Standardised testing and examinations at the primary level

Current thinking and practices

Martin Gustafsson
20 April 2018

Summary

A point of departure for the current report is the National Development Plan's call for '**reliable measures**' of learning outcomes per primary school to target support to schools and advance school accountability and effective planning in general. Specifically, the NDP calls for testing of one primary level grade on an annual basis to produce these 'reliable measures'.

Standardised testing, or learner assessments that will allow comparisons across, say, provinces and schools, and over time, is arguably the most contentious policy issue in basic education, not just in South Africa but around the world. **The aim of the current report is *not* to advance a particular policy position.** Instead, it is to provide an account of trends with respect to testing systems around the world, and the advice of global organisations which provide policy leadership. As is shown by the report, these organisations differ in their positions on standardised testing in many respects, in some cases to a large degree. Why this is the case is to some extent explored.

The underlying assumption of the report is that **the differing policy positions need to be understood by all who are engaged in the area of assessment policy**. Without understanding the policy narrative of, say, *both* Education International (the world federation of teacher unions) and UNESCO, it becomes difficult to participate constructively in the debates.

Turning to the NDP's call for 'reliable measures', there are **various possible policy responses**. A strictly correct response would be to re-introduce some improved form of the Annual National Assessments programme, with a special focus on one grade in every school. Grade 6 is sometimes considered the optimal grade in this regard. One option might be to make the existing year-end examination in Grade 6 a national examination, as opposed to having an ANA-like and externally administered set of tests. An alternative would be to rely on a sample-based testing system able to gauge trends down to the province level. As an interim measure, or to supplement a system of standardised testing, one could track trends in the data on school-based assessment (SBA), data which are available, but mostly under-utilised. Clearly, there are many conceivable options.

The various options all come with **complex advantages and disadvantages** relating to the availability of technical expertise in South Africa, financial costs, the scepticism of many stakeholders towards standardised testing, and sensitivities around what standardised tests may reveal about individual schools, districts or provinces.

A fundamental question which must be answered is whether there is **a global shift away from the kinds of standardised testing programmes which have been emphasised strongly in many countries for decades**. Certain sources do create the impression that this is the case. There is certainly considerable scepticism towards standardised testing, and this could create the impetus for a shift away from testing.

Perhaps the most useful expression of this scepticism is found in the resolutions of **Education International (EI)**. Its 2011 World Congress resolutions expressed **clear concerns about the *way* standardised testing is implemented by governments**. One could even interpret the

resolutions as reflecting an *a priori* opposition to standardised testing by EI, but this is debatable. Standardised testing, or certain types of it, is seen by EI as a part of the dangerous advance of neoliberalism, which also includes the privatisation of education and the 'casualisation' of labour in the sector. The EI's position on testing is in part, and justifiably, influenced by experiences of poor test design and clear mismanagement by Pearson, a company which provides testing services to a large section of the schooling system in the United States. This backs up the EI's concerns around the role of private players in large-scale assessments, coupled to weak controls over service providers by public authorities. The resolutions from EI's next World Congress, of 2015, makes no mention of standardised testing as such, though its positions on other issues such as privatisation and casualisation remain strong. It is not clear whether the omission of testing from EI's most recent resolutions represents a policy shift, though it would be fair to assume that this may be the case.

The **World Economic Forum (WEF)** has not published much on the matter of testing, or education policies in general, but what has been released by this influential body can be interpreted as **advice favouring a move *away* from testing**. Specifically, the WEF released, in 2017, an education policy position paper framed within the 'Fourth Industrial Revolution', meaning radical economic shifts occurring as a result of technology change, shifts which have a profound impact on the labour market and the demand for skills. The 2017 paper calls for a rethinking of the role of standardised testing as a means of improving educational quality. Moreover, it puts the 'Finland model' forward as an example countries should follow. This model, which has also been promoted outside the WEF, favours considerable teacher autonomy in deciding on, for instance, how learners are assessed, and de-emphasises standardised assessments (though Finland has a sample-based standardised testing programme aimed at gauging national trends). What is a bit confusing is that the WEF paper also calls for more 'competency-based credentialing', something which one assumes requires standardised assessments.

Three organisations, **UNESCO, the World Bank and the OECD,** organisations which are highly influential in policy circles and whose policy analyses easily run into the hundreds or thousands, adopt a position on standardised testing which is very different to that of the WEF (or EI). **All three organisations emphasise above all the importance of having good and comparable data on the performance of specific levels of the schooling *system***. To some extent, they consider comparable data on learning outcomes *per school* necessary for monitoring and management purposes. In other words, the three organisations clearly promote standardised testing, including in their most recent position papers. They draw extensively from evidence of what works, and what does not, in arguing that standardised testing is necessary, but go on to argue that how it is implemented and positioned within the broader range of education policies is critical for understanding the likelihood of a positive impact on learning outcomes. The World Bank acknowledges that the volume of standardised testing in, say, the United States seems excessive, while concluding that across most of the world there is too little, not too much, testing.

The **dominant policy advice from the global organisations thus remains that standardised testing is important**, while the design of testing systems must be appropriate and alignment across different policies is crucial.

This may be the predominant advice, but are countries following this advice? The evidence suggests that they are, and that **an increasing number of countries are adopting standardised testing programmes**. This is true both among rich OECD countries and developing countries. Of great importance, given China's size, is the introduction of a sample-based national assessment system in this country in 2015.

A key policy question is whether one should pursue a sample-based national assessment system, or a universal (or census-based) system. Here the policy advice does not lean strongly

in either direction. However, it is noteworthy that **most countries do appear to test or examine all learners, and not just a sample**, though some national system, be it a national assessment or an examination, at the primary level. A UNESCO database of 68 developing countries shows that 70% of these countries implement either a universal national assessment, or a national examination, at the primary level. An OECD list of 23 mostly rich countries running national assessments at the primary level indicates that only four rely only on sample data (one of the four is Finland).

One argument in favour of sample-based testing is that this approach permits better controls over the quality of the data, and hence comparability over time. While this is true, it is noteworthy that two developing countries, Brazil and Chile, appear to have succeeded over many years to produce **comparable school-level results, within *universal* systems**, by exercising sufficient controls, above all in the form of external test administrators. Chile's system is more effective than that of Brazil's insofar as it takes into account the socio-economic background of learners when drawing conclusions around the effectiveness of schools. Taking socio-economic status into account in this manner is something the literature emphasises strongly.

A critical question if automatic grade promotion is not pursued, is what information should guide the retention of learners. A paradox found in developing countries is that thresholds for adequate performance built into standardised assessment systems tend to be relatively high, implying that very large numbers of learners would have to repeat their grade if the principle were followed that learners should be ready for the next grade. Three countries examined in some depth – Brazil, Chile and Ghana – all **deal with this problem by implementing automatic grade promotion**, or something very close to this, at the primary level.

South Africa is clearly an outlier in the Southern African Development Community (SADC) when it comes to standardised assessments at the primary level. Other than during the years of the ANA programme, 2011 to 2014, there has been no standardised testing or examination at the primary level in South Africa. In contrast, virtually **every other SADC country tests or examines every primary-level learner through some national system**. The exception is Angola, which is still building its education system in the wake of its civil war. None of the 13 SADC countries which tests or examines its learner relies solely on a sample, and in 12 of these countries a national examination at the primary level exists – the one exception is Namibia, whose universal testing system is not considered an examination by UNESCO. Of the 12 SADC countries with examinations, eight *also* run national assessments, mostly sample-based.

Were South Africa to rely just on a sample-based system to gauge learning outcomes at the primary level, it would **remain an outlier in SADC as all other countries (bar Angola) test everyone**. However, it would not be alone in Africa. Ghana, one country examined in the current report in some detail, has a relatively advanced sample-based national assessment system at the primary level, and no examination (though Ghana does have a Grade 9 national examination).

Contents

1 Introduction

The National Development Plan (NDP, p. 311) says the following, under the heading 'Proposals for results oriented mutual accountability' and then the sub-heading '**Reliable measures**':

> Externally administer and mark the ANA for at least one primary school grade to ensure that there is a reliable, system-wide measure of quality for all primary schools. This will serve as a snapshot of the health of the system and help authorities to develop targeted interventions.

The discussion around this call in the NDP makes it clear that this is intended to enhance accountability, including **school-level accountability to parents**, but also to assist in targeting interventions to the right schools and gauging the extent which individual schools' learning outcomes are improving.

Given that the Annual National Assessments (ANA) were suspended in 2015, the question is how to interpret this proposal in the NDP. **Various possible responses exist.** One is to re-introduce a revised 'universal ANA' without the shortcomings that contributed to the 2015 suspension. Another is that it is sufficient to have a systemic and sample-based testing programme that measures trends in the system as a whole – in fact, unions and government agreed to such a programme during 2017[1]. Some have said that making the Grade 6 year-end examination, at least for a couple of key subjects, a national examination with external marking is the solution. Others have said that school-based assessment (SBA) data can be used to develop measures of school performance.

A key question is what the current thinking and practices around the world are with respect to standardised testing and examinations at the primary level. This paper addresses this question by looking at the most recent advice of global 'thought leaders': UNESCO, the World Bank, the OECD, the WEF, and Education International. Moreover, it looks at the current practices of, firstly, a few countries sometimes held up as exemplary in the area of education and, secondly, the countries of the SADC. The following sub-questions in the overall question stand out:

[1] See *The development of a system evaluation model for the basic education*, dated 11 November 2017.

1. How does one achieve **sufficient comparability across schools at one point in time, and across the same school at different points in time**? Obviously without sufficient comparability, a testing system comes to be viewed as unfair, improvements can erroneously come to be viewed as a deterioration, and vice versa.

2. How is the **trade-off between universal testing covering all schools and sample-based testing** managed? Universal testing carries the obvious advantage that one obtains statistics per school, which can then be used in engaging with schools for support and accountability purposes. However, many countries limit their national testing to a sample-based programme, largely to reduce costs, or because the schooling system is not ready, in terms of capacity or the politics of the system, to implement universal testing. Not only is universal testing more costly, it is also comes with greater risks that aggregate statistics at the national and sub-national levels will not be comparable. This is because the techniques that are necessary for sufficient comparability over time, in particular secure anchor items, become more difficult the greater the number of schools covered, because security becomes more difficult to maintain.

3. How does having standardised testing or examinations **contribute towards educational quality**? In particular, how are test results used to enhance the quality of schooling?

4. How are learner results used, or not used, to determine **promotion into the next grade**? A key contradiction in developing countries is that thresholds for what is considered sufficient performance are often not synchronised with thresholds actually used to determine grade promotion, the latter thresholds being considerably lower. This is mostly unavoidable. For instance, no schooling system wants to declare minimum literacy standards that look low, yet the reality is that actual teaching and learning is so weak in many schools that learners not reaching these standards must be promoted, to avoid unacceptably high levels of grade repetition.

The above questions guided the focus. Given time constraints and limitations in what is publicly available, there are inevitably gaps in what follows, yet it should provide adequate background information to guide the South African policy debates.

2 Answers to a few essential South Africa questions

Before the main analysis, there are few important South Africa-specific questions which can be answered empirically, and which influence the whole policy debate[2].

2.1 How standardised are SBA results?

The whole push for standardised testing and examinations is based on the premise that assessments conducted by schools, using the national curriculum but with teachers being allowed to determine their own assessment instruments and conducting the marking, produces **results which are not sufficiently standardised** for one to determine, for instance, that mathematics in taught better in one school than another.

But do school-based assessments (SBAs) really not produce results comparable across schools (or time)? The little evidence that exists in South Africa suggests the answer is no. A 2008 research paper focussing on the secondary level in Western Cape found that whether one

[2] One question which one could consider essential, but which is not addressed here, is the effectiveness of the Western Cape's universal standardised testing system. As seen in the case of Brazil (section 8.1), national policies on assessment can draw from regional or provincial experiences. A fairly comprehensive high-level evaluation of the Western Cape's system is available in a 2017 report produced for the National Treasury by the author of the current report – title *Generation and use of data in the Western Cape in the delivery of basic education* (28 February 2017).

was promoted to the next grade in a school was a bit like a 'lottery' in the sense that this was poorly correlated with actual competencies measured outside the school through a household survey[3]. A 2008 report published by Umalusi found that the correlation between Grade 12 SBA results and results from the national examinations were unacceptably low, a finding which contributed to a reduction of the weight of Grade 12 SBA results in the final Grade 12 mark for each subject[4]. There is no reason to believe the situation would be different at the primary level. A 2015 report, not publicly available, commissioned by the Western Cape Education Department finds that at the primary level, **standardised test results (from the provincial testing system) and SBA-based decisions on who is promoted are poorly correlated**[5].

2.2 What sample size is required to monitor district trends?

Monitoring trends at the district level is obviously important. There is often confusion about how best this is done because what constitutes a sufficient sample size is not well understood. Contrary to what many believe, **it is not the *percentage of the population* that needs to be considered, but *an absolute number of units*** (for instance schools). This is why Botswana and the United States have roughly the same sample sizes in the TIMSS programme, though the latter country has roughly 100 as many schools and learners. One needs around a hundred schools to obtain fairly reliable performance averages for a district. Thus in a district with 100 schools, one would need to 'sample' 100 schools. To illustrate, Seychelles is such a small country that in the SACMEQ programme the 'sample' consists of *all* the country's schools. Thus a *sample*-based testing system which produced sufficient performance statistics for all the 70 districts would need to collect data from at least 7,000 schools (70 times 100), or over a third of the country's 17,000 primary schools. It would be difficult to 'sell' such an arrangement to the public and stakeholders, who are bound to ask why one does not simply cover all 17,000 schools, as opposed to 7,000 schools. The 2017 agreement concluded with unions does not address the matter of sample size, and it is not clear that the aim is of the programme would be to gauge trends at the national and provincial levels only (this is presumably the aim, however, and it was the aim in the old Systemic Evaluation).

2.3 Were ANA scores sufficiently comparable across years?

Testing theory suggests strongly that the ANA national aggregates would not be comparable across years given that secure anchor items, and subsequent mark adjustments using item response theory (IRT) were not employed. In fact, it is easy to establish that **published ANA results produce trends which are too 'jumpy' and with differences which are too large to be an accurate reflection of what was actually occurring in schools**. To illustrate, in Grade 6 mathematics, declines across all provinces in the average ANA mark between 2011 and 2012, and then the subsequent improvements between 2012 and 2014, were all larger, in terms of multiples of standard deviations, than one would find using tests which are truly comparable. For instance, Free State's average score improvement between 2013 and 2014 was around five times as large, in terms of standard deviations, as the fastest annual improvements found in truly standardised tests around the world. These findings about ANA hold whether one uses the universal data or the data from the 'verification sample'. The verification sample was simply designed to verify whether universal results were reliable *within* a year, not to check the year-on-year comparability of results[6].

[3] Lam, Ardington and Leibbrandt, 2008.
[4] Van der Berg and Shepherd, 2008.
[5] Van der Berg, Van Wyk, Van Broekhuizen *et al*, 2015.
[6] Department of Basic Education, 2016: 33.

3 UNESCO (and the SDGs)

UNESCO's Global Monitoring Report, a large annual report on education released since 2002, is perhaps the best source available on the current policy consensus and trends relating to strengthening the schooling of children around the world. Each report zooms in on a specific theme, though there are core themes which recur in all reports. For the current report, two issues of UNESCO's report have particular relevance: the 2014 report focussing on educational quality (title *Teaching and learning: Achieving quality for all*) and the 2017 report focussing on accountability (*Accountability in education: Meeting our commitments*)[7].

Both the 2014 and 2017 reports **stress the importance of standardised testing**, including testing at the primary level, but also offer a few warnings. One warning relates to the definition of a 'national assessment', and the difference between this an examination. An examination can or should be used to determine grade-on-grade promotion, while a national assessment should be entirely diagnostic, in the sense that it diagnoses the *system*[8], where 'system' can mean the whole country, or even an individual school, but not a specific learner. Both universal and sample-based testing qualify as 'national assessments' in UNESCO's report. Universal systems referred to include those of Brazil and Uganda, while Finland and Indonesia are among many countries said to conduct sample-based national assessments. All these are examples from the primary level. Uruguay is an example of a country that mixes a universal and sample-based approach:

> Uruguay followed up an initial assessment of grade 6 pupils' language and mathematics learning in 1996 with sample assessments every three years. The assessment material and training sessions were also offered to non-sampled schools so that these schools could conduct the tests themselves; about 80% of schools per year volunteered to do so…[9]

The conclusion that **Brazil's universal national assessments** contributed to large educational improvements is reiterated in the 2014 report (this conclusion can be found in several other reports, by UNESCO and others):

> The ultimate test of national assessments as a diagnostic tool is whether the results are used effectively to help education ministries strengthen the policy mix so as to improve education quality and learning outcomes. Many national assessment systems are lacking in this respect. Brazil is an exception, having used national assessments to significantly improve education quality, especially for disadvantaged groups[10].

Good national assessments, be they universal or sample-based, are considered essential for all countries wishing to improve education outcomes. **The trend is for more countries to implement these assessments.** Japan, one of the few rich countries not to have had a national assessment, introduced a universal one in Grade 6 in 2007[11]. Very significantly, the Chinese authorities have partnered with a range of local and foreign partners to develop a sample-based system which was implemented for the first time in 2015[12]. By around 2015, about 40% of high income and upper middle income countries had what UNESCO would consider a national assessment at the primary level, with the figure rising to 53% and 66% for lower middle and low income countries respectively[13]. These figures would exclude countries with examinations but not national assessments. The fact that poorer countries have a greater probability of running these tests seems to be reflection of both relatively low levels of

[7] UNESCO, 2014; UNESCO, 2017.
[8] UNESCO, 2014: 6.
[9] UNESCO, 2014: 288.
[10] UNESCO, 2014: 90.
[11] UNESCO, 2017: 133.
[12] UNESCO, 2017: 127.
[13] UNESCO, 2017: 121.

learning at the primary level, a fact which makes monitoring more urgent, and the leverage of UNESCO and other development agencies in these countries.

The UNESCO Institute for Statistics (UIS) has been maintaining an international database of assessments and examinations found in 68 developing countries[14]. Around half the records in this database are 'assessments', the other half being 'examinations'. Of the assessments, which are all applied at the primary level, and which are found in 55 countries, **one-fifth are universal, or a 'census', with the remainder being sample-based**. However, it should be noted that of the countries with sample-based national assessments, half also run a national examination at the primary level. The bottom line is that of the 68 developing countries in the database, 49 test or examine *all* learners through some national system, an examination or assessment, at the primary level.

There are warnings against poorly designed national assessments. Nigeria's system is said to produce results which are not sufficiently reliable to guide policy. Comparability of results across years is considered important. A problem with South Africa's system, according to the 2017 report, is that **completely new items were used each year, meaning trends over time could not be established reliably**[15]. Remarkably, even Japan's system is said to not to produce reliable comparisons over time. In the case of Japan, the problem, like the problem in South Africa's ANA, is that secure anchor items repeated over time do not exist[16]. Both UNESCO reports are unfortunately thin on explaining what it is about well-designed systems that make them capable of producing sufficiently reliable results (though these details are readily available elsewhere). The point is made, however, that countries should learn from the international programmes such as TIMSS and PIRLS, which employ methods that to some degree need to be replicated in a national assessment.

The UN's **Sustainable Development Goals** (SDGs) are central in UNESCO's drive for more and better assessments across countries at the primary level. The indicators attached to these goals imply that countries need to measure the attainment of minimum proficiency levels in two learning areas – language and mathematics – at both the lower and upper primary levels[17].

4 The World Bank

To a large extent the **World Bank's position on standardised assessments mirrors that of UNESCO**.

Given the World Bank's economic focus, the emphasis on the following type of **cost effectiveness argument** should come as no surprise.

> Some people argue that assessments, particularly large-scale assessment exercises, are too expensive. In fact, the opposite tends to be true, with *testing shown to be among the least expensive innovations in education reform*, typically costing far less than increasing teachers' salaries or reducing class size. … even the most expensive state-level, test-based accountability programmes in the United States cost less than 0.25 percent of per-pupil spending. Similarly, in none of the Latin American countries reviewed … did testing involve more than 0.3 percent of the national education budget…[18]

A prominent contribution by the World Bank is its SABER[19] tools designed for evaluating a country's various education system component. One of these tools focusses on national

[14] http://uis.unesco.org/en/uis-learning-outcomes.
[15] UNESCO, 2017: 135.
[16] Kuramoto and Koizumi, 2016.
[17] United Nations, 2017; UNESCO: UIS, 2016.
[18] Clarke, 2012: 6.
[19] Systems Approach for Better Education Results.

assessments[20]. This tool, a 43-page document, provides a broad framework for understanding the purpose of **assessments and examinations, and their relationship with other factors in the schooling system**. This framework has been used to evaluate assessment systems in several developing countries. These evaluations are documented in short country reports available on the SABER web space[21]. What the framework does not do is to deal with hurdles that even high-level decision-makers need to understand relating to bringing about sufficient year-on-year comparability in assessments through anchor items and IRT scoring. The trade-off between sample-based and universal testing is moreover barely discussed. This is a pity.

SABER's concerns clearly span the primary level. Of the eight Sub-Saharan African countries covered in country-specific SABER reports dealing with assessments, all pay attention to what is happening at the primary level (but many also focus on the secondary level).

The SABER documentation refers to cases where standardised testing has had a net positive impact on educational quality, but **underlines that this impact is conditional on a number of factors**:

> …there is still much to learn about the optimal mix of incentives for test-based accountability models that will produce the best outcomes with the fewest negative side effects. To date, research models suggest that key determinants of whether the effects of test-based accountability exercises are more positive than negative include the technical quality of the tests themselves, the alignment between test design and the ways test results are used, and the extent to which supports are in place to help schools or teachers identified as underperforming…[22]

In short, the SABER and World Bank position is that assessments should be pursued, but that countries should exercise caution. This is also a message coming out of the World Bank's 2018 *World Development Report*, a report of special importance for the education sector insofar as its focus is on teachers. That report acknowledges that the levels of standardised testing imposed from above in schools in the United States has led to perversities in schools, but then goes on to argue that **in most of the world's countries there is too little, not too much testing**[23].

5 The OECD

The **Organization for Economic and Co-operation and Development (OECD)** is a policy research group serving and funded by, on the whole, rich countries which share democratic values. In identifying best practices, the OECD tends to draw from these countries. The OECD's best known education programme is PISA, or Programme for International Student Assessment, but this is applied at the lower secondary level, not the primary level.

An excellent source of the OECD's current positions on testing at the primary level (and testing in general) is the OECD's 2015 *Education Policy Outlook*[24]. This report pays considerable attention to testing. It indicates that **the use of standardised assessments has been on the increase in OECD countries**. This has been driven by 'greater decentralisation, more school autonomy, and rising expectations for better results and student outcomes'[25]. This would be in line with the notion that more autonomy for schools raises the need for the monitoring of quality from the centre. Australia and Spain are cited as countries which have

[20] Clarke, 2012.
[21] http://saber.worldbank.org/index.cfm?indx=6.
[22] Clarke, 2012: 9.
[23] World Bank, 2018: 17.
[24] OECD, 2015. A PDF version of this report can be obtained from the author of the current report.
[25] OECD, 2015: 94.

recently introduced universal standardised assessments in some primary grades[26]. Effective testing is one of several types of intervention promoted by the OECD report. Others include professional development for teachers and interventions aimed at improving the school environment.

Considerable attention goes to efforts to improve relations between the employer and teacher unions, clearly a critical matter given that unions are often not happy with new assessment programmes, in part due to workload concerns. The report recognises **the importance of a 'social dialogue' element in the employer-unions relationship**, and recommends countries pay more attention to deepening this aspect of the relationship. In other words, teachers should be recognised as important shapers of the country's political narrative, but they should also accept that how this is done must be part of a process of dialogue with stakeholders[27].

An earlier 2012 report published by the OECD provides further details on the extent of standardised testing found in OECD countries, and how test data are used at the primary level (in 2012). **Of 28 OECD countries for which data are provided, all but five do have 'standardised central assessments'**, and of the five only two – Portugal and the Slovak Republic – were not running sub-national standardised testing or planning to introduce a national system in the near future. Of the 23 countries *with* standardised assessments at the primary level in 2012, just four relied only on sample-based testing: Estonia, Finland, Netherlands and New Zealand[28].

6 World Economic Forum

The WEF is essentially an NGO which brings together leaders from the political, business and academic spheres each year in its famous Davos meetings. It was started in 1971 by Klaus Schwab, a German engineer and economist, who remains the WEF's Executive Chairman. Schwab is moreover largely responsible for popularising the term **'Fourth Industrial Revolution'** – in 2016 he published a book with this title. This concept helps to understand the likely impact of emerging technologies such as e-commerce and artificial intelligence on society, employment and the demand for skills.

The WEF has some research capacity (though not nearly of the level of UNESCO, the World Bank or OECD), and releases a number of standard reports, perhaps the best known being the ***Global Competitiveness Report***.

The **Davos meetings have been immensely influential** and at times dramatic. Mandela, De Klerk and Buthelezi held their first meeting outside South Africa in Davos in 1992, before a large audience. The WEF, to a much greater extent than, say, the World Bank or OECD, has been the target of anti-globalisation activists. In the education area, the *Global Competitiveness Report* has been criticised for ranking the educational quality of countries according to the opinions of business people within each country, without checks on the across-country comparability of the resultant indicator values, or checks against available test score averages[29].

The WEF has focussed very little on education policy, or testing, but a 2017 report released by the WEF (though not necessarily expressing the views of the WEF according to the report itself), attempts to outline best policy practices in education. Proposals are expressed largely in terms of the **responses required to address economic and societal changes associated**

[26] OECD, 2015: 103.
[27] OECD, 2015: 176.
[28] OECD, 2013: Annex 4.A2. A PDF version of this report can be obtained from the author of the current report.
[29] See http://resep.sun.ac.za/index.php/wef-rankings-on-education-unreliable for a South African critique, also Sabadie and Johansen (2010: 240) for a non-South African opinion.

with the Fourth Industrial Revolution. Much of the report deals with technical and vocational education at the secondary level, which is to be expected given the emphasis on skills for the labour market. However, the report does also focus on the primary level. The absolute importance of a solid educational foundation is emphasised, both in a focus on early childhood development and in the following:

> …there is a growing consensus that forward-looking curricula must focus on: the linguistic, mathematical and technological literacies all job roles will require in the future…[30]

In this sense, the WEF report is in line with the positions of UNESCO, the World Bank and the OECD.

However, the WEF's report differs when it comes to how to improve these foundational literacies. The **'Finland model' features prominently in the WEF report**. In this model, the emphasis falls strongly on teacher agency in determining how learning occurs, the assumption being that teachers are highly competent. Standardised testing is de-emphasised.

> The Finnish model is based on a number of core success factors … [a] light-touch approach to standardized curricula, based on systemic trust in high-quality teaching delivery … de-emphasizing the importance of testing and selection. … pilot-testing "self-assessments", whereby students are involved in determining their own progress, and "peer assessments"… abolish[ing] individual subjects, with students instead studying events and phenomena in an interdisciplinary format.[31]

One fact about Finland should be clarified. While it is true that there is no *universal* standardised testing at the primary level, there are **occasional, though not very frequent, sample-based standardised assessments** focussing on grades 3 and 5. It seems the most recent such testing occurred in 2012, with future testing planned for 2018 in Grade 6, 2019 in Grade 3 and 2022 in Grade 6[32].

In summarising its proposals, the report recommends, in relation to assessments, the following, without clarifying exactly what this means:

> Broaden assessment beyond traditional test-based approaches[33]

These are **proposals which differ substantially from those of the three other organisations discussed in sections 3 to 5**. Of course the World Bank and UNESCO are clearly directing their advice largely to developing countries, while one suspects that the WEF is focussing largely on what rich countries should do (though this is debatable, and the WEF report's aim in this regard is not made clear). However, even then the differences between the WEF and the OECD proposals are striking, given that the OECD is focussing on proposals for its members, meaning almost exclusively rich countries.

On a fundamental level, what appears to be happening is that the WEF report has fully appreciated ways in which the basic thinking around primary education has shifted in, say, UNESCO over the last twenty years. This shift is clearly reflected in the differences between the old Education for All (EFA) goals and indicators, and those of the new SDGs. While the former defined educational deprivation largely in terms of whether a child attends school, the latter defines it in terms of attendance *and* whether children are learning anything if they do attend school. **The WEF report, in stating what the basic educational problems are,**

[30] World Economic Forum, 2017: 8.
[31] World Economic Forum, 2017: 11.
[32] OECD (2013) and Finnish Education Evaluation Centre (2016).
[33] World Economic Forum, 2017: 13.

refers to the out-of-school challenge, but not the not-learning-in-school problem[34]. It is largely concerns about latter which underlies, say, UNESCO's emphasis on gathering data on learning outcomes.

There are **parts of the WEF report which one could argue are more line with the UNESCO approach**. It refers to the need for 'Competency-based credentialing and recognition systems'[35], which can be interpreted as reliable and standardised information on learner competencies.

To conclude, if one views this 38-page report released by the WEF in isolation from the broader body of knowledge and policy proposals, one might conclude that any country should aim to **emulate Finland at the primary level, and de-emphasise standardised testing**. While this may conceivably be a good *long-range* goal, one does clearly need to consider the various stages of development in a schooling system, each requiring its specific interventions. It is noteworthy that the OECD, of which Finland is of course a member, uses Finland to illustrate certain good practices, while it also draws from good practices in other developed countries. In the OECD's documents, there is no strong advocacy of the 'Finland model'.

7 Education International

Education International (EI) is the world's largest federation of unions. **Three South African teacher unions are members**: SADTU, NAPTOSA and SAOU. Unions from the great majority of countries are represented, though notable absences are China and Cuba, countries which do not permit teacher unions to operate freely in the typical sense.

Understanding EI's positions on education policy is critical for grasping the complex policy tensions in the sector. The resolutions for EI's last two world congresses, of 2011 and 2015, offer useful insights into the EI's positions. In a nutshell, the 2011 resolutions contained some very pointed opposition to the shift towards more standardised testing in schools across the world, and framed this opposition within a broader opposition to 'neoliberalism'. The 2015 resolutions retain the strong opposition to what are considered neoliberal education policies, while not expressing explicitly opposition to standardised testing. **Whether this represents a softening of EI's position on standardised testing is not clear.**

In places, the 2011 resolutions do seem to express **opposition to the very notion of standardised testing**:

> EI believes that a widespread abuse of the notion of quality to justify standardised forms of testing is harmful to the education system as a whole, as it attempts to reduce the teaching and learning process to quantifiable indicators. It is the standardization and one-dimensional approach to testing and evaluation of the teaching and learning processes to which EI objects strongly[36].

But other sections of the same set of resolutions clarify that the problem lies with **the inappropriate use of standardised testing**:

> When one form of evaluation designed for a particular purpose is used to serve a different purpose, the consequences can be unforeseen and damaging[37].

The **wider political concerns of EI** are captured in the following:

[34] Education International, 2011: 10.
[35] Education International, 2011: 9.
[36] Education International, 2011: 4.
[37] Education International, 2011: 1.

> The social values of education require public authorities to protect the education sector from the neo-liberal agenda of privatization and commercialisation. This negative agenda includes marketisation and trade in education and intellectual property, the casualisation of employment in the education sector, the application of private-sector management models on education institutions, the privatisation of provision, and the intrusion of for-profit motives or business interest in the governance of education institutions[38].

It appears as if for EI **standardised testing is not a central feature of neoliberal policies**, though the two are often closely associated with each other. As seen in the above extract, the trends EI is most concerned about are increasing marketisation, casualisation of labour and privatisation.

An ongoing and widely reported **dispute between EI and the company Pearson** is illustrative (SADTU has been actively involved in this dispute). EI's opposition to Pearson is largely due to Pearson's marketing of low-fee private schools in Africa and Asia, where costs are kept low through highly standardised curricula supported by software platforms[39]. But in addition, a long list[40] of mistakes and blunders in Pearson's testing systems used extensively by public schools in the United States has fuelled the notion that private involvement in testing is overly profit-driven, but also the notion that government's quality controls over service providers such as Pearson are lacking.

EI's 2015 resolutions make no mention of standardised testing, though the stand against, in particular, privatisation remains strong. However, it is significant that the **EI makes explicit its reservations about two organisations discussed above**:

> …serious concerns about the increasing role played since 2011 by some international institutions such as the OECD, the IMF and the World Bank which keep on influencing the policies implemented in the education sector by national governments[41].

The details behind these concerns are not too clear, though the concerns relating to the OECD probably have to do with PISA, whose role in education was criticised in the 2011 resolutions.

8 Schooling systems which are somehow exemplary

8.1 Brazil

Brazil is of great relevance to South Africa due to BRICS, and because in many respects Brazil shares South Africa's development challenges. Though Brazil's average scores in the PISA programme remain below those of many other developing countries, these average scores saw **larger increases than for any other country in the period 2000 to 2009** (with the 2009 level of performance roughly being sustained, but not improved upon, in the subsequent PISA runs of 2012 and 2015). In this sense, Brazil has been considered exemplary by several analysts[42].

Moreover, Brazil has **a testing system which has been considered exemplary**. For instance a World Bank report describes the system as 'superior to current practice in the United States

[38] Education International, 2011: 2.
[39] Article headed 'Teaching unions call for Pearson to sack John Fallon' at https://www.ft.com/content/59d35170-3025-11e7-9555-23ef563ecf9a; Nagarajan and Tabberer, 2016.
[40] Article headed 'Pearson's history of testing problems - a list' at https://www.washingtonpost.com/news/answer-sheet/wp/2016/04/21/pearsons-history-of-testing-problems-a-list/?utm_term=.5612f65a9f14.
[41] 'Resolution on trade union action to counteract neoliberal policies on education' at https://ei-ie.org/en/detail/14737/resolution-on-trade-union-action-to-counteract-neoliberal-policies-on-education.
[42] Carnoy *et al*, 2015.

and in many other OECD countries in the quantity, relevance, and quality of the student and school performance information it provides'[43]. Brazil's system has two 'legs'. The sample-based programme, known as SAEB, was introduced in 1990, while the universal leg of the system, known as Prova Brasil, was only introduced in 2005. This fifteen-year lag is interesting, and reflects the logistical, political and cost hurdles often associated with the introduction of universal testing (in South Africa too, the sample-based Systemic Evaluation was introduced in 2001, long before 2011 launch of ANA). SAEB has its roots in work by the World Bank in the 1980s on evaluating Brazilian schools. What is also noteworthy is that much of the testing occurring nationally today in Brazil is based on earlier innovations occurring in the state of Sao Paulo, Brazil's foremost state in economic terms.

Much of the documentation describing the Brazilian system is in Portuguese, though some is in English[44]. But even counting the documentation available in Portuguese, what is published on the **technical details of this important Brazilian system is less than what is necessary** to fully understand it.

Both SAEB and Prova Brasil cover Grade 5 at the primary level, are run every second year, use the same tests (Portuguese and mathematics), employ the same scoring system, and rely on **service providers external to the school to administer the tests**. The key difference between the two is that SAEB includes private schools within its sample, and thus produces representative statistics (down to the level of each of Brazil's 26 states), while Prova Brasil focusses on testing across all public schools (but not if there are fewer than ten Grade 5 learners)[45]. Reputable researchers who are familiar with the SAEB data consider the SAEB results comparable over time[46]. Similarly, official reports assume that Prova Brasil results are sufficiently comparable over time to draw conclusions on whether improvements occurred[47]. What is clear from the documentation is that since 1995 SAEB has used anchor items and statistical adjustments of scores based on item response theory (IRT), and that a matrix sampling approach has been followed, meaning different learners receive tests with different combinations of questions[48]. Since its inception, Prova Brasil has followed a similar approach[49].

While what is publicly available on the Brazilian system seems to be of a high technical standard, there are key questions that do not seem answered. Above all, one wonders how it is possible for a testing programme as large as Prova Brasil to implement **sufficiently tight security to ensure that tests do not become widely available to teachers and the public in general**. If this happens, the comparability of results over time would be compromised given that some items are repeated in different years. A further concern is that it seems missing schools in Prova Brasil varies somewhat from year to year. Clearly, this could also affect the comparability of, say, state-level averages.

How is Brazil's system used to improve performance? On the one hand, the authorities take advantage of the focus of the tests to disseminate items (questions) to schools and teachers which are similar to the ones used in the tests. Moreover, test scores are combined with learner retention figures to produce a composite performance score for every school, the Brazilian Education Quality Index (IDEB in Portuguese). This index has attracted much attention and been widely written about. It is used extensively in framing national debates on school improvement, and on setting school (and system) targets. It is also used to target

[43] Bruns, Evans and Luque, 2012: 7.
[44] See for instance Canen (2012).
[45] http://appprova.com.br/saeb.
[46] Carnoy *et al*, 2015.
[47] Brazil: Ministério de Educação, 2015a: 27.
[48] Brazil: Ministério de Educação, 2011: 9. De Andrade *et al*, 2000: 82-3; 137.
[49] Brazil: Ministério de Educação, 2015b: 30.

funding to the most needy schools[50]. It is possible to find the IDEB (and Prova Brasil) statistics, for several years, for each school, within a **school report card**, on the web[51]. One criticism that is made of the index is that it does not explicitly take into account the socio-economic disadvantage of each school's learners[52].

Brazil has **actively encouraged automatic promotion in the early grades** for over a decade, an intervention which has reduced grade repetition from levels above those seen in South Africa today (for instance 24% in Grade 1 in 2005 in Brazil) to levels well below the current South African ones, as manifested by average promotion rates at the lower primary level of 93% in 2015 for Brazil as a whole[53]. This would to a large degree have removed the dilemma of whether to use results from the externally administered tests to decide on promotion. The signal from the government has simply been to remove all grade repetition, regardless of academic performance, at least in the initial grades. It would in any case not be possible for schools to use Prova Brasil results for promotion purposes as schools are provided with school-level average results, through the school report cards, and not the results for individual learners.

Brazil's long-range education sector plan published in 2014 envisages enhancements to its system of standardised assessments. Specifically, it is envisaged that the upper secondary ENEM exit examination (the second-largest examination in the world in terms of participants) will be **redesigned along the lines of SAEB in order to improve the comparability of the examination results over time**[54].

The fee charged to private schools wishing to participate in Prova Brasil provides a sense of the minimum cost per school of implementing Prova Brasil. This fee comes to USD 1,200, or **14,000 Rand, per school for a school with 100 or more learners per grade**. This is roughly R140 per learner[55].

8.2 Chile

Chile is frequently put forward as an exemplary developing country. It has the **second-best 'control of corruption' value in Latin America** after Uruguay in the World Governance Indicators dataset (and exceeds the values of all Asian countries other than Japan and Singapore). Its Human Development Index is on a par with that of Poland and Portugal. Of nine countries in the Latin America and Caribbean region participating in PISA in 2015, Chile achieved the highest mathematics score. Yet Chile remains in many ways a developing country. Its educational performance in the international tests is below that of virtually all rich countries, and around 10% of the population is considered poor using World Bank measures.

The national testing system in Chile, known as SIMCE, was launched in 1996. It has been very ambitious in its coverage, but the realisation has grown that this could have been over-ambitious. As a result, the 2016 to 2020 **plans for SIMCE reflect some down-scaling**. Though annual universal testing in Grade 4 will continue, in Grade 2 what was universal testing has become sample-based testing, and in Grade 6 universal testing will be conducted every second year, and not every year. Unlike many other national testing system of this kind, SIMCE covers not just two or three subjects, but up to seven subjects. In Grade 4, for example, three subjects are covered: Spanish language, mathematics; and social sciences. The recent down-scaling efforts appear not to have affected the number of subjects covered.

[50] Cheng and Moses, 2016: 35.
[51] http://idebescola.inep.gov.br/ideb/consulta-publica.
[52] http://www.todospelaeducacao.org.br/educacao-na-midia/indice/30709/entidade-critica-ranking-de-escolas-feito-a-partir-de-avaliacoes-como-prova-brasil-e-pisa.
[53] Brazil: Ministério de Educação, 2015a: 10; Koppensteiner, 2013: 2-3.
[54] http://www.planalto.gov.br/CCIVIL_03/_Ato2011-2014/2014/Lei/L13005.htm.
[55] http://www.leiaja.com/carreiras/2017/05/25/governo-divulga-datas-da-prova-brasil-2017.

The quality of SIMCE's documentation is high. It is of a technically high standard, and problems which could affect comparability are made clear in reports on completed testing cycles. The emphasis on good documentation, but also other aspects of SIMCE, are explicitly driven by the ***Standards for Educational and Psychological Testing***, a guide (which one must buy) developed by three psychological and testing organisations in the United States, the most recent version of the guide being that of 2014. This guide is widely used by testing authorities in the United States.

All of the key **SIMCE tests are applied universally**, in other words not to a sample of schools or learners. Because the SIMCE tests are run in all public and private schools, Chile has no need for a sample to produce nationally representative statistics. The universal results produce these statistics. As in Prova Brasil, comparability across years is achieved through secure anchor items. As in Brazil, a matrix sampling approach is applied in the test design and scores are calculated using IRT[56].

One weakness of the Brazilian system is addressed in the Chilean system. SIMCE includes background questionnaires which collect data used to **determine the socio-economic (SES) level of the school community**. One element of this is questionnaires sent to parents through the learner, which come with an envelope that is sealed by the parent to protect privacy. National reports and school report cards make extensive use of SES categories. Moreover, schools are evaluated relative to other schools facing similar contexts.

Security controls in the test administration stage include placing test materials completely in the control of examiners working for a service provider at all points in time other than when learners are writing the test. Teachers from the school are not allowed entry into the test venue while the test is being administered. External examiners arrive at the school with the test scripts, and leave with test scripts plus learner responses, on the same day. There is **transparency around non-participation**, for instance in 2014 1.6% of schools did not participate in the Grade 2 test, and 16% of Grade 2 learners did not return the parent questionnaire[57] – clearly even in a well-functioning system reaching 100% is extremely difficult.

In SIMCE, 50 points is equal to one standard deviation[58]. Reports on long-range trends in Grade 4 reading reveal year-on-year improvements (and a few declines) which are within reasonable bounds. For instance, between 2005 and 2010 particularly large improvements were seen, but these come to around 0.06 of a standard deviation a year[59]. Truly rapid improvements hardly ever exceed 0.08 standard deviations a year, meaning **Chile's improvements are well within the historical 'speed limit' of countries**.

As in Brazil, schools receive only school averages from the administration, and not results for individual learners, meaning it would be **impossible to use SIMCE results to determine grade-on-grade promotion**.

It appears that **access to individual school report cards by the public, via the web, is no longer possible**, though some years ago it was. The reason for this is not provided, but one can speculate that the authorities have wanted to reduce the impact of poor interpretations of these report cards on parent decisions around where to send their children.

[56] Chile: Agencia de la Calidad de la Educación, 2014.
[57] Chile: Agencia de la Calidad de la Educación, 2014: 79, 88.
[58] Chile: Agencia de la Calidad de la Educación, 2014: 153.
[59] Chile: Agencia de la Calidad de la Educación, 2016.

In a separate programme known as SNED[60], Chile pays **financial incentives to whole schools displaying exceptional improvement** in their SIMCE results. This programme has been criticised by some analysts for not being particularly effective in the sense that it appears not to change the behaviour of schools. Thus while SNED acknowledges improvement, it does not appear to *cause* improvement[61].

8.3 Ghana

Ghana is exemplary insofar as it is among the top ten mainland Africa countries with respect to the World Governance Indicators 'control of corruption' variable. It probably also has the **technically most advanced report on its national assessments**, made public, of any African country.

Ghana's report, published in 2014, covers the 2013 wave of the National Education Assessment (NEA), which is run every second year in grades 3 and 6, with a focus on English and mathematics. The programme is sample-based and aimed at producing statistics at the national level and at the level of the ten regions in the country. The 2013 wave was the NEA's fifth wave (suggesting it began in 2007). USAID provides considerable funding and technical support for the programme. The **2013 results were the first results for which 'meaningful comparisons' could be made against the previous wave's results**[62]. This was possible due to the repeating of some items from the 2011 NEA tests. In addition, the 2013 tests include some TIMSS and PIRLS items.

Because the design of the NEA tests, and their administration, differed across 2011 and 2013, typical or ideal methods using IRT-based scores (as in Brazil and Chile) were not possible. Instead, four equating methods which all seem to make use of the common anchor items were attempted. One of these methods, the **'frequency estimation equipercentile method'** was considered best to make official comparisons across 2011 and 2013. While the work seems rigorous, there are gaps in the report which could raise questions around the credibility of the comparisons made. Above all, it is not completely clear whether 2011 raw mean scores were adjusted to a 2013 scale, or whether unadjusted 2011 statistics were simply considered sufficiently comparable to raw 2013 means to make adjustments unnecessary[63].

Furthermore, given the **importance of the test administration methods for providing confidence around the reliability of the results**, one might have expected more details in this regard. For instance, it is not clear if the external administrators (who were public officials) remained at the school for the full duration of the three-day testing period.

Confidence intervals are provided for national results but not regional results. In each region, 55 schools were sampled. Primary schools per region range from 2,900 to 500, but equal samples per regions would in fact be appropriate given the sampling issues discussed in section 2.2 above. However, it is very likely that **a sample below 100 per region would cause serious across-region comparability problems**, which is perhaps why confidence intervals for regions were not published. This is obviously concerning.

Ghana has **no examination at the primary level**. Learners take a national examination for the first time in Grade 9, at the 'junior secondary' level. The NEA thus represents the only gauge that the administration has at its disposal to monitor trends in the area of learning outcomes at the primary level.

[60] http://www.sned.mineduc.cl.
[61] Mizala and Urquiola, 2013.
[62] Ghana: Ministry of Education, 2014: ix.
[63] Moreover, it seems difficult to reconcile the statistics in Table 11 and Table C1.

Policies in Ghana on the determination of grade-on-grade promotion could not be found, but UIS statistics on grade repetition suggest that **Ghana is very close to practicing automatic promotion at the primary level**. The percentage of learners who are repeaters in the early grades is extremely low, at around 2%. This is much higher than in South Africa, where around 15% of Grade 1 learners are repeaters (though this drops to around 10% in subsequent grades)[64]. This does not mean that close to 100% of learners are truly ready for the next grade. As shown in the NEA report, 44% of Grade 3 and 35% of Grade 6 learners are considered 'non-readers'[65].

Ghana's report uses the NEA results to arrive at **policy conclusions in a range of areas**, including teacher development and the availability of materials in schools.

9 Southern African Development Community countries

The following table draws mostly from the **UIS database of assessments and examinations**[66]. Only SADC countries are considered. Clearly the database has not updated South Africa, as ANA is still said to be operating. The database is incomplete in the sense that there are developing countries with assessments and examinations not included in it. Of the 16 countries in the SADC, three were not included: Angola[67], Seychelles and Swaziland. Information for Swaziland was obtained separately, but no information was easily obtainable for the other two countries.

[64] Department of Basic Education, 2016: 14.
[65] Ghana: Ministry of Education, 2014: x.
[66] http://uis.unesco.org/en/uis-learning-outcomes.
[67] A 2009 SABER report by the World Bank on Angola indicates that at the time no national examination or assessment existed in the country, not even at the secondary level. Angola's extremely under-developed education system is very much a legacy of many years of civil war. See http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/CountryReports/SAS/SABER_SA_Angola_CR_Final_2009.pdf.

Primary-level assessments and examinations in SADC countries		
Country	*Name of assessment/examination (for assessments, whether it is sample-based or a census is indicated – no indication in brackets means an examination taken by all learners)*	*Grade*
Botswana	Standard Four Attainment Test (census)	4
	Primary School Leaving Examination	7
Comoros	Certificat d'études primaires élémentaires	6
Congo (DR)	Test national de fin d'études primaires	6
Lesotho	Lesotho National Assessment of Educational Progress (sample)	3, 6
	Primary School Leaving Examination	7
Madagascar	Évaluation des acquis scolaires des élèves de CM2 (sample)	5
	Certificat d'études primaires élémentaires	5
Malawi	Monitoring Learning Achievement (census)	2, 4, 7
	Early Grade Reading Assessment (sample)	2, 4
	Primary School Leaving Certificate Examination	8
Mauritius	Certificate of Primary Education	6
Mozambique	National Assessment (sample)	3
	National Examination Grade 5	5
	Primary Education Certificate	7
Namibia	Standardised Achievement Test (census)	5
South Africa	Annual National Assessment (census)	1, 2, 3, 4, 5, 6, 7
Swaziland[68]	Swaziland Primary Certificate	7
Tanzania	Early Grade Reading Assessment (sample)	2
	Standard Four National Assessment (census)	4
	Primary School Leaving Examination	7
Zambia	National Assessment Survey of Measuring Learning Achievement Levels (sample)	5
	Grade 7 Composite Examination	7
Zimbabwe	Zimbabwe Early Learning Assessment (sample)	3
	Grade Seven Examination	7

What are the noteworthy patterns one can draw from this table? Of the 14 countries covered, **all except for South Africa have some kind of national assessment or examination at the primary level** (this is of course assuming that other countries have not halted their systems, which seems unlikely). Of the 13 countries other than South Africa, all have either a universal assessment, or an examination, or both. In other words, all test all learners. None rely solely on a sample-based system. Moreover, of the 13 countries other than South Africa, just one does not have what UIS would classify as an 'examination' at the primary level. This country is Namibia, which is perhaps not surprising given the historical links between Namibia and South Africa. Of the twelve countries with an examination, eight have *both* an examination and, separately to this, what UIS would consider a 'national assessment'. Of these eight with a national assessment, six have only a sample-based assessment system (though they all also have a national examination).

[68] Umalusi, 2014.

References

Brazil: Ministério de Educação (2011). *PDE|SAEB: Plano do Desenvolvimento da Educação.* Brasilia. Available from: <http://portal.mec.gov.br/dmdocuments/saeb_matriz2.pdf> [Accessed April 2018].

Brazil: Ministério de Educação (2015a). *Resumo técnico: Resultados do Índice de Desinvolvimento da Educação Básica 2005-2015.* Brasilia. Available from: <http://download.inep.gov.br/educacao_basica/portal_ideb/planilhas_para_download/2015/resumo_tecnico_ideb_2005-2015.pdf> [Accessed April 2018].

Brazil: Ministério de Educação (2015b). *Avaliação Nacional da Alfabetização: Relatório 2013-2014.* Brasilia. Available from: <http://portal.inep.gov.br/documents/186968/484421/Relat%C3%B3rio+ANA+2013-2014+-+Da+concep%C3%A7%C3%A3o+%C3%A0+realiza%C3%A7%C3%A3o/8570af6a-c76e-432a-846f-e69bbb79e4b2?version=1.3> [Accessed April 2018].

Bruns, B., Evans, D. & Luque, J. (2012). *Achieving world class education in Brazil: The next agenda*. Washington: World Bank. Available from: <https://openknowledge.worldbank.org/bitstream/handle/10986/2383/656590REPLACEM0hieving0World0Class0.pdf?sequence=1> [Accessed February 2016].

Canen, A. (2012). *Assessment of schools in Brazil: Some reflections.* Rio de Janeiro: Universidade Federal do Rio de Janeiro. Available from: <http://www.nwu.ac.za/sites/www.nwu.ac.za/files/files/p-saeduc/3_Assessment%20of%20schools%20in%20Brazil_some%20reflections.pdf> [Accessed April 2018].

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

Cheng, X.J. & Moses, K. (2016). *Promoting transparency through information: A global view of school report cards.* Paris: IIEP. Available from: <http://unesdoc.unesco.org/images/0024/002463/246358e.pdf> [Accessed April 2018].

Chile: Agencia de la Calidad de la Educación (2014). *Informe técnico: SIMCE 2014.* Santiago. Available from: <http://archivos.agenciaeducacion.cl/InformeTecnicoSimce_2014.pdf> [Accessed April 2018].

Chile: Agencia de la Calidad de la Educación (2016). *Resultados educativos: 2016.* Santiago. Available from: <http://archivos.agenciaeducacion.cl/ResultadosNacionales2016_.pdf> [Accessed April 2018].

Clarke, M. (2012). *What matters most for student assessment systems: A framework paper*. Washington: World Bank. Available from: <https://openknowledge.worldbank.org/bitstream/handle/10986/17471/682350WP00PUBL0WP10READ0web04019012.pdf?sequence=1> [Accessed October 2015].

De Andrade, D.F., Tavares, H.R. & Valle, R. (2000). *Teoria de resposta ao item: Conceitos e aplicações*. Fortaleza: Universidade Federal do Ceará. Available from: <http://homes.ufam.edu.br/jcardoso/LivroTRI.pdf> [Accessed April 2018].

Department of Basic Education (2016). *Report on progress in the schooling sector against key learner performance and attainment indicators*. Pretoria. Available from: <http://www.education.gov.za/Portals/0/Documents/Reports/Education%20Sector%20review%202015%20-%202016.pdf> [Accessed September 2016].

Education International (2011). *Policy paper on education: Building the future through quality education*. Brussels. Available from: <http://www.ei-ie.org> [Accessed October 2011].

Finnish Education Evaluation Centre (2016). *National plan for education evaluation 2016-2019.* Helsinki. Available from: <https://karvi.fi/app/uploads/2016/06/National-Plan-for-Education-Evaluations-2016-2019.pdf> [Accessed April 2018].

Ghana: Ministry of Education (2014). *Ghana 2014 national education assessment: Technical report*. Accra. Available from: <https://globalreadingnetwork.net/sites/default/files/eddata/2013_NEA_Technical_Report_15May2014_wRecs.pdf> [Accessed April 2018].

Koppensteiner, M.F. (2013). *Automatic grade promotion and student performance: Evidence from Brazil.* Leicester: University of Leicester. Available from: <https://www.le.ac.uk/economics/research/RePEc/lec/leecon/dp11-52a.pdf> [Accessed April 2018].

Kuramoto, N. & Koizumi, R. (2016). Current issues in large-scale educational assessment in Japan: Focus on national assessment of academic ability and university entrance examinations. *Assessment in Education: Principles, Policy & Practice.*

Lam, D., Ardington, C. & Leibbrandt, M. (2008). *Schooling as a lottery: Racial differences in school advancement in urban South Africa*. Cape Town: SALDRU. Available from: <http://www.saldru.uct.ac.za/papers/wpapers/2008_18.pdf> [Accessed November 2008].

Mizala, A. & Urquiola, M. (2013). School markets: The impact of information approximating schools' effectiveness. *Journal of Development Economics*, 103: 313-335.

Nagarajan, A. & Tabberer, R. (2016). *The inner workings of affordable private schools: A guide for entrepreneurs and investors.* London: Pearson. Available from: <http://palf.io/annual-letter/the-inner-workings-of-affordable-private-schools.pdf> [Accessed April 2018].

OECD (2013). *Synergies for better learning: An international perspective on evaluation and assessment*. Paris.

OECD (2015). *Education policy outlook 2015: Making reforms happen*. Paris.

Sabadie, J.A. & Johansen, J. (2010). How do national economic competitiveness indices view human capital? *European Journal of Education*, 45(2): 236-258.

Umalusi (2014). *A comparative report on the education landscape of the countries in the Southern Africa Association for Educational Assessment*. Pretoria. Available from: <http://www.umalusi.org.za/docs/research/2015/harmonization.pdf> [Accessed April 2018].

UNESCO (2014). *Education for All global monitoring report 2013/4: Teaching and learning: Achieving quality education for all*. Paris. Available from: <http://www.unesco.org> [Accessed May 2015].

UNESCO (2017). *Global Education Monitoring Report 2017/18: Accountability in education: Meeting our commitments*. Paris. Available from: <http://www.unesco.org> [Accessed April 2018].

UNESCO: UIS (2016). *Laying the foundation to measure Sustainable Development Goal 4*. Montreal. Available from: <http://uis.unesco.org/sites/default/files/documents/laying-the-foundation-to-measure-sdg4-sustainable-development-data-digest-2016-en.pdf> [Accessed April 2018].

United Nations (2017). *Revised list of global Sustainable Development Goal indicators*. New York. Available from: <https://unstats.un.org/sdgs/indicators/Official%20Revised%20List%20of%20global%20SDG%20indicators.pdf> [Accessed October 2017].

Van der Berg, S., Van Wyk, C., Van Broekhuizen, H., Von Fintel, M., Gondwe, A. & De Villliers, P. (2015). *Learner flow-through patterns and performance in Western Cape schools and universities based on longitudinal cohort analysis.* Stellenbosch: Research on Socio-economic Policy. [Unpublished report dated 30 November 2015].

Van der Berg, S. & Shepherd, D. (2008). *Signalling performance: An analysis of continuous assessment and matriculation examination marks in South African schools*. Pretoria: Umalusi. Available from: <http://www.umalusi.org.za/ur/publications/UmalusiCASSReport2009%20LoRes.pdf> [Accessed August 2009].

World Bank (2018). *World Development Report 2018: Learning to realize education's promise*. Washington. Available from: <http://www.worldbank.org> [Accessed December 2017].

World Economic Forum (2017). *Realising human potential in the Fourth Industrial Revolution: An agenda for leaders to shape the future of education, gender and work*. Geneva. Available from: < http://www3.weforum.org/docs/WEF_EGW_Whitepaper.pdf > [Accessed April 2018].