04

The role of assessment in Foundation Phase improvement: The Annual National Assessments and beyond

CAROL NUGA DELIWE & SERVAAS VAN DER BERG

Abstract

Although it is a necessary part of delivering quality education at the classroom, learner, school, and system level, the assessment of learning outcomes at the individual level is a contested terrain in South Africa. To optimise resources targeted at improving learning outcomes, assessments must be conducted at the individual learner, classroom, and school levels. In addition, some assessments must provide information on performance at the national (or system) level, while other (more universal) assessments are more important for improvement at the learner, classroom and school levels.

The Grade 12 examination – a universal, summative learning assessment – has helped to galvanise resources and effort to improve instruction and learning in the higher grades of school. However, universal assessments in lower grades have been fraught with political and administrative difficulties, despite the need to measure foundational skills. The sudden demise of the Annual National Assessments (ANAs) in 2015, due to what unions regarded as their punitive use, is a clear demonstration of this.

In this chapter, we examine the international assessments that South Africa participates in, and national assessments such as the previous and planned Systemic Evaluations (SEs), the ANAs, and the school-based assessments (SBAs), as sources of information and

KEYWORDS
assessment, accountability, Annual National Assessments (ANAs), learning losses

pressure for accountability and improvement. In this chapter, we argue that increasing the use of *common SBAs* offers an opportunity to develop a comprehensive assessment system that includes examinations, SBAs, SEs, and international assessments.

With some adjustments and external moderation, common curriculum-aligned SBAs can be used for effective feedback and improvement at the classroom and learner levels. However, to prevent common SBAs from being used as a tool to punish schools, the mistakes made with the design and administration of the ANAs should be borne in mind.

1 Introduction

Measuring learning outcomes (what learners know and can do) has been a contested terrain for many education systems and researchers, despite universal acknowledgement that assessment plays an important role in curriculum implementation (UNESCO 2013; Darling-Hammond & Wentworth 2010; Department of Education [DoE] 1995). Venkat and Sapire (this volume) refer to the 'essential circuits' of education and the link between the curriculum, teaching practice, and assessment. Our focus is strictly on the Foundation Phase (FP), and where we refer to a specific subject, mathematics is our first concern. This chapter, therefore, only makes passing reference to the major external assessment, the National Senior Certificate (NSC) or matric examination.

We pay particular attention to the dual role of assessments as tools of accountability and important sources of information for many actors in the school system. There is a tension between these two roles, relating to how the information provided is used at different levels of the system, for different purposes. In the South African education system, school principals are required to both monitor and improve school performance, typically through learning assessments.

The objectives of assessment are to provide information to parents, learners, and teachers on the performance of individual learners, in order to generate pressure to be accountable for improving learners' progress and outcomes, and to report to relevant authorities on the school's performance as a place of learning and teaching (DBE 2010; 2015a; 2020). Chetty (2016) notes this dual role of assessments by contrasting the use of classroom- and school-based assessments in mathematics (to gather information that can be used to improve outcomes in maths) with that of systemic tests (including international assessments), which are used for accountability purposes at the national level. The texture, granularity, and form of classroom-level information that teachers might use to determine how to align their day-to-day teaching practice with the curriculum is different from the information acquired from sample-based[1] assessments that are intended to provide snapshots of the state of learning outcomes in the entire schooling system, for a broad audience.

Sample-based assessments take different forms. International assessments such as the Trends in International Mathematics and Science Study (TIMSS), Progress in

1. Typically, a sample of schools is drawn. Within these schools, learners from entire classrooms in a particular grade are assessed for their skills and knowledge levels.

International Reading Literacy Study (PIRLS), and the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), in which South Africa participates, provide well-known measures of learning at a system level. Systemic Evaluations (SEs), which are being reintroduced ostensibly to replace the Annual National Assessments (ANAs), are another form of sample-based assessment (Nuga Deliwe 2017).

This chapter starts with a bird's-eye view of the literature on assessment, particularly as it relates to South Africa. We then examine the different forms that assessment has taken in the country, mainly since the turn of the century.

We pay considerable attention to the ANA initiative as a policy intervention, highlighting some of its potential strengths and failings. To do this, we draw from earlier contributions (Hoadley & Muller 2014; Spaull 2015; Van der Berg 2015) relating to the so-called universal ANAs and the sample-based verification ANAs, among others. In line with the theme of this book, our analysis of performance patterns across grades, quintiles, and provinces focuses on Grade 3 mathematics. We try to draw some conclusions from anecdotal evidence on whether the ANAs may have improved assessment practices and perhaps also macro-pacing (the pace at which the curriculum is covered) in certain Foundation Phase classrooms (Goldsmith 2009). Thereafter, we look at international assessments such as TIMSS to evaluate their usefulness for improving policy and education outcomes before we briefly evaluate another universal testing system, the Western Cape Systemic Tests.

Next, we assess the new sample-based SEs introduced to replace the ANAs. We evaluate whether these planned tests could achieve some of the many objectives associated with the ANA initiative, namely to measure learning outcomes, to provide information to learners, parents and teachers at both a learner and class level, to provide information to schools and education authorities for reporting purposes, and to create pressure for accountability to improve learning outcomes. Finally, we argue that current forms of classroom-based assessment (usually called school-based assessment or SBA) have potential as an assessment tool that is both a source of information and a stimulus for accountability. In support of this argument, we look at a survey in which district officials were asked about the frequency of common assessments in schools within a province or district. This leads us to the conclusion that common SBAs, when used carefully in combination with SEs and international assessments, could provide a framework for improving the quality of information about learning, and raise accountability pressure to improve learning in schools.

Finally, we reflect on how SBAs and SEs can complement one another in fostering improved numeracy skills and knowledge in Foundation Phase classrooms, with some recommendations for strengthening the SBAs. The need to strengthen mathematics skills in the earliest grades in school is highlighted by Spaull et al. (this volume), who point out emerging evidence of weaknesses in outcomes well before the end of Grade 1.

This chapter, therefore, addresses two major research questions. Firstly, how have assessment practices evolved, and what is the current state of the Foundation Phase assessment system in South African schools? Secondly, can the dual roles of assessment as a source of information and for accountability purposes, be improved by making better use of common SBAs?

2 Literature on the role of assessments

2.1 Assessments and their roles

Alternative narratives on politics and ideology relating to assessment have always agreed on two things: firstly, the usefulness of assessment information to account for system performance and to assess learning at the level of the individual learner; secondly, teaching practice that uses feedback from assessment has a substantial effect on learning in classrooms and schools, when used to diagnose and remediate weaknesses in instruction and curriculum implementation (Lockheed & Verspoor 1991; Black & Wiliam 1998; Barber & Mourshed 2007; Ferrer 2006; Darling-Hammond & Wentworth 2010; Clarke 2012b; Darling-Hammond et al. 2014).

The main contention in the literature is about the extent to which assessment is used *outside* of classrooms to publicly account for and judge the quality of teaching in individual schools and classrooms, and by individual teachers (Phelps 2012; Conley 2015).

While the call for accountability creates pressure to use the information on performance to improve learning, information on its own is also useful for management and administrative purposes within schools. At the system level, accountability pressure can bring about genuine improvements in learning quality. However, pressure to account for one's performance can also lead to perverse responses focused on improving measured, visible outcomes without substantive improvements in underlying quality. These warped incentives are most likely to occur where there are high stakes attached to visible performance rather than to real improvement (Carnoy & Loeb 2002). For instance, in South Africa, high rates of learner retention at schools in the three grades before Grade 12 signal the strong pressure to be accountable for a school's good performance in the NSC examination (Van der Berg et al. 2021). Schools tend to hold back weaker learners so that they do not progress to the next grade, resulting in spikes in enrolment in Grades 9–11 as learners repeat a year and are discouraged from proceeding to Grade 12. The same pressure to be accountable for improvement does not exist in South African primary schools. However, many countries in southern and eastern Africa still conduct primary school exit examinations that may induce pressure, though the previously high stakes of these exams have largely reduced.

2.2 Measures of learning: sample-based assessments provide information to increase accountability pressure at the system level

Sample-based assessments collect information from selected schools, which makes it possible to analyse system performance. But since they do not provide information on all schools, learners, and classrooms, they cannot credibly be used as accountability

pressure tools to improve specific learning institutions.[2] Dixit (2012) notes that incentives to change weaken if accountability pressure is low, dispersed in a system, and is not attributable to a specific unit or organisation within a system.

All education systems have some form of SBA (Rosenkvist 2010; Black & Wiliam 1998). In a presentation to an Umalusi Colloquium in 2019, the Department of Basic Education (DBE) stated that SBA is the process of gathering valid and reliable information from the teacher about the ongoing performance of the learner against clearly defined criteria, using a variety of methods and tools (Umalusi 2019). Applying accountability pressure, using information from assessment, may induce individual schools to make improvements tailored to their needs, provided there is support and capacity for these improvements.

SBAs, including practical assessment tasks, and the Grade 12 NSC are the best-known forms of assessment in South Africa. The NSC is an excellent information tool, as it is both universal and aligned to the Grade 12 curriculum. Results from the NSC examination are used as an indication of education quality and therefore serve as a source of accountability pressure to promote improvement in higher school grades through what Braun and Kanjee (2006) refer to as the "backwash effects" of assessment results. The NSC examination does indeed influence curriculum implementation and teaching in earlier grades. However, for individual learners, the information comes too late, i.e. at the end of their school careers, long after they have written their exams.

There is no equivalent source of assessment data for primary schools that could induce a similar improvement in the system. To monitor numeracy skills and knowledge and to diagnose weaknesses and remedial actions at a learner, subject and classroom level, better measures of numeracy skills are required in the early grades. However, common assessments (SBAs that use the same test for a group of schools) are conducted, especially in Grades 3, 6, and 9, and they can provide similar information, provided that marking is sufficiently standardised through external moderation.

2.4 Measures of learning: universal assessments provide information and accountability pressure for improvement at school level

Information from universal assessments can be used to create accountability pressure to generate effort and resources for improvement at the individual learner, classroom and school level. The data from universal assessments in South Africa are typically aggregated into school-level reports, but are rarely used to communicate local-level information to parents. For example, the Western Cape Systemic Tests are universal, but the test reports only provide aggregated data at the *school* level and not routinely at the learner or classroom level. Reporting to parents, learners, and teachers on these tests is not common or standardised, though this kind of reporting can be done in

2. Because the identities of the participating schools and learners are usually not revealed, even the schools that participate in these surveys do not receive feedback.

the different curriculum domains, even if the test questions are kept confidential for future use.

The universal ANAs were useful in this respect. Reporting methods for local accountability included guidelines for interpreting and using ANA results at the school and classroom levels, based on individual learner assessments in numeracy and language skills; these were for parents and school governing bodies (DBE 2011b). Such report cards can be useful tools for learner- or classroom-level improvements (to adjust teaching and ultimately improve learning) and to provide information on performance at a local level to parents and learners. There is limited evidence of the systematic use of school reports on learning at the primary level using the ANAs, although the reporting of routine CAPS-aligned school-based assessments happens more often: every quarter.

In 2012, the DBE published guidelines for interpreting and using the ANA results to improve teaching and learning through feedback loops from the national assessment. The guidelines provide information on how teachers, principals, and district officials may practically use assessment data from the ANAs to develop strategies for improvement in classrooms, schools, and school communities, and in district support, monitoring, and oversight of teaching and learning. According to the guidelines, schools were expected to use learner-specific assessment information as the basis for developing plans, programmes, targets, and interventions to improve learning outcomes within classrooms and schools (DBE 2011a; DBE 2011b, 3, 10). Issued just after the first full implementation of the universal ANA, almost a third of the 18-page guideline on using and interpreting the ANA contains detailed instructions for teachers and officials on how to compute, analyse, and synthesise information on patterns, levels and distribution of performance at the learner, subject and grade level. The guidelines also provide practical guidance and examples of how to compute and compare the distribution and aggregate performance of learners in the grade under assessment review.

Universal assessments sometimes have high stakes at the learner, classroom, and school level (teachers and personnel) for all schools. Where the assessment system is not well developed and comprehensive enough to enable school-level improvement and system-level reporting, it becomes hard to maintain a balance between providing information for reporting, and providing information to secure accountability pressure for improvement at the learner level, respectively.

Teachers must be able to use assessment information to inform future improvements in teaching and learning (through effective feedback linked to the curriculum) and to guide their professional development. The lack of capacity to do assessments and to give feedback on them at schools, observed by policy-makers and researchers alike, may explain the findings that teachers often rely on summative rather than formative assessments, as these are more convenient to replicate and administer (Umalusi 2019). Inevitably, policy-makers focus on reporting and accountability at the school level, while teachers are more interested in learning improvements in their classrooms at the individual learner level (Best et al. 2013).

In the absence of a comprehensive assessment system for credibly monitoring school performance and providing information for accountability, performance in

universal ANAs in different years was used inappropriately to sanction or reward schools, even though the test results were not technically valid for such comparisons (SADTU 2014).[3] The ANAs triggered opposition by unions to this method of learning assessment (SADTU 2011). Unions were concerned about the administrative burden, the inability to use assessment information for improvement, and an unhelpful focus on naming and shaming individual teachers and schools due to the ANA results being reported in the public domain (SADTU 2014).

Despite these shortcomings and the fact that the ANAs were only in place for four years, these assessments created the space for discussions about accountability in schools (Taylor 2015).

2.5 Towards an assessment system: the assessment diversity

Clarke (2012a, 2012b) defines an education assessment system as a collection of policies, structures, practices, and tools for generating and using information on students' learning for decision-making and policy-support. Most countries with mature education systems administer universal assessments at the school level for formative feedback and summative assessments for decisions on grade promotion. These assessments are supplemented by sample-based systemic assessments, international assessments for benchmarking, and examinations for certification and selection (Clarke 2012a, 2012b; Rosenkvist 2010). In any such system, measuring learning outcomes is central (Nuga Deliwe 2017).

The ideal national assessment system is diverse, comprising a combination of sample-based assessments (useful for system-level reporting and accountability pressure, with high stakes for policy-makers and low stakes for individual schools) and universal assessments (useful for school-level accountability pressure and communication and for mobilising improvement, with relatively high stakes for learners, school personnel, and parents). Assessment test results can be used formatively to guide instruction within individual classes, with limited consequences for learners, teachers, and the school. On the other hand, summative assessment at the end of a grade or cycle can guide learners' decisions about subject choice, possible post-school career choices, and further educational opportunities. Summative assessments (for example, examinations at the end of the grade or common assessments) can influence the behaviour and performance of learners, teachers, and schools as these are associated with important educational decisions.

3 The SADTU National Congress held on 5 October 2014 re-affirmed the 2013 National General Council resolution: "… That ANA should remain a systemic evaluation with clear time frames that would allow for prompt feedback to be given to schools before the results are publicized followed by meaningful intervention programmes; that ANA should not be abused to label teachers and schools, thereby demoralising and de-professionalising them; and, that ANA should be reviewed as an annual assessment as of 2015, and be substituted by a [three]-year cycle of assessment."

2.6 'Learning poverty' and numeracy in the Foundation Phase

Learning poverty is defined by the World Bank as a state of being unable to read and understand a short, age-appropriate text by the age of ten. It includes learners who have not achieved minimum reading proficiency and children who are not in school. While the World Bank acknowledges that all foundational skills are important, it motivates that reading is an appropriate proxy for foundational learning, which is intuitively understood by the public and media, and a useful proxy for quality of learning, just as physical stunting is widely recognised as one of the indicators of early childhood development (World Bank 2018).

Implementing this concept of learning poverty (that singles out reading) may unintentionally take focus away from the importance of developing numeracy skills in the early grades, and negates the vital role of mathematics in sustainable development. Furthermore, the United Nations (UN) Sustainable Development Goal indicator 4.1.1 specifically refers to young people reaching a certain minimum proficiency in reading and mathematics by Grades 2 and 3 (Azevedo & Montoya 2021; ACER 2019; UNESCO 2015).

The UN's minimum benchmarks for proficiency at the lower-primary school level for reading and numeracy are as follows:

> *Nutshell statement for reading learning area: Students read aloud and comprehend many single written words, particularly familiar ones, and extract explicit information from sentences. They make simple inferences when longer texts are read aloud to them (Azevedo & Montoya 2021, 12).*
>
> *Nutshell statement for mathematics: Students demonstrate skills in number sense and computation, reading simple data displays, shape recognition and spatial orientation (Azevedo & Montoya 2021, 25).*

South Africa will not come close to meeting its international commitments to sustainable development if attention to reading skills is favoured to the detriment of numeracy. Although President Ramaphosa's State of the Nation Address in 2020 stated that every child in the country aged ten should be able to *read* for meaning, the development of South African children's numeracy as well as mathematics skills should receive the same urgent attention.

3 Analysis and findings

3.1 Assessments and education policy after the political transition

Education policy after 1994 focused on curriculum reform, while assessment reform came much later in 1998. Participation in national and international assessments

has been vigorous in the country since the early 1990s (DoE 2003a, 2003b, 2005, 2008). Founding education legislation set down system-level norms, standards, and monitoring expected of the Minister of Education and the DBE (Republic of South Africa National Education Policy Act 1996).

The phased implementation of the National Curriculum and Assessment Policy Statement (CAPS) between 2010 and 2014, following three cycles of curriculum review, was followed by the review of Schedule 4 of CAPS. The curriculum reviewers were concerned about the weak specifications of the curriculum and ineffective assessment practice and use in schools, particularly in those serving learners from poor households.[4] Schedule 4 specified assessment tasks more clearly, describing SBAs as formal and informal assessment tasks responding to curriculum needs (DBE 2010, 2020). The ANAs were launched with the CAPS and combined two versions of assessment in a new education reform package: a universal ANA for instructional improvement and a sample-based ANA for monitoring learning outcomes at the system level. The sample-based ANA was a sub-sample of universal ANA assessment test responses, subjected to stricter external moderation to assure the quality of the results emerging from the universal ANA.

3.2 The ANA experiment: aligned with the curriculum but not developed into a policy

The ANA was introduced in 2009, following some testing and the launch of the Foundations for Learning Campaign in 2008, but it only became fully operational in 2011 (Chetty, 2016). The universal assessment (separate from the sample-based or systemic version) entailed assessing both numeracy/mathematics and literacy/language skills in Grades 1 to 6 and in Grade 9. It was a highly ambitious logistical undertaking to test seven million learners in two subjects each and then analyse and capture the test results. The multiple objectives of the ANAs perhaps contributed to the failure of the initiative (Van der Berg et al. 2020, v).

One perspective was that ANA was introduced to improve accountability throughout the system (Taylor et al. 2013, 264–265; National Planning Commission 2012). However, the foreword by Minister Motshekga to the 2014 ANA report (DBE 2015b) points out that the tests were intended as a formative assessment tool so that gaps in teaching and learning could be identified and classroom practices adjusted accordingly. According to the DBE (2011b), ANAs were expected to improve learning by exposing teachers to best assessment practices, making it possible to target interventions at schools that needed it most, by allowing schools to measure their improvement, and by giving parents better information on their children's performance.

Gustafsson (2015) concluded from anecdotal reports that the universal ANAs did indeed increase teachers' exposure to and capacity for assessment. Similarly, Nuga Deliwe (2017, 138) reflects that teachers' involvement in the administration and setting

4. Interview with Dr Rufus Poliah, Chief Director: National Examinations and Assessment, DBE, September 2014.

of questions in the curriculum-aligned NSC and the ANAs made a positive contribution to "assessment capital" in schools.

However, in some cases, universal ANAs were used to sanction and penalise schools for poor performance: this was based on comparing schools within the same year of assessment, despite the ANAs not being designed for this purpose – a point repeatedly raised by teachers' unions. Eventually, the fact that neither version of the ANA programme was ever formally developed into educational policy, combined with the technical and social validation deficiencies of the programme, led to its demise.

Table 1 demonstrates some of the deficiencies of the ANAs, both as a tool for accountability and as a source of information. It shows the average scores for all grades tested in mathematics for three of the years that the universal ANA was fully functioning. Increases in average mathematics marks from 41% to 56% (Grade 3) or from 27% to 43% (Grade 6) in two years are clearly not credible. But the performance across different grades varied greatly, with Grade 4 results remaining unchanged across the three years. The Grade 9 results were particularly far out of line, creating the incorrect impression that South African mathematics performance was adequate in the lower grades but far from acceptable in Grade 9. This event gave rise to calls for a strong focus on Senior Phase mathematics. Yet Gustafsson (2015) pointed out that the ANA results for Grade 9 showed little correspondence with the Grade 12 matric results. In 2013, of the 43 public schools in which 80% or more of all Grade 12 enrolment passed mathematics, 18 had not registered any passes in ANA mathematics in Grade 9, and the other 25 had an average of only 38% passing ANA Grade 9 mathematics. Another indication that the ANA Grade 9 mathematics results (shown in Table 1) exaggerated the differences in learners' performance across grades, is that 36% of Grade 5 learners and 47% of Grade 9 learners achieved the Low International Benchmark of 400 in TIMSS. The performance of learners in the rigorously standardised international assessment, TIMSS, was higher than the performance indicated in the ANAs.[5]

Table 1: Average percentage scores in ANA mathematics tests by grade (2012–2014)

	2012	2013	2014
Grade 1	68	60	68
Grade 2	57	59	62
Grade 3	41	53	56
Grade 4	37	37	37
Grade 5	30	33	37
Grade 6	27	39	43
Grade 9	13	14	11

Source: Van der Berg 2015, 3.

5. Authors' own calculations. Note that these benchmarks are not necessarily directly comparable. Also, note that South African learners were tested in Grades 5 and 9, whereas other countries tested in Grades 4 and 8.

At the end of 2015, the ANAs were discontinued due to strong opposition by unions; many members appeared to be threatened by what they saw as a new form of accountability-policing. Several educationists also objected to the ANAs and argued that they could lead to an undesirable approach to teaching, such as "teaching to the test" (Van der Berg & Hofmeyr 2018, 16). Some researchers had objections to the content of tests, and the inconsistency of the results over time made them unreliable as a measure of school performance. There was no strong coalition supporting the idea behind the ANAs, and parents and some education officials were largely in the dark about their role (Cartwright 2013).

3.3 International assessments

South Africa has participated in several sample-based international assessments of learning since the political transition of 1994. While these assessments provide a credible snapshot of system performance and learning, they are of limited value in securing direct and specific learner-level and classroom-level improvements. At the primary school level, the country only participated in the 2015 and 2019 TIMSS assessments designed for Grade 4 (but in South Africa, the tests were administered to Grade 5 learners). Mathematics performance did not change much between 2015 and 2019. Scores declined very slightly from 376 to 374, a statistically insignificant decrease, while the percentage of learners who achieved the Low International Benchmark score of 400 also declined slightly from 39% to 37%. In contrast, the Grade 9 TIMSS results increased substantially over the same period, from 372 to 389, after strong gains from 289 in 2003 and 352 in 2011 (Reddy et al. 2022).

SACMEQ is the only other international mathematics assessment undertaken in primary schools. Although this test is only conducted in Grade 6, the results also indicate the quality of earlier learning, and the performance of South African learners in this test is poor. Available results indicate that South Africa's performance is not much above the SACMEQ average. Five of the 14 countries participating (Botswana, Kenya, Seychelles, Eswatini, and Uganda) outperformed South Africa in mathematics. Furthermore, mathematics teachers in the same five countries and Zimbabwe performed better than South African teachers on a very similar test to the ones used for testing learners (SACMEQ 2021, 87).

3.4 The Western Cape Systemic Tests

The Western Cape Education Department has maintained a system of Systemic Tests (sometimes called Diagnostic Tests) in Grades 3, 6, and 9 for almost two decades. These tests are universal, and the test system has provided useful information at a provincial level, especially since 2011 when the Centre for Evaluation and Assessment at the University of Pretoria was brought in to evaluate and update the testing instruments to maintain standards over time. Yet these tests are not exploited optimally at the learner

or classroom level. While schools do receive feedback, it does not appear to be sufficient to inform changes in teaching. Test pass rates are set at 50%, and average test results rose from 47.2% in 2011 to 58.1% in 2019. In her statement on progress made from 2011 to 2019, the Western Cape Member of the Executive Council (MEC) responsible for Education, Debbie Schäfer, noted that Grade 3 learners showed improvement regarding patterns, algebra, and functions but that they were still struggling with measurement (Schäfer 2020).

The value of the Western Cape Systemic Tests was further highlighted in a recent assessment of learning losses and learner dynamics in the Western Cape during the Covid-19 pandemic. It was found that Grade 3 mathematics scores for all the questions that appeared in the 2019 and 2021 papers dropped from 59.5% to 50.7%, a decline of 36% of a standard deviation. Furthermore, the percentage of learners failing to achieve 50% for questions that were common to both the 2019 and 2021 tests increased from 32% to 47% of all learners (Van der Berg et al. 2022). Lastly, the authors found clear declines in skills relating to number operations and relationships, and a further drop in performance in the measurement domain.

3.5 The old and new national Systemic Evaluations

Much like the earlier Systemic Evaluations, the 2022 SE will be carried out every three years on a nationally representative sample covering Grades 3, 6, and 9.

Average marks achieved in the Systemic Evaluation assessment in mathematics administered to Grade 3 learners in 2001 and 2007 ranged from 23% (the lowest) in the Northern Cape, and 35% (the highest) in KwaZulu-Natal in 2001; in 2007, Limpopo, where learners averaged 29%, was the lowest, and the Western Cape, with 48%, was highest.

In 2007, 53,972 Grade 3 learners from 2,327 primary schools participated in the Systemic Evaluations. Of these, 8,537 learners (15.8%) from 290 schools performed at or above both the literacy and numeracy benchmarks of 50%. Another 3,976 learners (7.3%) performed at or above the numeracy benchmark only, and 4,057 learners (7.5%) at or above the literacy benchmark only. Only 2,706 learners (5%) in 80 schools countrywide achieved 70% or more in numeracy (DoE 2008).

The new Systemic Evaluations will be administered to Grade 4, 7, and 10 learners in the first half of 2022 (instead of to Grade 3, 6, and 9 learners), with results anticipated a year later. A highly complex research design will be used, with support from a specialist international education assessment agency. A matrix-sampled test administration method makes it possible to cover many more items in the assessment without increasing the test burden faced by each learner, as each learner deals with only some items. The intention is not to report individual test scores but to estimate scale score distributions for groups of learners using Item Response Theory (IRT) methods. The assessments will cover work from all school terms and include tests with questions from each of the grades making up each phase (e.g. the Grade 3 test will have questions covering Grades 1, 2 and 3 curriculum domains). The 2022 tests will include

an investigation of school-level processes associated with implementing the Whole School Evaluation (WSE) and district support in a sub-sample of schools.[6]

In summary, the sample-based Systemic Evaluations will report on learning outcomes at the national and the provincial level, but unlike the ANAs or Grade 12 NSC, not at the school, learner or class level. These assessments will, therefore, not provide detailed information for accountability pressure and improvement. Learning outcomes and learner performance will need to be measured by other means at the individual and school levels. This is where the SBAs come in – they provide an opportunity to secure the school-specific, accountability-based improvement described in the National Development Plan (National Planning Commission 2012).

3.6 Opportunities for school-level accountability through common SBAs

SBAs comprise practical assessment tasks and end-of-term or end-of-year examinations to give learners, parents, and teachers an indication of what learners know, understand, and can do. These results are also used to determine whether learners can move on to the next grade. They are therefore high-stakes results, though decisions typically rest on a combination of more than one single assessment task or test. They are typically developed, administered, and marked by teachers and are therefore generally aligned with the curriculum. However, there are concerns about their standardisation and quality. They therefore suffer from some of the same problems as the ANAs, namely concerns about score variation, the quality and standard of tasks, reliability, the possibility of parental assistance, item quality and mark inflation, and moderation or standardisation (Umalusi 2019).

Assessment literacy is defined loosely as an individual understanding of the assessment concepts, tools, and procedures likely to influence educational decisions (Popham, 2011). Poor assessment literacy among teachers has long been a serious concern (Umalusi 2004; DoE 2003b; DoE 2005; Van der Berg & Shepherd 2010; DBE 2010; Carnoy et al. 2008, 2012). Therefore, strengthening formative assessment is crucial, and common assessments done by districts and provinces provide an important opportunity to do this.

Common SBAs can be good measures of learner-level and school-level performance. Provided that they are externally moderated, they can be made credible enough to use to create accountability pressure at the school level without the results being at risk of manipulation. SBAs can allow for better teacher development, planning of school improvement, and classroom-based remediation of learning weaknesses. The level of detail that SBAs provide can supplement the information in the newly-designed sample-based Systemic Evaluations to target the teaching and learning support provided by education officials to specific schools, teachers, and classrooms.

A questionnaire of 28 questions administered by the first author at the district

6. These are processes concerned with basic functionality, governance, leadership and management, quality of teaching, learning and educator development, curriculum provisioning, school safety, infrastructure, and parental engagement.

directors' meeting with the Minister of Basic Education convened in June 2016 provides a glimpse into the underlying status of assessment in the country, even though only 44 out of 70 district officials returned the questionnaire at the meeting's end. This survey relates particularly to so-called common assessments, i.e. SBAs for particular grades written by all schools in a district or province. Fifty per cent of the district managers who responded indicated that common assessments were administered in Grades 3, 6, and 9 in their province, and 43% of district managers reported that common assessments were administered in their districts. In other words, common assessments were fairly widely administered, and most district officials indicated that these assessments were undertaken quarterly.

Assessments were, first of all, marked almost exclusively (86%) by a teacher of the same grade as the learner, before moderation. Interestingly, district and provincial common assessments focused on the higher grades, although provincial assessments were more frequently found in Grades 3, 6, and 9. Common assessments also took place in other grades, but not to the same extent as in the highest three grades (Grades 10 to 12). It is also worth noting the high levels of moderation of assessments by district or provincial officials, pointing to an appreciation of the importance of some level of control and standardisation in the provincial- and district-level assessments.

Data from schools were predominantly captured electronically, with the majority of respondents (91%) using the DBE-issued SA-SAMS software. Most district officials reported that schools had received formal feedback on their assessments, and that this information was used to identify and support underperforming schools. For about one in five of the common assessments, assessment data were separated at the item level, which is best for providing helpful feedback to learners and teachers.

SBAs are already marked by teachers. Strengthening the external moderation of the marking of the SBAs provides an opportunity to strengthen the credibility of decisions made with SBA results and allows an existing form of assessment to inform school improvement and accountability. The use of SBA results to gauge the real levels of learning in schools within a particular district is possible, provided that these are externally moderated, and the moderated results are used to adjust school-level results (if necessary) in the more general universal learning assessments administered in schools.

Common assessment results provide the opportunity to improve the evidence-based planning and targeting decisions made within districts in which the same assessment papers are administered. In the 2017 school monitoring survey, 64% of primary school principals and 95% of secondary school principals indicated that they participated in common assessments set by the province or district. By province, principals of primary schools indicated participation in common assessments least in the Western Cape (44%) and Free State (50%), and most in the Eastern Cape (74%). Common sets of SBAs administered in a group of schools in a district, or indeed within a group of districts in a province, can be used to provide information for improvement, provided that external moderation of marks is strengthened, to ensure SBA consistency within schools and comparability across schools. As comparability in SBAs has been variable, especially in schools with low performance, recent policy guidelines that focus on improving SBA administration and moderation are a step in the right direction. A dedicated focus on the primary grades is needed to compensate

for the limitations in the design features of SBAs versus more systemic assessments like the ANA tests (DBE 2020; 2022; Chetty 2016, 252–254).

As the DBE works to improve the diversity and depth of assessment types in the national assessment system, SBAs can be strengthened. They can be used to provide better information to learners, educators, and parents on how children are learning. Helping subject advisors in decision-making and targeting their school-level support can be one tool in challenging learning poverty in South Africa's schooling system. Schools and the teachers and learners within them can then use the information they produce at the learner and classroom level to improve instruction, teacher practice, and learning outcomes more credibly.

Venkat and Sapire (this volume) note the increase in South African research on early grade mathematics assessment since 2010, and cite a DBE-led evidence-based programme that has been rolled out to support FP teachers in dealing with the learning losses incurred in 2020, and strengthen learners' capabilities on core topics in mathematics (Rhodes University 2021). To support improvements in FP numeracy, the programme uses FP diagnostic assessments to promote effective teaching and learning of mental calculation strategies for Grade 3 mathematics (Venkat & Sapire, this volume).

The analysis in this chapter confirms that there is an appreciation for common SBAs within the schooling assessment system. If the data systems can allow subject advisors to access information from the common assessments within a given district or province, it will allow subject advisors and teachers to tailor teaching to the needs of individual learners and schools in the specific learning areas identified in the particular district or province.

Better quality assurance of school-based assessment items and tests will strengthen the tracking of the actual learning progress of individual learners, and the targeting of resources and support required for principals, teachers, and subject advisors to develop skills and knowledge in numeracy in the FP. Packaging credible data from the common SBAs for subject advisors could help them determine where to focus their efforts in developing numeracy skills. Critically, it must be understood that this information must primarily be used to support and improve teaching and learning, and not to sanction or punish individual primary schools.

4 Conclusion and the way forward: towards an assessment system

It is clear that the South African assessment system is still evolving. At the classroom level, it includes assessments in the form of SBAs that provide feedback to learners, teachers, and parents. If the use of common assessments is harnessed well, then SBAs can also start to provide more feedback to school management and education authorities at the system level. It remains important, though, to guard against common SBAs becoming a source of high-stakes accountability, as this is likely to undermine its acceptance and value in assessment.

The new Systemic Evaluations are due to be implemented nationally by mid-2022, with indications of learning trends and performance to be released in 2023. International assessments remain important for benchmarking purposes and for tracking performance over time, but as these surveys are carried out every four or five years, they are too infrequent to adequately influence teaching and learning policy and practice, although they provide a valuable cross-sectional analysis of education system performance for policy and planning purposes.

Turning to what we know about mathematics performance in the Foundation Phase, the available assessments and literature indicate that performance is generally poor and that many learners lack the foundational skills they need to build on in the Intermediate Phase (Venkat & Sapire, this volume). While there is limited evidence about trends, the measured progress in the Western Cape Systemic Tests contrasts with the stagnation seen in the national performance of Grade 5 learners in TIMSS from 2015 to 2019. What progress there might have been, if any, would clearly not weigh up against the much bigger losses caused by lockdowns, school closures and rotational school attendance during much of 2020 and 2021 due to Covid-19.

Yet even when such an assessment system grows to its full potential, there will still be an important shortcoming in our progress. This is what the Sustainable Development Goals Report (United Nations 2019) refers to as learning poverty: the lack of minimum proficiency in reading and mathematics by the time a learner reaches the middle of primary school.

Despite the importance of mathematics in the early grades, it is not yet systematically assessed. We propose that, until it is, SBAs in the primary grades should be strengthened to support the work of subject advisors through better moderation practices at district level, and processes for adjusting original marks, using the moderated scores to get a more accurate reflection of learning levels and progress, should also be strengthened.

The findings that there are already considerable learning difficulties in mathematics by the end of Grade 1 are sobering (Spaull et al. 2022). Opportunities to deal with these difficulties include developing assessment tools to determine school preparedness in pre-numeracy skills and knowledge in the reception grade (Grade R) (which now enjoys near-universal participation), well before young children enter Grade 1. In addition, we propose that research on standards for numeracy and pre-numeracy be carried out to generate content for SBAs for Grade R, and to generate the information required to boost and track numeracy and mathematics outcomes later on.

Finally, administrative data systems will need to be responsive to this need for a more focused and refined use of SBAs. This should improve numeracy outcomes within the system and also allow researchers to track trends in learning poverty in mathematics in the Foundation Phase and beyond.

References

ACER Centre for Global Education Monitoring (ACER-GEM) (2019). Minimum proficiency levels: described, unpacked and illustrated. Version 2. UNESCO Institute for Statistics. https://research.acer.edu.au/monitoring_learning/47

Azevedo, J.P. & Montoya, S. (2021). How the SDG 4.1.1 Framework and learning poverty can help countries focus their education policy response to COVID-19. UNESCO Institute of Statistics.

Barber, M. & Mourshed, M. (2007). How the world's best-performing school systems come out on top. McKinsey & Company.

Best, M., Knight, P., Lietz, P., Lockwood, C., Nugroho, D., & Tobin, M. (2013). The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries. Final report. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), 7–74.

Braun, H. & Kanjee, A. (2006). Using assessment to improve education in developing nations. *Improving education through assessment, innovation, and evaluation*, 1–46.

Carnoy, M. & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational evaluation and policy analysis*, 24(4), 305–331.

Carnoy, M., Chisholm, L., & Baloyi, H. (2008). Towards understanding student academic performance in South Africa: a pilot study of Grade 6 mathematics lessons in Gauteng province. South Africa: Human Sciences Research Council.

Carnoy, M., Chisholm, L., & Chilisa, B. (2012). The low achievement trap: Comparing schooling in Botswana and South Africa. Human Sciences Research Council Press.

Cartwright, F. (2013). Review of the national ANA programme, South Africa. World Bank.

Chetty, M. (2016). Comparing school-based assessments with standardised national assessments in South Africa. Doctoral dissertation: University of the Witwatersrand.

Clarke, M. (2012a). Measuring learning: how effective student assessment systems can help achieve learning for all (No. 10058). World Bank.

Clarke, M. (2012b). What matters most for student assessment systems: A framework paper. SABER–Student Assessment Working Paper, 1.

Conley, D. (2015). A New Era for Educational Assessment. *Education policy analysis archives*, 23, 8.

Darling-Hammond, L. & Wentworth, L. (2010). Benchmarking learning systems: Student performance assessment in international context. Stanford Center for Opportunity Policy in Education. Stanford University, CA.

Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014). Accountability for College and Career Readiness: Developing a New Paradigm. *Education policy analysis archives*, 22(86),1.

Department of Basic Education (DBE). (2010). Action Plan to 2014: Towards the Realisation of Schooling 2025. Pretoria: DBE.

Department of Basic Education (2011a). Report on the Annual National Assessments of 2011. Pretoria: DBE.

Department of Basic Education (2011b). Annual National Assessments 2011: A guideline for the interpretation and use of ANA results. Pretoria: DBE.
Department of Basic Education (2015a). Action Plan to 2019: Towards the Realisation of Schooling 2030. Pretoria: DBE.
Department of Basic Education (2015b). Annual National Assessment 2014. Diagnostic Report: Foundation Phase. Mathematics and Home Language. Pretoria: DBE https://www.education.gov.za/Portals/0/Documents/Reports/REPORT%20ON%20THE%20ANA%20OF%202014.pdf?ver=2014-12-04-104938-000.
Department of Basic Education (2019). School Monitoring Survey 2017/18: Technical report. Pretoria: DBE.
Department of Basic Education (2020). Action Plan to 2024: Towards the Realisation of Schooling 2030. Pretoria: DBE.
Department of Basic Education (2022). Guideline on the quality assurance of School Based Assessment (SBA) for Grades 10 to 12, March 2022. Pretoria: DBE.
Department of Education (DoE) (1995). White Paper on Education and Training. Notice 196 of 1995. Pretoria: DoE.
Department of Education (2001). National Policy on Whole-School Evaluation. Government Gazette Vol.433, No. 22512 of July 2001. Pretoria: DoE.
Department of Education (2003a). National Report on Systemic Evaluation: Foundation Phase (Learners with Disabilities in Special Schools. Pretoria: DoE.
Department of Education (2003b). Systemic Evaluation: Foundation Phase (Mainstream). Pretoria: DoE.
Department of Education (2005). Grade 6 Systemic Evaluation: Intermediate. National Report. Pretoria: DoE. http://www.hsrc.ac.za/research/output/outputDocuments/3580_Grade6National.pdf
Department of Education (2008). Grade 3 Systemic Evaluation 2007. Leaflet. Pretoria: DoE.
Department of Education. Undated leaflet text. Systemic Evaluation: Grade 3 Literacy and Numeracy results. Released 2008. Pretoria: DoE.
Dixit, A. (2012). Bureaucracy, its Reform and Development. *Review of Market Integration*, 4(2), 135–157.
Ferrer, J.G. (2006). Educational assessment systems in Latin America: Current practice and future challenges. PREAL.
Goldsmith, J. 2009. Pacing and time allocation at the micro-and meso-level within the class hour: Why pacing is important, how to study it, and what it implies for individual lesson planning. *Bellaterra Journal of Teaching & Learning Language & Literature*, 1(1), 30–48.
Gustafsson, M. (2015). Some more angles on ANA Grade 9 mathematics versus the Grade 12 examinations. Unpublished note.
Hoadley, U. & Muller, J. (2014). Testing, testing: Investigating the epistemic potential of systemic tests. Mimeograph. Cape Town: University of Cape Town https://documents1.worldbank.org/curated/en/339291529320964248/pdf/127304-Education-in-South-Africa.pdf
Lockheed, M.E. & Verspoor, A.M. (1991). Improving primary education in developing countries. Oxford University Press for World Bank.
National Planning Commission (2012). National development plan 2030: Our future – make it work. Pretoria.
Nuga Deliwe, C. (2017). An analysis of the measurement of the progress in learning outcomes at the country level: the case of South Africa. MEd thesis. Johannesburg: University of the Witwatersrand. https://hdl.handle.net/10539/25944

Phelps, R.P. (2012). The effect of testing on student achievement, 1910–2010. *International Journal of Testing*, 12(1), 21–43.
Popham, W.J. 2011. Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, 46(4), 265–273.
Reddy, V., Winnaar, L., Harvey, J., Hannan, S., Isdale, K., Arends, F., & Juan, A. (2022). The South African TIMSS 2019 Grade 5 Results. Building Achievement and Bridging Achievement Gaps. Pretoria: Human Sciences Research Council.
Republic of South Africa (1996). The National Education Policy Act (Act 27 of 1996). Pretoria: Government Printer.
Rhodes University (2021). Education Matters. Issue 1, September 2021.
Rosenkvist, M.A. (2010). Using Student Test Results for Accountability and Improvement: A Literature Review. OECD Education Working Papers, No. 54. OECD Publishing (NJ1).
SACMEQ (2021). The SACMEQ IV Project in International: A study of the conditions of schooling and the quality of education. SACMEQ: Gaborone.
SADTU (2011). SADTU'S discussion document in response to ANA report. Retrieved 15 Aug 2016 from: http://www.sadtu.org.za/docs/disc/2011/ana_report.pdf
SADTU (2014). SADTU: Over-emphasis on tests, assessments. Retrieved June 2016 from: http://www.news24.com/SouthAfrica/News/Sadtu-Over-emphasis-on-tests-assessments-20140928
Schäfer, D. (2020). 2019 Systemic Test results show Western Cape education strategies are working. Statement by MEC, Western Cape Education Dept. https://wcedonline.westerncape.gov.za/news/2019-systemic-test-results-show-western-cape-education-strategies-are-working
Spaull, N. (2015). An Overview and Interrogation of the Annual National Assessments in South Africa. Presented at the Catholic Parliamentary Liaison Office round table discussion, 4 May 2015. http://www.cplo.org.za/wp-content/uploads/2018/02/BP-391-The-Annual-National-Assessments-Sept-2015-1.pdf
Taylor, S. (2015). Why have annual national assessments? *GroundUP*, 30 Sept 2015. https://www.groundup.org.za/article/why-have-annual-national-assessments_3348/
Taylor, N., Van der Berg, S., & Mabogoane, T. (2013). Creating effective schools. Cape Town: Pearson.
Umalusi (2004). Investigation into the Standard of the Senior Certificate Examination: A Report on Research conducted by Umalusi. Pretoria.
Umalusi (2019). Colloquium presentation on Policy and School Based Assessment. https://www.umalusi.org.za/wp-content/uploads/2019/09/DBE_Dr-Poliah-SBA-Colloquium.pdf
UNESCO (2013).Training Tools for Curriculum Development: A Resource Pack. Geneva: UNESCO-IBE.
UNESCO (2015). Education 2030 Incheon Declaration and Framework for Action: Towards inclusive and equitable quality education and lifelong learning for all. Paris: UNESCO.
United Nations (2019). The Sustainable Development Goals Report 2019. New York: United Nations.
Van der Berg, S. (2015). What the Annual National Assessments can tell us about learning deficits over the education system and the school career. *South African Journal of Childhood Education*, 5(2), 28–43

Van der Berg, S., Gustafsson, M., & Malindi, K. (2020). Education and skills for the economy and links to labour markets in South Africa. Report for the Economic Task Team of the National Planning Commission, released as Position Paper. Research on Socioeconomic Policy (RESEP), Stellenbosch University. https://resep.sun.ac.za/wp-content/uploads/2020/06/Education-and-skills-for-the-economy-and-links-to-labour-markets-in-South-Africa.pdf

Van der Berg, S., Hoadley, U., Galant, J., Van Wyk, C., & Böhmer, B. (2022). Learning Losses from COVID-19 in the Western Cape. Evidence from systemic tests. Research on Socioeconomic Policy (RESEP), Stellenbosch University.

Van der Berg, S., & Hofmeyr, H. (2018). Education in South Africa. Background note to World Bank's Systematic Country Diagnostic, An Incomplete Transition: Overcoming the Legacy of Exclusion in South Africa. https://documents1.worldbank.org/curated/en/339291529320964248/pdf/127304-Education-in-South-Africa.pdf

Van der Berg, S. & Shepherd, D. (2010). Signalling performance: Continuous assessment and matriculation examination marks in South African schools (No. 28/2010).

Van der Berg, S., van Wyk, C., Selkirk, R. & Hofmeyr, H. (2021). Learner flows through schools. Available at SSRN 4009654.

World Bank (2018). What is Learning Poverty? Available at: https://www.worldbank.org/en/topic/education/brief/what-is-learning-poverty (Accessed 8 May 2022)