C., Fleisch, B. & Reddy, V. (2017). *The Early Grade Reading Study: Impact evaluation after two years of interventions*. Johannesburg: JET.

UIS (2019). *How fast can levels of proficiency improve? Examining historical trends to inform SDG 4.1.1 scenarios*. Montreal.

UNESCO (2014). *Education for All global monitoring report 2013/4: Teaching and learning: Achieving quality education for all*. Paris.

Van der Berg, S. & Gustafsson, M. (2017). *Quality of basic education: A report to Working Group 1 of the High Level Panel on the Assessment of Key Legislation.* Cape Town: Parliament.

Wills, G. (2020). Teachers' unions and industrial action in South African primary schools: Exploring their impacts on learning. *Development Southern Africa,* 37(2): 328-347.

World Bank (2012). *Learning Outcomes in Thailand: What Can We Learn from International Assessments?* Washington.

Annexure: Descriptive statistics for 33 explanatory variables

Table 8: Means and standard deviations

	2011		2016		2011-2016 change		2016 std. deviation	
	All	Q1-4	All	Q1-4	All	Q1-4	All	Q1-4
L assets	-0.03	-0.23	0.27	0.13	0.3	0.4	0.65	0.59
L home comp.	2.84	2.82	2.51	2.44	-0.3	-0.4	0.63	0.64
L internet	0.26	0.21	0.33	0.29	0.1	0.1	0.22	0.20
L is girl	0.48	0.48	0.48	0.47	0.0	0.0	0.10	0.08
P area	0.55	0.51	0.58	0.52	0.0	0.0	0.50	0.50
P pre-school	0.47	0.44	0.50	0.48	0.0	0.0	0.27	0.27
P infrastructure	2.43	2.29	2.47	2.30	0.0	0.0	1.08	1.03
P library	0.41	0.33	0.39	0.31	0.0	0.0	0.49	0.46
P lib. stock	2.57	2.53	2.56	2.53	0.0	0.0	1.16	1.17
P materials	2.43	2.31	2.48	2.32	0.0	0.0	1.08	1.01
P parent involve.	2.62	2.53	2.75	2.68	0.1	0.1	0.95	0.92
P T skills	2.64	2.59	2.67	2.66	0.0	0.1	1.10	1.09
P enough space	2.36	2.26	2.46	2.28	0.1	0.0	1.14	1.12
P T present	3.27	3.23	3.40	3.31	0.1	0.1	0.75	0.78
P T punctual	3.06	2.98	3.20	3.14	0.1	0.2	0.81	0.81
P T success	2.56	2.47	2.77	2.67	0.2	0.2	0.72	0.70
T age	44.62	44.68	45.59	46.06	1.0	1.4	9.56	8.50
T class test	2.73	2.74	2.76	2.76	0.0	0.0	0.44	0.45
T collaborate	2.71	2.71	1.83	1.75	-0.9	-1.0	0.79	0.75
T ext. tests	2.23	2.29	2.51	2.51	0.3	0.2	0.59	0.59
T homework	3.55	3.53	3.47	3.48	-0.1	0.0	1.09	1.09
T groups	3.14	3.21	3.15	3.22	0.0	0.0	0.78	0.75
T remedial	0.61	0.58	0.78	0.77	0.2	0.2	0.41	0.42
T dev hours	10.91	11.02	20.09	20.30	9.2	9.3	18.27	18.37
T is female	0.84	0.86	0.79	0.76	-0.1	-0.1	0.40	0.42
T L behind	1.81	1.79	1.97	1.96	0.2	0.2	0.43	0.44
T L disruptive	2.00	2.03	2.03	2.04	0.0	0.0	0.57	0.58
T lib. corner	0.70	0.66	0.54	0.50	-0.2	-0.2	0.49	0.49
T read aloud	2.61	2.69	2.69	2.74	0.1	0.0	0.50	0.48
T safe	3.27	3.18	3.30	3.26	0.0	0.1	0.83	0.84
T writing	2.95	3.05	3.18	3.25	0.2	0.2	0.72	0.68
In class	38.14	39.83	41.22	43.38	3.1	3.5	14.44	14.67
In school	42.87	44.36	44.63	46.37	1.8	2.0	18.14	18.42