AOSIS

Beyond test scores: A framework for reporting mathematics assessment results to enhance teaching and learning

CrossMark
click for updates

Authors:
Meshack Moloi[1]
Anil Kanjee[1]

Affiliations:
[1]Department of Primary Education, Tshwane University of Technology, South Africa

Corresponding author:
Meshack Moloi, moloimq@tut.ac.za

In this article we propose a framework for reporting mathematics results from national assessment surveys (NAS) such that effective use of the resulting reports can enhance teaching and learning. We explored literature on factors that may contribute to non-utilisation of assessment data as a basis for decision-making. In the context of South Africa, we identified the form and formats in which results of NAS are reported as a possible limiting factor to the effective use of summative assessment results for formative purposes. As an alternative, we propose a standards-based reporting framework that will ensure accurate measurement of, and meaningful feedback on, what learners know and can do. We illustrate how, within a properly designed reporting framework, the results of a NAS in mathematics can be used for formative purposes to enhance teaching and learning and, possibly, improve learner performance.

Background

National assessment surveys (NAS) have been implemented in South Africa since the abolishment of the apartheid education system in 1996, and have evolved over time, changing in name, purpose, design, scope and frequency (Department of Education [DOE], 2005; Kanjee, 2007). National assessments are defined as 'regular and systematic measurement exercises designed to determine what students have learned as a result of their educational experiences' (UNESCO, 2000, p. 14). They are different to public examinations in that their goal is to inform policy for the education system as a whole, rather than to certify individual learners. These assessments may be administered to an entire cohort (census testing) or to a statistically chosen group (sample testing) and may also include background questionnaires administered to learners, teachers or education officials to obtain additional information for use in interpreting learner scores. Braun and Kanjee (2007) note that the utility of the data generated from these assessments depends on the quality and relevance of the assessment, the thoroughness of the associated fieldwork, as well as the expertise of those charged with the analysis, interpretation, reporting and dissemination of results.

Between 1996 and 2015, the form, format and frequency of NAS in South Africa have changed significantly – from national sample-based surveys administered in selected grades to assess mathematics and language performance every 3 to 4 years to annual national census-based assessments (Department of Basic Education [DBE], 2013). In addition to NAS, individual provinces such as the Western Cape and North West also administer provincial assessments and common tests, respectively, which focus on different subject areas and grades (Hoadley & Muller, 2016). While there have been marked improvements in the administrative and logistical processes of the assessments, a challenge that remains unresolved pertains to the meaningful reporting and effective use of the results from these assessments for enhancing teaching and learning.

The phenomenon of non-utilisation or under-utilisation of national assessment data in influencing decision-making in South Africa has been noted as a matter of concern (Kanjee & Moloi, 2014; Kanjee & Sayed, 2013). Yet there has been a growing body of research which indicates that, when the results of NAS are reported, disseminated and utilised properly, there are observable improvements in learner performance (Klinger, DeLuca & Miller, 2008; Ravela, 2005; Schiefelbein & Schiefelbein, 2003). It would appear, therefore, that one challenge facing teachers in South Africa is the inadequacy in meaningful reporting and effective utilisation of evidence from assessment. Meaningful reporting includes finding effective ways of converting raw data into information that could inform decision-making. At classroom level, 'meaningful information'

refers to information that the teacher could use for determining what learners at a particular grade level know or do not know, and can or cannot do, and to develop relevant interventions to address specific learning needs of learners.

In this article, we propose a framework for reporting results from NAS for use at the school level, and demonstrate how this framework can be applied to identify specific learning gaps of learners and provide guidelines to address identified learning gaps. Although the reporting framework is exemplified in mathematics, its applicability extends to any school subject. First, we contextualise the proposed framework by providing a brief overview of reporting of assessment results as regulated in the South African Curriculum and Assessment Policy Statement (CAPS). Next, we provide a conceptual framework for reporting and using assessment data, highlighting the challenges impeding effective use of data. This is followed by a description of the proposed reporting framework, its underlying philosophy and an exemplar school report, in which we highlight its practical application and implications for enhancing teaching and learning. We conclude the article by listing areas for further research to optimally use summative assessment results for formative purposes.

Reporting mathematics results

The view taken in this article is that mathematics as a subject embodied in most school curricula is often characterised as a hierarchical cumulative body of knowledge. As such, the foundations of relevant mathematics content at a particular grade level are developed in the previous grade and the acquisition of complex capabilities builds on relatively basic concepts. For instance, young children progressively develop a 'number concept' often demonstrated by first being able to organise concrete objects before they can manipulate abstract concepts. Given this unique nature of the subject, assessment and use of assessment results in mathematics seem to present specific challenges to mathematics teachers (Webb, 1997).

In order to enhance learning of mathematics knowledge and skills, as well as to identify and address specific learning gaps revealed by assessment results, teachers must have full mastery of the mathematics content area as well as a thorough understanding of the hierarchical nature of the subject. Similarly, for assessment data to be useful for teachers to enhance learning in mathematics, it becomes more critical that the data be organised and reported in a manner that reflects the nature of mathematical knowledge and how learning in mathematics takes place. In practice, this implies that learner performance results reported with the intention of enhancing teaching and learning must, at a very minimum, provide information on what learners at a particular point know and can do and, at the same time, what they are potentially ready to learn (Vygotsky, 1962).

One limitation in reporting the results of NAS is the tendency to adopt a norm-referenced approach in which schools, and even learners, are ranked and compared with one

another according to their performance in the tests (Green, 2002). The 'league tables' that often emanate from norm-referenced reporting are notorious for attracting resistance to assessment and evocation of negative feelings among teachers. This undesirable phenomenon was reported in the United Kingdom (Goldstein, 2001) and was also observed in South Africa when teacher unions boycotted the administration of the Annual National Assessment (ANA) because they perceived the assessment as 'an onslaught on teachers with no intention to improve the [*education*] system' (Nkosi, 2015).

In this article we argue that the vital element that links NAS results to enhancing teaching and learning is a reporting framework that provides accurate measurement and meaningful feedback on what learners know and can do (Griffin, 2009). Importantly, the reporting framework must reflect the structure of mathematical knowledge as well as the process of learning in mathematics. It must embrace what Griffin (2009) defines as 'criterion-referenced interpretation' and involve measurement coupled with 'skills audits' in which responses to clusters of items in a test are interrogated to identify an underlying construct. For example, a Grade 6 learner who only responds correctly to test items that involve counting forward with whole numbers is demonstrating mathematical understanding that is at a lower level than a learner who, in addition, also responds correctly to items that involve doing calculations using fractions.

We are aware of the critical distinctions that some make between NAS and school-based assessments in terms of how the assessments are impacted upon differently by the socio-economic contexts within which learning takes place (Nichols & Berliner, 2007). Disparities among school-based testing procedures (Webb, 1997), possible variations in curriculum coverage across schools and other differences may lead to questioning the fairness of the NAS. Within the limits set by these caveats, we take the view advanced by Dunne, Long, Craig and Venter (2012) that a good balance between NAS and school-based assessment is possible with proper test design and effective reporting of results. Proper test design encapsulates considerations of the extent to which the test adequately elicits meaningful information on what learners know, can or cannot do in the subject area of interest. Effective reporting involves 'packaging' and presenting the results in ways that enable the target users to initiate appropriate interventions for improvement. In particular, the South African ANA model, where all learners in a grade, and not just typical representative samples, participate in the NAS, enhances both the feasibility and the practicability of the balance that Dunne et al. recommend. A standards-based reporting framework (SRF) that allows criterion-referenced interpretation of test results in these conditions stands to benefit policymakers, teachers, parents and even learners.

It is important to recognise that the value of the results of an assessment is optimal when they are used within the confines

of the purpose for which the assessment is designed. On the one hand, school-based assessments include formative assessments whose purpose is to inform teaching and learning while a lesson is in progress and are, therefore, developmental in design. On the other hand, schools also conduct summative assessments which basically measure the extent to which learning has taken place after several lessons were delivered. Testing that characterises NAS falls under the latter category of assessments. Our argument is that, within a properly designed reporting framework, the results of summative assessments can be used for formative purposes to enhance the quality of teaching and learning.

Curriculum and Assessment Policy Statement reporting framework

Assessment results in basic education in South Africa, both school-based as well as results of common examinations and NAS, are recorded and reported according to a framework that is prescribed in the CAPS document. The 'framework' has three key elements designed in seven levels, namely rating codes, descriptions of competence and percentages (Table 1). We examined the CAPS framework against our proposed conceptual model and noted some disparities which we consider to be of material importance.

The CAPS framework prescribes that assessment data will be organised into fixed percentage bands with the lowest band ranging from 0% to 29% and the highest from 80% to 90%. Within this framework, a learner obtaining a minimum score of 50% is deemed to be functioning at the 'adequate achievement' level (DBE, 2011). We argue that by organising and summarising results using percentages, the CAPS framework does not provide any information on the specific knowledge and skills that learners have or have not mastered. For example, a score of 56% provides no information on what should be done for enhancing teaching and learning. We extracted a table that summarises NAS results in a typical ANA report compiled according to the CAPS framework to point out some of the conceptual challenges that compromise CAPS-based reports (Table 2).

In Table 2, which contains information that was put in the public arena, the raw score bands have been summarised using the seven codes and the corresponding descriptions of competencies. No substantive qualitative analysis has been

presented that provides detailed information on what learners at each score band in Table 2 know and can do. In a survey to assess the extent to which South African teachers used the ANA results, Kanjee and Moloi (2014) reported that up to 26% of the teachers in their study were of the view that the ANA reports did not provide any new information that they did not already know. An inference that could be made from these perceptions was that these teachers were, logically, not likely to utilise these results. Our view is that perceptions of inadequacy in the content of the NAS reports could contribute to non-utilisation of the results for enhancing teaching and learning, which in turn could lead to perpetuation of underperformance in the system.

The fixed percentage bands as exemplified in Table 2 do not accommodate variations in the difficulty of tests. For instance, learners who score in the range of 0% – 29% are categorised as functioning at the 'Not achieved' (L1) level and those who score in the range of 80% – 100% as functioning at the 'Meritorious' (L7) level, regardless of the difficulty of the specific test. We should be aware that on an easy test percentage correct raw scores tend to be higher than in a difficult test; however, in the same test learners of higher ability are expected to score higher than their counterparts of lower ability (Bond & Fox, 2007). So, a meritorious achievement in an easy test may not necessarily be the same in a difficult test. It is also not possible to set two different tests that have exactly the same level of difficulty, even if the exact same test specifications are followed. It is for this reason that test equating measures have been introduced to adjust for differences in test difficulty (Kolen & Brennan, 1995).

The net effect of these inconsistencies is that the users of CAPS-based reports may have either superficial or distorted knowledge about the performance of learners. Moreover, the use of this reporting framework implies that a higher conceptual workload is placed on teachers and school leaders by expecting them to be able to record, report, categorise and address learner needs across seven levels of performance. It could prove unrealistic to expect a teacher to keep track of and provide differentiated support across seven categories of learners in a class. To mitigate the observed shortcomings

TABLE 1: Curriculum and Assessment Policy Statement framework for reporting assessment results.

Rating codes	Description of competence	Percentage
7	Outstanding achievement	80–100
6	Meritorious achievement	70–79
5	Substantial achievement	60–69
4	Adequate achievement	50–59
3	Moderate achievement	40–49
2	Elementary achievement	30–39
1	Not achieved	0–29

Source: Department of Basic Education. (2011). *Curriculum and assessment policy statement. Grades 7–9. Mathematics* (p. 158). Pretoria: DBE. Retrieved from https://www.education. gov.za/LinkClick.aspx?fileticket=uCNqOwfGbmc%3d&tabid=573&portalid=0&mid=1629

TABLE 2: Percentage of Grade 6 learners by achievement level in mathematics.

Province	L1	L2	L3	L4	L5	L6	L7
Eastern Cape	47.5	19.0	16.7	8.5	5.0	2.0	1.2
Free State	31.7	19.3	22.5	12.4	7.7	3.7	2.6
Gauteng	25.6	16.7	19.0	14.1	11.5	7.2	5.8
KwaZulu-Natal	30.4	18.6	20.6	13.3	9.5	4.6	3.0
Limpopo	49.0	18.9	16.7	7.9	4.7	1.8	1.0
Mpumalanga	45.6	21.0	17.3	8.1	4.9	1.9	1.1
Northern Cape	44.1	19.1	16.2	9.4	6.3	2.8	2.1
North West	39.6	21.0	18.7	10.3	6.0	2.7	1.7
Western Cape	27.1	17.9	17.3	12.7	10.7	7.2	7.2
National	36.2	18.7	18.6	11.3	8.0	4.2	3.1

Source: Adapted from Department of Basic Education. (2013). *Report on the annual national assessment of 2013: Grades 1 to 6 & 9* (p. 59). Pretoria: DBE. Retrieved from https://www.education.gov.za/Portals/0/Documents/Reports/ANA%20Report%202013%20(2).pdf?ver=2013-12-05-123612-000

L1, L2, … refer to the rating codes in Table 1.

of the CAPS reporting framework and ensure that assessment results are reported in ways that provide quality information to support users and enhance the teaching and learning process, in this study we propose an alternative model that is underpinned by a theory of data use proposed by Breiter and Light (2006).

Conceptual model for using assessment data

Breiter and Light (2006) developed a conceptual model for using data to inform decision-making in the management of education districts. Central to their model is a definition of decision-making as a 'highly complex, individual cognitive process that can be influenced by various environmental factors' (Breiter & Light, 2006, p. 208). They discourage notions of decision-making that require innumerable disparate pieces of data and suggest rather that decision-making involves intelligibly reducing (collecting and organising) large amounts of data, converting the data (summarising and analysing) into information and transforming the information into context-related knowledge to inform action (prioritising and synthesising). Their model comprises four key elements, namely data, information, knowledge and decision-making. While not necessarily focusing specifically on assessment data, the model also accounts for the multiplicity of data and data sources that decision-makers in education must deal with. We adapted the model by Breiter and Light and developed a conceptual model to report assessment results so that the information can be used to enhance the quality of teaching and learning in schools (see Figure 1).
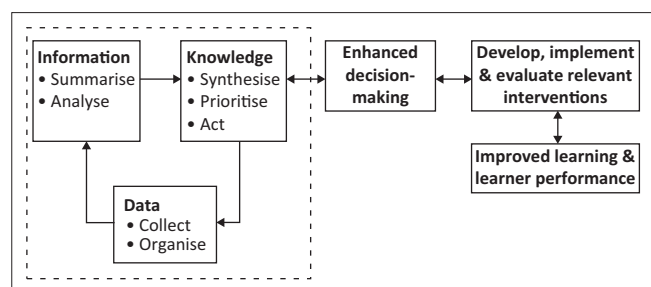
The basic element of the model by Breiter and Light (2006) is *data*. This includes, but may not be confined to, raw statistical data like test scores, for instance. Once teachers, school leaders or other decision-makers become aware of a situation of educational importance that needs to be addressed such as the persistent underperformance in mathematics and related issues, for example language ability or home background, appropriate data, often presented in numerical formats, need to be collected and analysed to gain detailed insight into the nature of the phenomenon. We agree with Breiter and Light that once data is collected it must be organised in

Source: Adapted from Breiter, A., & Light, D. (2006). Data for school improvement: Factors for designing effective information systems to support decision-making in schools. *Journal of Educational Technology & Society*, 9(3), 206–217 (p. 210). Retrieved from https://www.j-ets.net/ETS/journals/9_3/18.pdf

FIGURE 1: Conceptual model for using assessment data.

ways that will make it meaningful to the users. But data do not speak for themselves; hence the continued reporting of mathematics assessment results in raw scores in our schools seems to have influenced neither teaching nor learning. Data must be reported in ways that allow key users, such as district officials, school leaders and teachers, to decode the data (Coburn, Honig & Stein, 2009).

Information, in the model, refers to data that has been appropriately analysed and summarised so that it sheds light on the nature and extent of the identified problem. Thus, any report must communicate relevant *information* that will either add to what is known or will illuminate a new area of interest or further investigation. For example, a mathematics school report should provide information on what individual or groups of learners know or do not know and can or cannot do in mathematics, which domains of mathematics pose particular challenges to learners and whether different groups of learners (e.g. boys vs girls or rural vs urban) display comparable levels of proficiency. Later in this article we show how reporting assessment data using meaningful performance standards provides information that empowers key users to make relevant decisions about the challenges of teaching and learning in schools.

Knowledge builds on available information by synthesising what is new with what is already known or available to change the undesirable situation and weighing what the priorities are. For example, a teacher who interprets assessment results and identifies relevant teaching strategies to address revealed learning gaps and explores possible interventions to rectify the situation has *knowledge*. We contend that there is a relationship between the depth and quality of *knowledge* about the education system and the quality of available information. Assessment information that is either incomplete or inaccurate will lead to partial or distorted knowledge about the education system and is likely to result in ineffectual interventions for improvement.

Decision-making is the deployment of acquired knowledge to impact the situation as desired and, in the case of knowledge that comes from assessment, to improve learning outcomes. Breiter and Light (2006) argue that decision-making does not begin with data but with knowledge of needs, for instance needs of learners, teachers or even district officials. It is knowledge that directs the decision-maker to the types of data to collect, the time of collecting it and the methods of transforming the data into actionable decisions. It is important to note that, because of the dynamic nature of the education enterprise, there is a dialectical relationship between knowledge and the context in which teaching and learning take place. On the one hand, there is knowledge of what the assessment results reveal and what needs to be done to turn things around. On the other hand, there is knowledge of new phenomena that may require the collection of new data to understand their nature and thus begin a new cycle of data

collection, generation of information and development of necessary knowledge to make relevant interventions. Decision-making involves leveraging on existing knowledge and prioritising what needs to be done to achieve the desired goals. For instance, when it is known that learner performance in mathematics in a school or district is particularly and continually unacceptable and the factors that contribute to the situation are also known, policymakers and practitioners are confronted with deciding on the best action to take to remedy the situation and count on existing evidence to justify their interventions.

Development and implementation of relevant intervention for any decision to have an impact on practice, relevant interventions that address the key challenges identified must be developed and implemented. In practice, the nature, extent and duration of the intervention may vary depending on what decisions are taken across different contexts. For example, interventions to improve mathematics performance could focus on a specific phase or grade, for example Foundation Phase or Grade 3; these interventions could address specific content areas, for example geometry, or groups of students, for example second language speakers, or the interventions could be conducted as additional lessons before new concepts are introduced or as additional exercises during lessons. The key point is that the intervention developed must be based on addressing challenges identified from the information collected, provided the information is clear, meaningful, easy to read and relevant. In addition, it is critical that some form of evaluation be conducted to monitor progress in implementing interventions.

Improved learning is the ultimate goal within classrooms and schools. Within a learner-centred paradigm, improvement in learning and realisation of observable learner performance hinge largely on the quality of feedback that is given to learners (Saddler, 2010). While Breiter and Light (2006) were specifically referring to feedback in formative assessment in classrooms, we argue that the principle applies to test results as well. When feedback, in the form of information-rich assessment reports, is clear and specific in terms of where the learners are and what the expectations are such that learners are enabled to take control of their learning, it can serve to move learners to the next step. Feedback that provides evidence of what knowledge and skills learners have mastered and which they have not guides teachers to support learners meaningfully and relevantly according to their identified needs (Sloane & Kelly, 2003). It creates a conducive environment wherein teachers and learners work together to realise their shared instructional goals. In this environment learner performance is highly likely to improve.

The implications of adopting the proposed conceptual model for use of assessment data to enhance learning in mathematics are twofold. Firstly, an assessment framework that is based on this conceptual model must have a facility that makes it possible to transform assessment data to information and add value to information to convert it to knowledge. Secondly, because our focus is on mathematics, the framework must be sensitive to the nature of mathematics as a body of knowledge and a school subject and to how learning in mathematics takes place. We argue that these requirements can be met by a standards-based reporting framework.

Challenges to the use of assessment results

Some of the reasons identified for the non- or under-utilisation of information from NAS include poor or non-dissemination of the findings, lack of confidence in the validity of such information among those who have to act upon it, and lack of capacity and absence of appropriate tools to help teachers use the data (Kellaghan, Greaney & Murray, 2009). Other researchers (Hambleton & Pitoniak, 2006; Hambleton & Slater, 1997; Underwood, Zapata-Rivera & Van Winkle, 2010) also blame reports from national assessments for being complex, often couched in statistical jargon that users cannot decipher, difficult to read, and even more difficult to interpret. In South Africa, Kanjee and Moloi (2016) reported that, although the results of NAS had been considered in some policy-related decisions, there had been limited focus on using the results to support improvements in teaching and learning.

More pertinent to the objectives of this article was the finding from Kanjee and Moloi (2016) that a significant percentage of the teachers endorsed 'Agreed' and 'Strongly agreed' when presented with the statement: 'Teachers do not know how to use ANA results to assist learners'. What was more concerning was that approximately 60% of the teachers in this category were teaching in affluent schools that were reputed for high performance. The corresponding percentage of teachers in poorer and often under-performing schools went up to 85%. An inference that could be made from these perceptions was that these teachers were not likely to utilise the results of these national assessments. The situation could be exacerbated by the finding that in the same study, up to 65% of the teachers strongly negated a statement that district officials provided guidance and training on the use of ANA results. Effectively, it would appear that teachers are left to their own devices when it comes to interpreting and using the ANA results.

In his study on how provinces, districts and circuits utilise data from ANA, Govender (2016) reported wide variations in the two provinces and districts that he sampled. Although the education officials were aware of the utility value of the data, Govender (2016) notes that the majority reported that they lacked technical and practical capacity to analyse and interpret the data in meaningful ways. Again, like in the case of Kanjee and Moloi (2014), this finding implies that district officials are unable to provide relevant guidance and support to schools and teachers to enhance their use of assessment results for improving teaching and learning.

In another study Kanjee and Mthembu (2015) explored the extent to which Foundation Phase teachers in one district in South Africa demonstrated understanding of concepts and practices related to both formative and summative assessment. Their study sample included teachers from schools serving communities that ranged from low to high socio-economic status. Kanjee and Mthembu (2015) reported that the teachers demonstrated very low levels of assessment literacy, more so in formative than in summative forms of assessment. Although the sample was quite small and not representative, it is to be noted that these findings were in agreement with the observations that Govender (2016) made. Both district officials and teachers in South Africa appear to lack adequate capacity to utilise assessment data in ways that will enhance the quality of teaching and learning.

Overall, research suggests that, invariably, the interpretation and utilisation of assessment results to enhance teaching and learning in schools are often limited by the competencies of key users, including teachers, school leaders and education department officials (Griffin, 2009; Timperley, 2009). The implication is that reports presenting assessment results must not be dependent on assumed competencies of key users. Thus, these reports should be easy to read, easy to understand, and provide some indication of possible 'next steps' that users can follow to identify and address specific learning gaps or support learners in improving on their current levels of performance. However, limited information currently exists on how such reports need to be developed, nor on what type of analysis is required or how information is best presented to increase the utility value of these reports for teachers.

Exploring the use of standards-based reports

To address the limitations of reporting as discussed in the previous sections we propose a standards-based reporting framework. Green (2002) notes that a standards-based report presents assessment results according to demonstrable mastery of knowledge and skills displayed by learners as evidence of achieving expected learning outcomes. A standards-based report does not 'average' learner scores in a test but identifies what learners know and can do in relation to what the expected standard specifies. Implicit in a standards-based report is a *priori* statement of what is expected of a learner at a particular level or grade. Drawing from the analysis of observed learner scores, the report provides easy-to-read, easy-to-understand and clear guidelines or clues on next steps for teachers (Ravela, 2005). The 'knowledge and skills' expected from learners are generally referred to as 'standards' (Goodman & Hambleton, 2004, p. 148).

In educational circles, a distinction is made between 'content standards' and 'performance standards' (Cizek, 1996; Rodriguez et al., 2011). Rodriguez et al. (2011, p. 18) define 'content standards' as 'what students need to learn'. In the context of South Africa, 'content standards' are spelt

out in the CAPS by grade and by subject (DBE, 2011). 'Content standards' specify the nature and scope of content knowledge, including skills, that a learner must acquire in a given grade. Hambleton (2000) defines performance standards as:

> well-defined domains of content and skills and performance categories for test score interpretation [*that*] are fundamental concepts in educational assessment systems aimed at describing what examinees know and can do. The primary purpose [*of the affected assessments*] is not to determine the rank ordering of examinees, as is the case with norm-referenced tests, but rather to determine the placement of examinees into a set of ordered performance categories. (p. 2)

Thus, while content standards answer the 'what' question, performance standards answer the question about 'how much'. An apt description of the purpose of performance standards proffered by Hambleton (2000) is that they are qualitative and descriptive statements of how much learning has taken place and how much of it is 'good enough'. Our interpretation of Hambleton is that performance standards provide a framework of evidence to be used for placing learners at particular points on a continuum of knowledge and skills according to what they are able to demonstrate when given opportunity to do so, like in a test.

Important features of performance standards are performance levels (PLs) and performance level descriptors (PLDs). Zieky and Perie (2006) describe PLs as:

> general policy statements that indicate the official position on the desirable number and labels of categories to be used in classifying learners according to their knowledge and skills in a particular subject and grade. (p. 3)

Because the knowledge and skills are mapped on a continuum that stretches from low to high, carefully selected scores, known as *cutscores*, are determined to mark and distinguish two consecutive levels of competence or PLs on the continuum (Kaftandjieva, 2010). PLDs are defined as detailed descriptions of 'the knowledge, skills, and abilities to be demonstrated by students who have achieved a particular PL within a particular subject area' (Zieky & Perie, 2006, p. 4). Morgan and Perie (2005) affirm that PLDs are 'working definitions of each of the performance levels [*that*] … define the rigor associated with the performance levels' (p. 5).

Standard setting

The link between 'standards' and effective reporting of learner performance is provided by the process of standard setting. Cizek and Bunch (2007) define standard setting as:

> a process of establishing one or more cutscores on a test for purposes of categorising test-takers according to the degree to which they demonstrate the expected knowledge and/or skills that are being tested. (p. 13)

A cutscore is defined as a point on a score scale which distinguishes two consecutive levels of competence
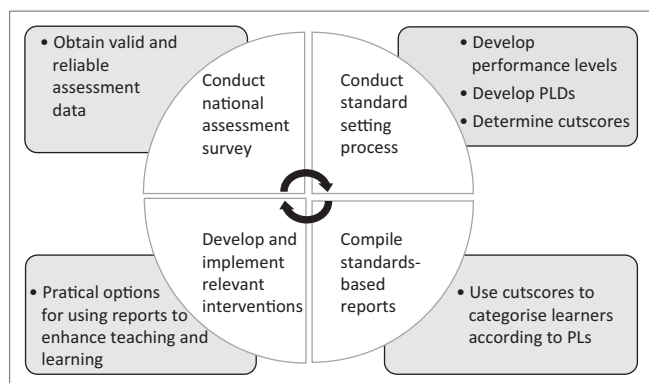
PLs, performance levels; PLDs, performance level descriptors.

FIGURE 2: Standards-based reporting framework for national assessment surveys results.

(Kaftandjieva, 2010). From this definition learners who obtain scores lower than the cutscore will typically be less competent in the affected subject than those with scores above the cutscore.

In practice, the standard setting process involves both quantitative and qualitative inputs. It involves technical analysis of learner responses in raw data form as well as content expert inputs from teams of professionals from relevant stakeholder groups who serve to validate the technical results (Tiratira, 2009). Teams of content experts, preferably teachers of mathematics in this case, develop concise descriptions of typical knowledge and skills that characterise a learner who functions at a particular level. In addition, the teams suggest implications for progression and intervention for learners who, from their test scores, are categorised to be functioning at a particular level. The standard-setting exercise serves to transform raw data into meaningful information as envisaged in our conceptual model in Figure 2. Details on the technical processes of standard setting are available in Moloi (2016).

Framework for standards-based reporting

Having pointed out the shortcomings in the raw score reporting system as currently used in South Africa (DBE, 2013), and highlighted the value of using performance standards as an alternative, we now propose a framework to implement a standards-based reporting system. The primary purpose of this SRF is to present user-friendly reports that promote the formative use of summative NAS results to enhance teaching and learning. For example, end-of-year annual national assessment results that are reported using performance standards should provide teachers with detailed information on specific learner strengths and weaknesses. This information can then be used by teachers to plan and prepare lessons that address identified gaps or reinforce specific knowledge of learners.

The SRF comprises five key sections: (1) rationale for the SRF and how it should be applied in practice, (2) guidelines on the quality, form and format of the raw data obtained from

NAS, (3) process to be followed for conducting standard setting exercises, (4) content and key elements required to compile standards-based reports and (5) practical proposals on how teachers could use the reports to enhance learning and teaching.

Rationale for, and application of, the standards-based reporting framework

The purpose of the SRF is to propose key specifications and practical guidelines for developing information-rich reports for users to provide relevant feedback to enhance teaching and learning in all schools. The SRF addresses the limitations of current guidelines and reporting practices specified in the national curriculum documents for data from NAS.

The framework establishes a coordinated system of processing results from NAS to compile relevant reports of high utility value for use at the different levels of the education system. Depending on the purpose and focus of the NAS, the reports can be used by officials at the national, provincial and district levels as well as by school leaders and teachers.

The practical application of the SRF requires: (1) obtaining valid and reliable data from NAS, (2) conducting appropriate standard setting exercises, (3) compiling relevant standards-based reports for the targeted audience and (4) using reports to develop and implement relevant options to enhance teaching and learning. A diagrammatic representation of the key elements and the flow of the SRF is shown in Figure 2.

Data obtained from national assessment surveys

A primary requirement for the application of the SRF is the availability of relevant and valid data from NAS. The raw data must be in the form of item-level learner results and could be in a scored or raw data format. More importantly, the data must be linked to specific teachers, schools, districts or provinces, depending on the level for which reporting will be conducted. Complete data will include scores for all the schools and learners who participated in the NAS. The data must be valid in the sense that, in the case of mathematics, the learners' scores are inclusive of all the domains of mathematics that the assessment covered, have high reliability coefficients, and be free of errors. The quality of the reports that are based on the SRF depends on the quality of the data used.

Implementation of a standard setting process

The standard setting process involves selected panels of subject content experts, for example teachers of mathematics or mathematics curriculum specialists, to establish both quantitative and qualitative indicators of expected performance standards for the subject. The process provides information to identify those learners who meet the standards, those who fall below the expected standard and

those who exceed the standard. It is important to ensure that the panels are representative across the varying contexts within which teaching and learning takes place in the education system. For instance, in the case of South Africa proportionate representation of urban, semi-urban, rural and farm schools, or quintile categories, in the panels is necessary. The panels receive training on how to develop PLs and PLDs and how to determine valid cutscores. The panels develop generic PLs that categorise learners according to thoroughly discussed generic competencies with clear indications of what the implications are for intervention and progression at each level.

In addition, the panels develop subject-specific PLs and PLDs. They must determine and validate cutscores that mark transitions from one PL to the next in terms of subject knowledge and skills. Using the PLs, PLDs and cutscores determined by the panels, relevant standards-based reports can be developed for national, provincial, district or school level use.

Development of standards-based reports

Developing standards-based reports and reporting accordingly constitute the realisation of the aim of the SRF. Standards-based reporting presents assessment results in report formats that are easy to read, easy to understand and easy to use in decision-making. Compilation of standards-based reports is a process of converting the information from the assessment into useful knowledge that is synthesised and prioritised in ways that enable users to make evidence-based decisions as envisaged in the conceptual model presented in Figure 2.

A standards-based report comprises the following sections: (1) particulars of the institution, (2) how to use the report, (3) performance level definitions and implications, (4) subject-specific performance levels and descriptors and (5) presentation of results by performance levels. Each of these sections is described below, its purpose clarified and, where appropriate, its potential in contributing to enhancing teaching and learning is specified. In order to demonstrate its practical application, an exemplar of a standards-based report is provided in Appendix 1 focusing on reporting at the level of a school. Similar reports can also be compiled at the level of provinces, districts and classrooms.

Particulars of the institution

For ease of identification, basic particulars such as the name of the school, the district and province under which the school falls, the grade and the subject must appear on the first page of the report. As noted in the exemplar in Appendix 1, this report has been compiled for 'City Primary School' which is located in Southern province using Grade 6 mathematics data from the 2015 National Assessment Study.

How to use the report

A note on how the standards-based report should be used is included to specify the steps that teachers should follow to enhance teaching and learning in their classrooms.

Performance level definitions and implications

The use of performance levels, their definitions and implications in standards-based reports is a seismic shift from traditional raw score reporting. A standards-based school report will clarify up front what these features mean and how they help the teacher address learning needs in a differentiated approach as opposed to traditional one-size-fits-all approaches.

Subject-specific performance levels and descriptors

In a standards-based school report all the results from a particular assessment are presented according to performance levels and, preferably, in iconic formats such as pie charts, bar and linear graphs for visual impact. Performance levels do not only enable teachers to adopt differentiated approaches to interventions, but, in the case of mathematics, they mirror the hierarchical nature of the subject. Requisite knowledge and skills at a particular performance level lay the foundations for the next higher level. For more nuanced analysis, school results in a standards-based report are usually disaggregated by sub-groupings such as gender, school poverty quintile category, subject domains, cognitive levels, urban-rural sub-divisions and others. An illustrative example of a standards-based school report with some of the features discussed in this article is shown in Appendix 1.

Presentation of results

The results in the standard-based report are aggregated by PL and may be presented by specific sub-groups (e.g. boys and girls) as well as by subject-specific sub-domains and cognitive levels. Moreover, a summary of the results, for example by school or class, should also be provided to provide an overview of performance, while additional comparisons by district or province, where available, should also be reported in order to provide schools with a context within which to interpret results.

Use of reports by teachers for enhancing learning and teaching

Non-utilisation of information from NAS is one criticism that led to the development of the SRF and dissemination of standards-based reports. Standards-based reports are designed with the needs of the end-user in mind. Greater value from the report will be derived if teachers operate within professional learning communities than if they work as individuals. It is recommended that the report be reviewed and discussed by all school staff responsible for mathematics, including the Head of Department and school management team members.

For instance, in the Exemplar School Report in Appendix 1 (A1), key information is contained and presented in a user-friendly format in the generic PLs and the subject-specific PLDs. By referring to the PLs and the PLDs, the teacher can easily detect whether learners in the school or class are performing at the requisite grade level, identify learners with specific learning needs and plan targeted interventions. From the overall performance the report drills down to performance disaggregated by relevant sub-categories such as gender, subject domains, content cognitive demand levels, and others as necessity dictates. The report places at the disposal of a teacher powerful and specific information that they can use to decide on what to prioritise and how to differentiate interventions.

As reported in Figure 1-A1 of the school report, 25% of Grade 6 learners in City Primary School are functioning at the Partly Achieved level. According to Table 1-A1, 'Partly achieved' means that these learners demonstrate partial understanding of the knowledge and skills required to function at the Grade 6 level for mathematics (Table 2-A1), and are 'unlikely to succeed in the next grade without support'. Moreover, these learners require specific intervention to address their identified knowledge gaps, and additional support to progress to the required grade, that is, the 'Achieved' level. Table 3-A1 indicates that the performance of these learners is similar to other learners in the district and province, while Table 4-A1 indicates that there are more boys than girls in this PL category. The findings reported in Table 5-A1 indicate that these learners have performed relatively well for 'Number and operations' but need more assistance with 'Probability', 'Data handling', and 'Measurement'. In addition, results in Table 6-A1 also indicate that these learners struggled the most with questions that focused on 'Application' and 'Reasoning'.

From the reported findings on learner knowledge and skills, teachers can draw specific learner performance trends in their school and then proactively plan and prepare their lessons, and assessments to address identified learning deficiencies or improve on learner's strengths. In practice, three options exist for using the results of summative assessments in a formative manner to support learners to address their learning gaps or improve on their strengths. Planned interventions by teachers can be implemented: (1) at the beginning of a school year or of a school term, depending on whether results are from previous year or term, (2) just before a teacher introduces a new topic for which findings from the reports have shown pose particular challenges to learners or (3) using both options (1) and (2).

The discourse in this article highlights important seminal work on how standards-based reporting can influence effective utilisation of summative assessment results to enhance teaching and learning. While there is educationally sound motivation from research literature for the potential efficacy of standards-based reports from NAS, especially within the context of developed nations, empirical research on the aspects of the framework and its application in a setting like South Africa would provide necessary evidence to serve as a basis for taking the framework forward. In this regard, some of the specific areas for further research into the efficacy of the framework and how the use of standards-based reports could be sharpened need to be highlighted.

Areas for further research

For ease of implementation of the SRF and to ensure that assessment data obtained from summative assessments can be optimally used for formative purposes, we suggest four areas for further research. First, the SRF must be implemented in practice to: (1) determine its utility value across the different school types that characterise the education system in South Africa and (2) identify specific challenges and successes in its application by different role players, that is, teachers, school leaders, and education department officials at the district, provincial and national level.

Second, exploration is needed of the use of SRF to focus greater attention on, and implement specific interventions for, addressing the challenge of equity in classrooms, schools and districts. The use of standards-based reports can provide teachers, school leaders and education department officials with more useful and valid indicators that move beyond accountability measures that highlight specific disparities between learners at the different PLs. In this regard, additional research is required to explore the use of standards-based reports as indicators for determining the support that is needed to reduce the percentage of learners functioning at the lowest PLs, that is, 'Not achieved' and 'Partially achieved', and to increase the percentage of learners functioning at the 'Achieved' and 'Advanced' PLs. Thus, instead of monitoring change using mean percentage scores, which obscure which learners are progressing, the use of SRF focuses on those learners who need the most support, that is, the poor and marginalised.

Third, we recommend the development of a formatted Excel spreadsheet that teachers can populate with test results, and the software compiles a typical standards-based report that allows teachers to: (1) easily identify the specific questions on which learners perform poorly, (2) identify which learners need more assistance and (3) decide on possible next steps to follow for using summative data in a formative manner. For example, a report generated in the form of Figure 3 provides a shaded item map that shows whether learners got an item correct (unshaded) or incorrect (shaded). From Figure 3, teachers can immediately see which learners performed 'well' and in which test questions learners had the 'most difficulty'. In this case, learners had the most difficulty in questions 4, 5, 6, 9, 12, 13 and 15.

Similarly, the software should also generate a table that provides teachers with ideas for next steps, tailored to learners for each performance level. For example, the information generated in Table 3 provides intervention ideas for learners functioning at the 'Partly achieved' level.

Question	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total	Percent
Learner 1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	13	87
Learner 2	1	1	1	0	0	1	0	1	1	0	0	1	0	1	0	8	53
Learner 3	1	1	1	0	0	0	1	1	1	1	1	0	0	1	0	9	60
Learner 4	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1	12	80
Learner 5	1	1	1	0	0	1	0	0	0	1	0	0	1	0	0	6	40
Learner 6	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	11	73
Learner 7	1	0	1	0	1	0	1	0	1	0	1	0	0	0	0	6	40
Learner 8	0	1	1	0	0	0	1	1	0	1	1	0	1	1	0	8	53
Learner 9	0	1	1	0	1	0	1	1	0	1	1	0	0	1	0	9	60
Learner10	0	1	1	0	0	1	0	1	0	0	1	0	0	1	0	6	40

FIGURE 3: Example of Excel data entry sheet showing an item correct map.

TABLE 3: Example of an Excel sheet indicating next steps for learners functioning at the 'Partly achieved' level.

'Difficulty' questions for learners functioning at the Partly achieved level		
Question	Skills/competencies assessed	Workbook page numbers
4	Choose the correct multiple of 10	2–5, 30
5	Choose the correct prime number	15, 16
6	Choose the correct decimal.	94, 95, 170–173
9	Count forwards using decimals (to 3 decimal places)	67, 68, 89
12	Rewriting a number from words to symbols	7,8, 54, 102 & 103
13	Find the correct number value (in units)	6, 55–57, 104 & 105
15	Find the correct number value (in thousands)	

The information presented in Table 3 indicates the question number that most learners got incorrect, the specific competency or skill assessed in the question, and the pages in the DBE workbook or commercial textbook which teachers can use as revision exercises for addressing specific learning gaps.

Fourth, we are of the view that the introduction and use of performance standards in reporting assessment data paves the way for individualised testing. The question raised by Kingsbury, Freeman and Nesterak (2014) is appropriate in this regard: 'If we believe that education should meet each student's academic needs, why wouldn't we use assessments that adjust to their individual achievement levels?' (p. 1). With the assistance of enabling ICT in general and appropriate item response theory techniques in particular, the prospects of strategies such as Computerised Adaptive Testing (CAT) are growing and need to be exploited optimally (Weiss & Betz, 1973; Weiss & Kingsbury, 1984). In CAT an individual learner does not have to respond to all items in a test. Instead, the learner responds to an item that is considered to be either easy or of medium difficulty from a pool of items of a wide range of levels of difficulty. If the learner answers the item correctly, a more difficult item is administered until the probability of answering a more difficult item is shown to be at its lowest and testing is discontinued. The learner is then assumed to be functioning at the level of the most difficult item that they answered correctly.

Several features seem to make CAT a preferable option in terms of enhancing the utilisation of assessment data. Firstly, CAT has short turnaround times as the candidate does not have to answer a fixed number of test items and, therefore,

allows time for immediate intervention and remediation. Secondly, CAT is a relatively more cost-effective testing option than traditional testing as it uses a limited number of test items. Thirdly, computer applications like CAT allows for a wider range of items to be used, items that are more likely to approximate practice and real-life contexts, ensuring the validity of the testing process. We recommend that every report must be accompanied by both the generic PLs and the subject-specific PLs and PLDs. While information in the report may be presented in suitable graphical and tabulated formats, reading and interpreting the results against the PLs and PLDs enhances the significance and implications of what the results show. The ultimate goal is to ensure that teachers will find the detailed information in the PLDs more meaningful and thus be able to utilise this information for providing relevant support and feedback to improve chances of effective learning for all learners in their classrooms.

Conclusion

We reviewed literature that shows that, while the phenomenon of national assessments has been on the increase, there has not been convincing evidence that the results of these summative assessments optimally influence what happens in the classrooms in terms of teaching and learning. In the context of South Africa, we explored how prescriptions for recording and reporting the results of these assessments tend to fall short of the main purpose of assessment, which is to provide evidence-based feedback that will enable appropriate interventions to enhance teaching and learning. We pointed out how the reporting framework that is prescribed in the national curriculum may limit the extent to which NAS results could be used meaningfully as evidence to inform decision-making for planning and delivering appropriate interventions to enhance teaching and learning in the classrooms.

In particular, we highlighted that the reporting approach that averages assessment results in raw scores such as 'percentage correct responses' is deficient in information. It lacks necessary qualitative information on what learners know and can do as evidenced from the assessment. Consequently, users of the reports, particularly teachers, are not empowered to intervene in ways that will enhance teaching and learning and, potentially, improve performance.

We propose an alternative reporting framework that is intended to add value to reporting of results from NAS, and how this information can be used to provide quality feedback to inform evidence-based decision-making at different levels. The key design features of the proposed SRF are a clear rationale for providing relevant feedback to all users, the requirement for standard-setting exercises to enrich quantitative data with qualitative expert-provided information and easy-to-read guidelines on how standards-based reports should be compiled and used.

Using a mathematics school report as an exemplar, we demonstrated in a fair amount of detail how information from summative national assessments – presented and disaggregated by subject domains and various learner categories as guided by the SRF – can be used effectively for differentiated formative purposes to address identified learner needs at strategic stages during the school year. This kind and level of detail in reporting, derived from a carefully designed standards-based framework, goes beyond sheer traditional raw test scores, provides information-rich feedback and has potential to enhance teaching and learning in all schools.

Through the use of carefully defined hierarchical subject-specific performance levels and descriptors, the SRF leads to generating reports that adequately reflect the hierarchical nature of mathematics where knowledge of basic concepts lays foundations for understanding complex concepts. In the same vein, the hierarchy suggested in the framework reflects how learning in mathematics should be facilitated and carefully planned to provide 'scaffolding' that helps learners continually move to the next cognitive level.

We recognise that the alternative SRF needs to be piloted to obtain adequate empirical feedback about its efficacy. This is a limitation that we plan to address in a large-scale pilot of the SRF. The known challenges of capacity among teachers cannot be ignored; hence, we propose that the use of the SRF must be coupled with professional support, monitoring of the implementation progress, especially as it pertains to the needs of teachers and learners in low-resourced schools, and, where necessary, provision of appropriate ICTs to reduce workloads so that teachers can spend most of their time on effective utilisation of assessment results to enhance learning.

Acknowledgements

Competing interests

We declare that we have no financial or personal relationships that might have inappropriately influenced our writing of this article.

Authors' contributions

M.M. did most of the literature review and writing of the manuscript. A.K. led the conceptualisation of the article, made significant contributions to the writing and revising of the manuscript.

References

Bond, T.G., & Fox, C.M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed). London: Lawrence Erlbaum.

Braun, H., & Kanjee, A. (2007). Using assessment to improve education in developing nations. In J.E. Cohen, D.E. Bloom, & M.B. Malin (Eds.), *Educating all children: A global agenda* (pp. 303–353). Cambridge, MA: MIT Press.

Breiter, A., & Light, D. (2006). Data for school improvement: Factors for designing effective information systems to support decision-making in schools. *Journal of Educational Technology & Society, 9*(3), 206–217. Retrieved from https://www.j-ets.net/ETS/journals/9_3/18.pdf

Cizek, G.J. (1996). Standard-setting guidelines. *Educational Measurement: Issues and Practice, 15*(1), 13–21. https://doi.org/10.1111/j.1745-3992.1996.tb00802.x

Cizek, G.J., & Bunch, M.B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. London: Sage.

Coburn, C.E., Honig, M.I., & Stein, M.K. (2009). What's the evidence on districts' use of data? In J. Bransford, L. Gomez, D. Lam, & N. Vye (Eds.), *Research and practice: Towards a reconciliation* (pp. 67–87). Cambridge, MA: Harvard Educational Press.

Department of Basic Education. (2011). *Curriculum and assessment policy statement. Grades 7–9. Mathematics*. Pretoria: DBE. Retrieved from https://www.education.gov.za/LinkClick.aspx?fileticket=uCNqOwfGbmc%3d&tabid=573&portalid=0&mid=1629

Department of Basic Education. (2013). *Report on the annual national assessment of 2013: Grades 1 to 6 & 9*. Pretoria: DBE. Retrieved from https://www.education.gov.za/Portals/0/Documents/Reports/ANA%20Report%202013%20(2).pdf?ver=2013-12-05-123612-000

Department of Education. (2005). *Intermediate Phase systemic evaluation report*. Pretoria: DOE. Retrieved from https://www.gov.za/sites/default/files/DoE_GR_6_Intermediate_Phase_Systemic_Evaluation_Report_0.pdf

Dunne, T., Long, C., Craig, T., & Venter, E. (2012). Meeting the requirements of both classroom-based and systemic assessment of mathematics proficiency: The potential of Rasch measurement theory. *Pythagoras, 33*(3), a19. https://doi.org/10.4102/pythagoras.v33i3.19

Goldstein, H. (2001). *Using pupil performance data for judging schools and teachers: Scope and limitations*. London: University of London.

Goodman, D.P., & Hambleton, R.K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education, 17*(2), 145–220. https://doi.org/10.1207/s15324818ame1702_3

Govender, D.A. (2016). *The use of Annual National Assessments by provinces and districts to improve teaching and learning*. Unpublished doctoral dissertation, Tshwane University of Technology, Pretoria, South Africa. Retrieved from http://tutvital.tut.ac.za:8080/vital/access/services/Download/tut:2574/SOURCE1

Green, S. (2002, July). *Criterion referenced assessment as a guide to learning – the importance of progression and reliability*. Paper presented at the International Conference of the Association for the Study of Evaluation in Education in Southern Africa, Johannesburg, South Africa.

Griffin, P. (2009). Teachers' use of assessment data. In C. Wyatt-Smith & J.J. Cumming (Eds.), *Educational assessment in the 21st century: Connecting theory and practice* (pp. 183–208). Dordrecht: Springer.

Hambleton, R.K. (2000). *Setting performance standards on educational assessments and criteria for evaluating the process*. Lab Report 377. Amherst, MA: University of Massachusetts.

Hambleton, R.K., & Pitoniak, M.J. (2006). Setting performance standards. In R.L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 433–470). Westport, CT: Praeger.

Hambleton, R.K., & Slater, S.C. (1997). *Are NAEP executive summary reports understandable to policy makers and educators?* CSE Technical Report 430. Amherst, MA: University of Massachusetts.

Hoadley, U., & Muller, J. (2016). Visibility and differentiation: Systemic testing in a developing country context. *The Curriculum Journal, 27*(2), 272–290. https://doi.org/10.1080/09585176.2015.1129982

Kaftandjieva, F. (2010). *Methods for setting cuts cores in criterion-referenced achievement tests: A comparative analysis of six recent methods with an application to tests of reading in EFL*. Arnhem: European Association for Language Testing and Assessment. Retrieved from http://www.ealta.eu.org/documents/resources/FK_second_doctorate.pdf

Kanjee, A. (2007). Improving learner achievement in schools: Applications of national assessments in South Africa. In S. Buhlungu, J. Daniel, R. Southall, & J. Lutchman (Eds.), *State of the nation: South Africa 2007* (pp. 470–499). Pretoria: HSRC Press. Retrieved from http://www.hsrcpress.ac.za/product.php?productid=2183

Kanjee, A., & Moloi, Q. (2014). South African teachers' use of national assessment data. *South African Journal of Childhood Education, 4*(2), 90–113. https://doi.org/10.4102/sajce.v4i2.206

Kanjee, A., & Moloi, Q. (2016). A standards-based approach for reporting assessment results in South Africa. *Perspectives in Education, 34*(4), 29–51. https://doi.org/10.18820/2519593X/pie.v34i4.3

Kanjee, A., & Mthembu, A. (2015). Assessment literacy of foundation phase teachers: An exploratory study. *South African Journal of Childhood Education, 5*(1), a346. https://doi.org/10.4102/sajce.v5i1.354

Kanjee, A., & Sayed, Y. (2013). Assessment policy in post-apartheid South Africa: Challenges for improving education quality and learning. *Assessment in Education: Principles, Policy & Practice, 20*(4), 442–469. https://doi.org/10.1080/0969594X.2013.838541

Kellaghan, T., Greany V., & Murray, T.S. (2009). *Using the results of a national assessment of educational achievement*. Washington, DC: The World Bank.

Kingsbury, G.G., Freeman, E.H., & Nesterak, M. (2014). The potential of adaptive assessment. *Educational Leadership, 71*(6). Retrieved from http://www.ascd.org/publications/educational-leadership/mar14/vol71/num06/The-Potential-of-Adaptive-Assessment.aspx

Klinger, D.A., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy, 76*, 1–34. Retrieved from https://journalhosting.ucalgary.ca/index.php/cjeap/article/view/42757

Kolen, M.J., & Brennan, R.L. (1995). *Testing equating methods and practices*. New York, NY: Springer.

Moloi, M.Q. (2016). *A national framework for reporting the results of large-scale surveys in South Africa*. Unpublished doctoral dissertation, Tshwane University of Technology, Pretoria, South Africa. Retrieved from http://tutvital.tut.ac.za:8080/vital/access/services/Download/tut:2377/SOURCE1

Morgan, D.L., & Perie, M. (2005). *Setting cut scores for college placement*. New York: College Board. Retrieved from https://research.collegeboard.org/publications/content/2012/05/setting-cut-scores-college-placement

Nichols, S.L., & Berliner, D.C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.

Nkosi, B. (2015). SADTU: Boycott national assessments in schools. *Mail&Guardian*. Retrieved from https://mg.co.za/article/2015-09-02-sadtu-calls-for-boycott-of-national-assessments

Ravela, P. (2005). A formative approach to national assessments: The case of Uruguay. *Prospects, 35*(1), 21–43. https://doi.org/10.1007/s11125-005-6816-x

Rodriguez, M., Rubio, F., Landsdale, J., Vukmirovic, Z., Meckes, L., & Gysling, J. (2011, April). *Standard setting issues and practice in Latin America*. Paper presented at the Symposium on Standard Setting in an International Context: Issues and Practice, at the annual meeting of the NCME, New Orleans, LA. Retrieved from http://edmeasurement.net/research/Rodriguez%20Rubio%20Landsdale%20Meckes%202011.pdf

Saddler, D.R. (2010). Beyond feedback: Developing student ability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535–550. https://doi.org/10.1080/02602930903541015

Schiefelbein, E., & Schiefelbein, P. (2003). From screening to improving quality: The case of Latin America. *Assessment in Education, 10*(2), 141–154. https://doi.org/10.1080/0969594032000121252

Sloane, F., & Kelly, A. (2003). Issues in high-stakes testing programs. *Theory into Practice, 42*, 12–17. https://doi.org/10.1207/s15430421tip4201_3

Timperley, H. (2009). Using assessment data for improving teaching. In *Australian Council for Educational Research Conference Proceedings* (pp. 21–25). Perth: ACER. Retrieved from http://research.acer.edu.au/cgi/viewcontent.cgi?article=1036&context=research_conference

Tiratira, N.L. (2009). Cutoff scores: The basic Angoff method and the item response theory method. *The International Journal of Educational and Psychological Assessment, 1*(1), 27–35. Retrieved from https://docs.google.com/open?id=0ByxuG44OvRLPWUtIRlZNS2FuRms

Underwood, J.S., Zapata-Rivera, D., & Van Winkle, W. (2010). *An evidence-centred approach to using assessment data for policymakers*. Princeton, NJ: Educational Testing Service.

UNESCO. (2000). *Assessing learning achievement. Education for all: Status and trends 2000*. Paris: UNESCO.

Vygotsky, L.S. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Webb, N.I. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Washington, DC: Council of Chief State School Officers.

Weiss, D.J., & Betz, N.E. (1973). *Ability measurement: Conventional or adaptive?* Research Report 73–1. Minneapolis, MN: Department of Psychology, University of Minnesota. Retrieved from http://www.iacat.org/sites/default/files/biblio/we73-01.pdf

Weiss, D.J., & Kingsbury, G.G. (1984). Applications of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 31*, 361–375. https://doi.org/10.1111/j.1745-3984.1984.tb01040.x

Zieky, M., & Perie, M. (2006). *A primer on setting cut scores on tests of educational achievement*. Educational Testing Service. Retrieved from https://www.ets.org/Media/Research/pdf/Cut_Scores_Primer.pdf

Appendix 1
Example school report

Source: Adapted from Moloi, M.Q. (2016). *A national framework for reporting the results of large-scale surveys in South Africa*. Unpublished doctoral dissertation, Tshwane University of Technology, Pretoria, South Africa. Retrieved from http://tutvital.tut.ac.za:8080/vital/access/services/Download/tut:2377/SOURCE1

2018 National Assessment: Grade 6 Mathematics results

City Primary School   Southern province

The evidence presented in this report is intended to support school leaders and teachers identify learner strengths and weaknesses, and plan appropriate interventions for improvement.

How to use this report

It is recommended that the report be reviewed and discussed by all school staff responsible for mathematics, including the Head of Department and school management team members. The information provided in Table 1-A1 to Table 6-A1 and in Figure 1-A1 below should be carefully reviewed to:

(i)   Determine whether learners in the school are performing at the requisite grade level.
(ii) Identify specific learning needs of learners who at risk and those who are on track.
(iii) Plan targeted interventions for supporting all learners improve learning, based on their learning needs identified, especially those at the lower performance levels.

Performance level definitions and implications

Table 1-A1 below provides information on the four levels used to report the mathematics performance of Grade 6 learners, and it implications for progression and interventions for improving learning.

TABLE 1-A1: Performance levels for Grade 6.

| Performance level	Level definition	Progression implications	Intervention implications
Advanced	Performance at this level indicates that a learner demonstrates **comprehensive understanding** of the knowledge and skills required to function at this grade level	Learner has high likelihood of success in the next grade	Learner requires little or no academic intervention but need to be provided with more challenging tasks to maximise their full potential
Achieved	Performance at this level indicates that a learner demonstrates **sufficient understanding** of the knowledge and skills required to function at this grade level	Learner has a reasonable likelihood of success in the next grade	Learner may require some assistance with complex concepts to progress to the advance level
Partly achieved	Performance at this level indicates that a learner demonstrates **partial understanding** of the knowledge and skills required to function at this grade level	Learner unlikely to succeed in the next grade without support	Learner requires specific intervention to address knowledge gaps, and additional support to progress to the required grade (achieved) level
Not achieved	Performance at this level indicates that a learner demonstrates **very limited understanding** of the knowledge and skills required to function at this grade level	Learner unlikely to succeed in the next grade without significant support	Learner requires specific intervention to address knowledge gaps, with extensive and continued support to progress to the required Achieved level

Table 2-A1 lists the specific mathematics knowledge and skills that learners functioning at each performance level are expected to demonstrate.

TABLE 2-A1: Mathematics knowledge and skills at each performance level.

Not achieved	Partly achieved	Achieved	Advanced level
A learner at this level can recognise basic number systems, and can: • count forward only with whole numbers • count objects not exceeding 10 • add whole numbers up to 10 • draw simple pictures of objects • measure length of lines • name few SI units • measure area & perimeter of objects • draw simple bar graphs	A learner at this level can, in addition to skills and knowledge in the lower PL can: • count forward and backwards in decimals • recognise place value up to 9 digits • round off number up to 1000 • add and subtract up to 9 digits • do simple calculations using ordinary fractions and decimals • read digital and analogue time • measure using basic SI units • draw pictographs	A learner at this level can, in addition to skills and knowledge in the lower PLs: • count, recognise and do calculations using fractions & decimals • represent multiples, factors & prime numbers • find percentages of whole numbers • solve problems involving finances and measurement • compare rate and ratio • identify & describe numeric and geometric patterns • use and describe transformations • locate and describe movement on a grid • differentiate between sample & population • draw & interpreting bar and pictographs	A learner at this level can, in addition to skills and knowledge in the lower PLs: • critically read, interpret and analyse with awareness of sources of error and manipulation to draw conclusions and make predictions • list possible outcomes for simple experiments including tossing a coin, rolling a die and spinning a spinner • distinguish between volume, surface area & dimensions of rectangular prisms. • solve problems involving different time zones • estimate, record, compare & convert between SI units (including mass, temperature, distance and capacity) • organise & record data • calculate the median & mode of data • list possible outcomes & predict 'likelihood' of events

Results for school

Figure 1-A1 below presents the overall percentage of Grade 6 learners functioning at each of the performance levels in mathematics for your school. It shows learners who are at risk (i.e. at the Not achieved and Partly achieved levels) and those who are on track (Achieved and Advanced levels).
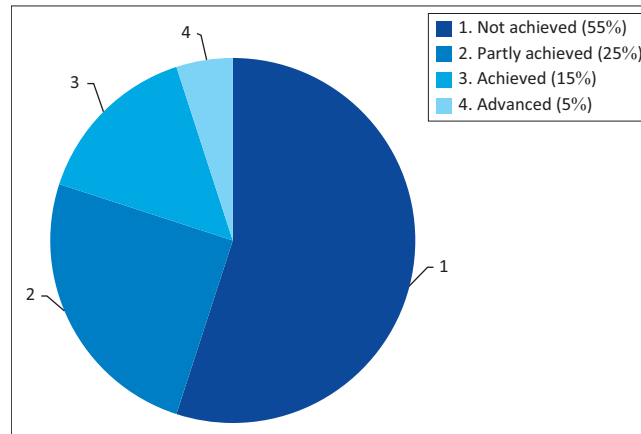
FIGURE 1-A1: Distribution of learners across mathematics performance levels.

Table 3-A1 compares the overall percentage of Grade 6 learners in this school functioning at each of the performance levels in mathematics against the district and province results.

TABLE 3-A1: School performance (%) in mathematics by district and province.

Variables	Not achieved	Partly achieved	Achieved	Advanced
School	55	25	15	5
District	39	21	22	18
Province	47	23	17	13

Table 4-A1 provides information on the percentage of boys and girls functioning at the different performance level within the school, the district and the province.

TABLE 4-A1: Mean score (%) by district, province and gender.

Variables	Not achieved		Partly achieved		Achieved		Advanced	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
School	30	20	25	20	30	45	15	15
District	20	10	35	25	30	45	15	20
Province	35	25	25	20	20	20	20	35

Table 5-A1 lists the mean scores of learners functioning at each of the performance levels in the five mathematics content domains. This information indicates the knowledge levels of learners in each performance level for the different content domain areas.

TABLE 5-A1: Mean score (%) by content domain and performance level.

Content domain	Not achieved	Partly achieved	Achieved	Advanced
Number & Operations	47	64	71	85
Functions & Algebra	35	46	60	70
Measurement	19	33	49	89
Data handling	14	34	68	67
Probability	10	25	60	63

Table 6-A1 presents mean scores of learners functioning at various cognitive levels. This information indicates the extent to which learners are demonstrating complex cognitive capabilities in mathematics.

TABLE 6-A1: Mean score (%) by cognitive level and performance level.

Cognitive Level	Not achieved	Partly achieved	Achieved	Advanced
Knowledge	20	57	74	95
Application	21	44	78	89
Reasoning	6	34	87	97