A NATIONAL FRAMEWORK FOR REPORTING THE RESULTS OF LARGE-SCALE ASSESSMENT SURVEYS IN SOUTH AFRICA

By

MESHACK QETELO MOLOI

for a

Doctoral Degree in Education

in the

Department of Educational Studies

FACULTY OF HUMANITIES

TSHWANE UNIVERSITY OF TECHNOLOGY

Supervisor: Prof Anil Kanjee

December 2015
COPYRIGHT
This study proposes a National Reporting Framework (NRF) to enhance the recording, reporting, dissemination and utilisation of results of large-scale assessment surveys (LSAS) such as TIMSS, PIRLS, SACMEQ and others, and uses Annual National Assessments (ANAs) as a specific example, in South Africa. The study describes the process followed in developing a standards-based NRF, discusses the rationale behind the proposed use of performance standards in reporting learner performance and compares three different methods of setting performance standards to identify the most suitable method for the South African context.

South Africa administers ANAs to monitor the levels of performance in basic education. While the main purpose of ANAs is to provide data that must inform effective intervention strategies to improve the quality of teaching and learning in schools, after more than five years, the impact of ANA on teaching and learning effectiveness remains unconfirmed.

From the reviewed literature, some of the factors that were identified as possible impediments to effective use of assessment results were the reporting of results in information-deficient raw score percentages, inadequate capacity for data use in the education sector generally and among teachers in particular, and a pervasive absence of a culture of data use. At the same time there was evidence, both in South Africa and from international studies, which showed the superiority of performance standards over classical statistical formats such as percentage of correct answers in a test, in the reporting of assessment results.

The study posits a ‘theory of action’ which holds that providing a standards-based NRF within a criterion-referenced paradigm, presenting assessment results in performance standards and formats that are user-friendly, meaningful and information-rich, accompanied with relevant capacity building in a sustained culture of data use, will lead to effective data use at all levels and result in improved performance in schools.

A comparison of three methods of standard-setting, the Angoff, Objective Standard Setting (OSS) and the Objective Borderline Method (OBM), was conducted in order to recommend a method that would be most suitable for the South African context. Key criteria that were used to compare the different standard setting methods were: a) a facility to generate consistent cutscores in different contexts; b) the relative impact of cutscores on categorising learners according to
performance; c) the cost; and, d) the amount of subject content linked to identified performance standards.

An evaluation of the three different SS methods in terms of the selected criteria showed that the OSS method: a) cost less and had a shorter turnaround time than the Angoff method; b) involves technical and conceptual processes that are easier to be understood and implemented by standard-setting panellists, resulting in reduced variability of cutscores compared to the Angoff method; and, c) generates cutscores and performance standards that are the richest and the most appropriate in terms of curriculum and subject content. The OBM was found to be a purely statistical technique that ensures precision of cutscores but produces performance standards that lack curriculum or subject content.

A survey conducted on education officials in South Africa indicated marked preference for a standards-based reporting design over reporting learner performance in raw score percentages. Some of the motivating statements for the preference shown by officials were that standards-based reports communicated more information than merely raw score percentages. Standards-based reports were regarded as being more useful for teachers to identify and address specific needs of learners.

The NRF developed in this study outlines the rationale for use of performance standards in reporting assessment results, the processes that must be followed after every national assessment and provides templates of standards-based reports that must be generated for national, provincial, district, school and individual learner levels. It is recommended that the OSS method be used for setting performance standards because of the advantages that set it apart from the other methods. There are recommendations on processes that should be followed in terms of interpreting assessment results, setting improvement targets and mounting appropriate interventions for improving teaching and learning as well as overall learner performance.

One research implication from this study is the need to replicate the standard setting processes using larger panels of standard setters. Published research indicates that larger panels produce more reliable cutscores than smaller panels. There is also a need to run a pilot study of the implementation of the NRF. A limitation of this study was that the target users of the NRF, who are mainly the officials who manage curriculum and assessment implementation, could not be
interviewed for detailed information on what they perceive as possible challenges in implementing the NRF. Their inputs would serve to improve the utility of the NRF.
EXAMINERS’ SIGNATURES
ACKNOWLEDGEMENTS

This study could not have been completed without the support that I enjoyed from my DBE colleagues, family, and my supervisor. In particular Deva and Sabie gave me wonderful moral support and brilliant ideas that helped this work crystallize. May their kind increase.

Professor Kanjee was simply amazing in the way he patiently kept me focused on this work. I thank him immensely.

While I take final responsibility for the quality of this work, Cynthia Ayeza Mutabaazi did sterling work in going through each line and page of this document with a ‘fine comb’! Thank you, Cynthia.

I had a very strong and loving support from my wife and children. May God bless them.
# TABLE OF CONTENTS

1. **INTRODUCTION** ................................................................. 1
   1.1 Background ........................................................................ 1
   1.2 Annual National Assessment results ..................................... 4
   1.3 The potential of LSAS ............................................................. 5
   1.4 Guidelines for LSAS data use ................................................. 7
   1.5 Use of performance standards ................................................ 8
   1.6 The study thesis ................................................................. 10
   1.7 Statement of the research problem ......................................... 11
   1.8 Purpose of the study ......................................................... 11
   1.9 Objectives of the study ....................................................... 12
   1.10 Research questions ............................................................ 12
   1.11 Significance of the study .................................................... 13
   1.12 Definition of terms ............................................................... 13
   1.13 Limitations of the study ...................................................... 14
   1.14 Organisation of the study .................................................... 15

2. **LITERATURE REVIEW** ......................................................... 17
   2.1 Conceptualisation and use of LSAS in education ...................... 17
   2.2 Factors affecting utilisation of LSAS ...................................... 23
   2.3 Case studies of effective utilisation of LSAS ............................. 28
   2.4 The value and challenges linked to feedback from LSAS ............ 34
   2.5 The quest for standards ....................................................... 35
      2.5.1 Content standards ......................................................... 35
      2.5.2 Performance standards .................................................. 36
      2.5.3 Setting and reporting performance standards ...................... 37
      2.5.4 Approaches and methodologies to standard setting ............. 41
   2.6 The traditional Angoff method .............................................. 42
      2.6.1 Sequence of steps in the Angoff method ......................... 42
      2.6.2 Rasch-based Angoff SS methods .................................... 43
   2.7 The OSS method ............................................................... 46
   2.8 The OBM ........................................................................... 49
   2.9 Historical and current practices of LSAS in South Africa ............ 53
      2.10 Systemic evaluations of 1998 - 2003 ................................... 53
      2.11 Systemic evaluations in nodal zones in 2004 - 2007 ................ 54
      2.12 Foundations for Learning Campaign and ANA in 2008 - 2009 ... 54
2.13 ANA from 2010 onwards........................................................................................................54
2.14 Use of various reporting formats ..........................................................................................56
2.15 Frameworks for data-driven decision-making .................................................................65
2.16 Key principles ..........................................................................................................................71
2.17 Design and quality of tests for SS.........................................................................................72
  2.17.1 Validity of tests ....................................................................................................................73
  2.17.2 Item and person fit ..............................................................................................................74
  2.17.3 Test targeting .....................................................................................................................75
2.18 Summary ..................................................................................................................................76
3.  RESEARCH DESIGN AND METHODOLOGY ............................................................................78
  3.1 Introduction .............................................................................................................................78
  3.2 Research questions ................................................................................................................78
  3.3 Research design .....................................................................................................................79
  3.4 Types and sources of data .......................................................................................................81
    3.4.1 Learner data ......................................................................................................................82
    3.4.2 Test item rating scores ......................................................................................................82
    3.4.3 SS-process observation scores .........................................................................................82
    3.4.4 Guidelines on use of assessment data ..............................................................................82
  3.5 Instrumentation .......................................................................................................................83
    3.5.1 Mathematics and language tests ......................................................................................84
      3.5.1.1 Validity and reliability of the tests ..............................................................................84
      3.5.1.2 Test targeting .............................................................................................................85
    3.5.2 Instruments used in SS .....................................................................................................87
      3.5.2.1 Instruments used in Angoff SS ...................................................................................87
      3.5.2.2 Instruments used in the OSS .....................................................................................88
      3.5.2.3 SS evaluation instrument .........................................................................................88
      3.5.2.4 Survey questionnaire ...............................................................................................89
  3.6 SS processes and techniques ..................................................................................................89
    3.6.1 Selection of SS panellists .................................................................................................89
    3.6.2 Development of PLs and PLDs .......................................................................................90
  3.7 Data collection .........................................................................................................................94
    3.7.1 Angoff SS data collection .................................................................................................94
      3.7.1.1 Training of panellists .................................................................................................96
      3.7.1.2 Rating of test items ....................................................................................................98
      3.7.1.3 Capturing of data .......................................................................................................99
3.7.1.4 Calculation of final cutscores ........................................................................................................ 99
3.7.2 OSS data collection ............................................................................................................................. 99
  3.7.2.1 Training of panellists ..................................................................................................................... 100
  3.7.2.2 Rating of test items ....................................................................................................................... 101
  3.7.2.3 Capturing of data .......................................................................................................................... 102
  3.7.2.4 Calculation of cutscores ............................................................................................................... 103
3.7.3 OBM SS method .................................................................................................................................. 104
3.8 Data simulations .......................................................................................................................................... 105
3.9 Evaluation of the various SS methods ........................................................................................................ 106
3.10 Necessary conditions for data use .......................................................................................................... 107
3.11 Survey on reporting results using performance standards ...................................................................... 108
3.12 Statistical tests of significance ............................................................................................................... 109
3.13 Summary .................................................................................................................................................. 110
4. RESEARCH FINDINGS ............................................................................................................................. 112
  4.1 Introduction ............................................................................................................................................... 112
  4.2 Findings on Research Question 1 ........................................................................................................... 112
    4.2.1 Mathematics cutscores .................................................................................................................... 113
    4.2.2 Language cutscores ........................................................................................................................ 115
    4.2.3 Cutscores from real and simulated data .......................................................................................... 118
      4.2.3.1 Mathematics cutscores from real and simulated data .............................................................. 119
      4.2.3.2 Language cutscores from real and simulated data .................................................................. 121
    4.2.4 Impact of cutscores ........................................................................................................................ 123
    4.2.5 Summary of key findings on cutscores ............................................................................................ 126
      4.2.5.1 Mathematics cutscores ............................................................................................................. 126
      4.2.5.2 Language cutscores ................................................................................................................. 126
    4.2.6 Reliability test for the SS methods .................................................................................................... 126
      4.2.6.1 Reliability of SS method for mathematics ............................................................................. 127
      4.2.6.2 Reliability of SS method for FAL ......................................................................................... 127
    4.2.7 Impact of cutscores ........................................................................................................................ 127
    4.2.8 Evaluation of the SS methods by panellists .................................................................................... 127
    4.2.9 Summary of findings from Research Question 1 ............................................................................ 131
  4.3 Findings on Research Question 2 ........................................................................................................... 132
    4.3.1 Conditions for data use .................................................................................................................. 132
    4.3.2 Summary of findings from Research Question 2 ........................................................................... 135
  4.4 Findings on Research Question 3 ........................................................................................................... 135
7.4.2 Guidelines and support to provinces and districts .................................. 256
7.4.3 Views of education officials on standards-based reporting ....................... 256
7.5 Conclusion ........................................................................................................... 257
7.6 Recommendations ............................................................................................. 258

REFERENCES ............................................................................................................. 260

Appendix A: Extract from a FAL rater sheet used by panellists in Angoff SS .......... 276
Appendix B: Grade 6 mathematics rating sheet and summary (Angoff SS) ............. 277
Appendix C: Extract from a FAL rating form used for the OSS ............................ 279
Appendix D: Standard-setting Evaluation Instrument ............................................. 280
Appendix E: Questionnaire and templates used for survey ................................... 281
Appendix F: Grade 3 PLs and PLDs ......................................................................... 284
Appendix G: Grade 6 PLs and PLDs ......................................................................... 285
Appendix H: Grade 9 PLs and PLDs ......................................................................... 287
Appendix I: Grade 6 mathematics item difficulties from OSS method ............... 289
Appendix J: Relationship between “range of error” and cutscores in OBM .......... 293
Appendix K: MS-DOS commands executed in data simulation ............................. 294

List of figures and tables

Figures

Figure 1: Distribution of ability scores in a test ....................................................... 50
Figure 2: Example of calculating a cut score in OBM ............................................. 52
Figure 3: Distribution of grade 6 learners across performance levels .................... 59
Figure 4: SACMEQ levels for grade 6 mathematics .............................................. 60
Figure 5: Level 2-emergent numeracy item ............................................................. 61
Figure 6: Level 5-competent numerate ................................................................. 62
Figure 7: Example of a report based Scale Scores .................................................. 63
Figure 8: The SACMEQ Framework for Educational Evaluation .......................... 67
Figure 9: Conceptual Model of data-driven decision-making ................................. 68
Figure 10: Grade 6 mathematics Item Map ............................................................. 86
Figure 11: Grade 6 FAL Item Map ......................................................................... 87
Figure 12: Mathematics cutscores calculated from three SS methods ................... 114
Figure 13: Language scores calculated from three SS Methods ............................ 117
Figure 14: Mathematics cutscores calculated from real and simulated data ........... 119
Figure 15: Language cutscores calculated from real and simulated data ............... 121
Figure 16: Distribution of mathematics learners across the performance levels ....... 123
Figure 17: Distribution of language learners across the performance levels .......... 124
Tables

Table 1: Illustration of steps and key concepts in calculating OSS cutscores ........................................ 48
Table 2: Levels and score ranges in the DoE reporting framework ..................................................... 56
Table 3: DBE performance levels ........................................................................................................ 57
Table 4: ANA 2013 national results .................................................................................................... 58
Table 5: Profile of participants in setting PLs and PLDs ................................................................. 91
Table 6: Grade 6 mathematics PLs and PLDs .................................................................................... 92
Table 7: Grade 6 FAL PLs and PLDs ................................................................................................ 92
Table 8: Participants in the Angoff SS process ................................................................................... 95
Table 9: SSEI, the instrument used to evaluate SS methods ............................................................ 107
Table 10: Distribution and responsibilities of respondents to questionnaire ................................. 109
Table 11: Mathematics cutscores that violate the PL hierarchy ..................................................... 113
Table 12: Summary of significant differences among math cutscores by SS method ................. 115
Table 13: FAL cutscores that violate the PL hierarchy ................................................................. 116
Table 14: Summary of significant differences among FAL cutscores by SS method ................ 118
Table 15: The SSEI results on the Angoff, OSS and OBM SS methods .......................................... 128
Table 16: Summary of necessary conditions for data use ............................................................. 133
Table 17: Summary of national and provincial views on T1 and T2 (n=39) ........................................ 136

ACRONYMS

CAPS  Curriculum and Assessment Policy Statement
CITO  (Dutch) National Institute for Educational Measurement
GET   General Education and Training
LSAS  Large-scale assessment surveys
NDP   National Development Plan
NRF   National Reporting Framework
OBM   Objective Method
OSS   Objective Standard Setting
UNESCO United Nations Education and Science Organisation
UNICEF United Nations Initiative for Child Education and Fellowship
US    United States of America
1. INTRODUCTION

1.1 Background

A common finding of all studies on the quality of education in South Africa, as measured through learner achievement, has been that achievement is not only low, but the country is also performing far lower than neighbouring countries that spend significantly less on their education systems (Van der Berg, 2008; Fleisch, 2008). This has raised questions as to what the persistent low national performance could be ascribed to, and more importantly, what the government and the Department of Basic Education (DBE) should do to turn the situation around.

Investment in and prioritisation of education worldwide have been premised on, and supported by, research findings that indicate that levels of education, represented by the number of years of formal schooling, have a positive rate-of-return expressed in workers’ earnings (Mincer, 1974). However, research also shows that the impact of education on economic productivity of workers is enhanced significantly when the quality, not only the quantity as expressed in years of schooling, is included in the economic modelling of the relationships (Hanushek & Kimko, 2000). The realisation that access to education on its own is inadequate for both social and economic growth prompted the international community to shift the focus from access to education (UNESCO, 2014a) to requiring that all learners achieve measurable educational outcomes of a high quality by 2015 (UNESCO, 2014b), particularly in the key foundational skills of literacy and numeracy.

As a consequence of the focus on measurable learning outcomes, there has been a phenomenal increase in the number of countries that either conduct large-scale assessments of their learner populations or collaborate with others in international studies to benchmark the quality of their educational outcomes (Greaney & Kellaghan, 2008; Kanjee & Sayed, 2013). A definition of large-scale assessment surveys (LSAS), also often referred to as national assessments of learner achievement, is given by Kanjee (2007:13) as:

“The process of obtaining relevant information from an education system to monitor and evaluate the performance of learners and other significant role-players as well as...
the functioning of relevant structures and programs within the system for the purpose of improving learning."

Greaney and Kellaghan (2008:7) advance a similar definition and add that these assessment surveys involve testing representative samples or populations of learners of a particular age or grade to obtain an estimate measure of the level of achievement in the education system as a whole. There have been instances where, in addition to completing questionnaires on contextual factors that might impact on teaching and learning in schools, teachers were also requested to write tests to assess their levels of content knowledge in the subjects that they teach (Moloi & Chetty, 2010).

Greaney and Kellaghan (2008:15) distinguish between national examinations which are high stakes tests, administered for purposes of certifying and selecting individual learners, and LSAS which are low stakes tests whose purpose is to provide feedback on the quality of education, mainly to policy makers as well as to teachers. It is important to note that LSAS as contemplated and defined in this study, exclude assessments such as the Dutch CITO system (Maslowski, 2001). In the Dutch CITO system schools and individual learners participate by choice, not through a process of probabilistic selection to represent others of the same kind. CITO tend to be high stakes examinations whose results can determine the institutions where individual learners may or may not be admitted (Maslowski, 2001:150,169). ANAs and other LSAS as defined in this study do not carry such consequences. A single country can conduct its own LSAS but often groups of countries collaborate to conduct these studies together.

In South Africa, Kanjee (2007) chronicled a series of LSAS that were introduced following the demise of the apartheid regime. Specifically, the government of South Africa promulgated the national Assessment Policy (1998) which, among others, mandated the implementation of periodic systemic assessments of learner achievement, also known as ‘systemic evaluations’, to track progress in learner achievement over time and report on achievement of the then transformational goals of access, redress, equity and quality.

The first LSAS by the post-apartheid Department of Education (DoE) to monitor the quality of the education system was implemented in 1999 (Kanjee, 2007). With the support of the UNESCO-UNICEF collaboration, the Department of Education implemented the Monitoring Learning Achievement (MLA) survey in 1999 (Strauss, 1999). In 2001, the MLA was
followed by the first Systemic Evaluation (SE) study (Kanjee, 2007). Since then, LSAS has been evolving alongside a growing culture of a monitoring and evaluation (M&E) system that emphasises the use of quantitative indicators of educational effectiveness. The assessments started as periodic three-year-cycle SE studies which were intended to be administered on representative samples of schools and learners at the levels of grades 3, 6 and 9 (DoE, 2003a; DoE, 2005) but were later expanded to include all learners in all schools in grades 1-6 and grade 9 and have been conducted on an annual basis since 2011 (DBE, 2011b).

Furthermore, since 1994 South Africa has participated in both regional and international benchmarking studies, like the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) (Moloi & Strauss, 2005), the Trends in International Mathematics and Science Studies (TIMSS) (Reddy, 2006), and the Progress in International Reading and Literacy Studies (PIRLS) (Howie, Venter, Van Staden, Zimmerman, Long & Scherman, 2007). In these studies, learner achievement is used as a proxy for measuring the quality of the education system. Popular subjects in which learners are tested are language and mathematics although, as is the case with TIMSS, subjects such as science are sometimes included (Greaney & Kellaghan, 2008:8).

The results of both national and international assessment surveys continued to show that the performance levels of South African children were unacceptably low and, in the grades that participated in international studies, were lower than performance in neighbouring countries that invest significantly lower levels of resources in their education system (DBE, 2011a:74).

Responding to broader issues of poor quality in the education system, as manifested in the persistent learner under-achievement, the DBE introduced and led the development of the Education Sector Plan, “Action Plan 2014: Towards Schooling 2025”, which was a strategy to focus on improving the quality of basic education for all children of South Africa (DBE, 2011a). The Action Plan required all sections of the DBE to set measurable goals for short-, medium- and long-terms and specify how each goal would, directly or indirectly, help address the common problems for education – unacceptable levels and quality of learner achievement.

Through the Sector Plan, the DBE identified laudable intervention strategies to improve the quality of basic education and thus set measurable goals that were planned to be achieved by
2014 (DBE, 2011a). The Action Plan prioritises long-term planning and specifies 27 measurable goals that the Education Sector must achieve by 2025. Of relevance to this study is that the top three goals in the Action Plan are about increasing the numbers of learners who master “minimum literacy (language) and numeracy (mathematics) competencies” in grades 3, 6 and 9, respectively (DBE, 2011a:9). The Action Plan notes that performance of learners will be measured through the results of Annual National Assessments (ANA), which have been identified as the main indicator of successful achievement of most of the goals in the Sector Plan towards establishing “… a schooling system that offers quality education to all young South Africans …” (DBE, 2011a:17).

1.2 Annual National Assessment results

The ANAs were introduced following the Presidential injunction in the State of the Nation Address in which the President of South Africa announced that the country was going to monitor the quality of basic education to ensure that at least 60% of learners in grades 3, 6 and 9 achieved “acceptable levels” of performance in the key foundational skills of language and mathematics by 2014 (South African Parliament, 2010). The assessments were subsequently included as the primary mechanism for measuring key learning targets specified in the Education Sector Plan (2011a).

Between 2011 and 2013, the DBE extended participation in ANAs to include all learners in grades 1 to 9 who are in public schools (DBE, 2013a). Learners are tested in language and mathematics. The results are reported in raw scores (percent correct), as well as in percentage of learners who achieve a minimum of 50% of correct responses in each test, and are disaggregated by province, language of learning and teaching, gender and poverty quintile of the school. The results of ANA 2013, in average percent marks, ranged from 14% in grade 9 to 60% in grade 1 mathematics, 43% in grade 9 to 60% in grade 1 Home Language (HL) and 33% in grade 9 to 39% in grade 4 First Additional Language (FAL). These results indicated that overall performance in the system was generally low, with extremely poor performance noted in the higher grades.

According to Kanjee (2007), there have been observable improvements in South Africa in the methodological and technical aspects of the studies which include recognition of a “systems approach” to monitoring performance whereby issues of access, quality, equity and efficiency
are viewed not in isolation but as interrelated and interdependent design features of educational quality. Other improvements have been in methods of instrument administration, data management and data analysis, thus increasingly lending more confidence and credibility to the study findings. Kanjee’s (2007) reflection on the evolution of LSAS in South Africa showed that most of the assessments produced information that was mainly intended for use by policy makers and very few were designed to inform teachers and teaching for effective learning. Where there were innovative initiatives to package information for use by teachers, as was the case in the grade 6 ‘systemic evaluation’ (DoE, 2005), the innovations were neither replicated nor sustained, arguably because they were not required by policy nor specified in any guidelines or frameworks regarding the reporting of scores from LSAS.

Moreover, while there have been improvements in the administration of ANA over the years (DBE, 2013a), what has been demonstrably lacking in South Africa has been effective utilisation of the findings from LSAS to inform decision-making at various levels of the education system. Specifically, the key challenge is to enhance the use of results from LSAS to inform decision-making that will lead to appropriate interventions to improve the quality of education (Kanjee & Sayed, 2013).

This study maintains that interventions that can lead to improvement in educational quality must affect what happens in the classrooms and influence the support that is given to schools. This proposition has been corroborated by the work of Kellaghan, Greaney & Murray (2009) on the under- and non-utilisation of the results of LSAS. Kellaghan, Greaney & Murray (2009) observed that irrespective of the economic development stage of the countries that either conduct or participate in assessment studies, there was concern that feedback from these studies either does not reach the intended audiences, or does not reach them in forms and formats that are easily understandable, and often does not translate into expected change in practice.

1.3 The potential of LSAS

There has been a growing body of research on the impact of LSAS in systems that have implemented these assessments. Senghor (2014) reported on how in The Gambia they used assessments effectively to turn their system around. Part of their strategy was to present
assessment data on school scorecards that showed, among others, performance of each school in relation to schools in the same district and region and training teachers on how to use the data to improve teaching in their classrooms. Schools were provided with guidelines on how to use the data to address identified individual learner needs and were given standardised tools for recording and reporting the performance of their learners in turn. According to Senghor (2014), this was the first time that assessment results were also presented to parents and school communities in reader-friendly formats, including use of visuals. The Gambia then started a new practice of holding schools accountable for improving learner performance.

In a Latin American context, Ravela (2005) describes how Uruguay used the results from their census-based national assessments, similar to the South African ANA, to inform training of teachers which in turn led to improved teaching strategies and significant improvements in national performance. According to Ravela (2005:1) their strategy was premised on an agreed understanding that they were only going to conduct national assessments on condition that this process would directly impact on teaching practice in the classrooms. Uruguay reported national assessment results in five performance levels (PLs) that concisely summarised what a learner at each level was expected to be able to do, a fairly detailed description of competences that characterise each PL, and the distribution of learners across the PLs. They also provided tabulated comparative data where schools could see how their performance compared with schools in the same area as well as all schools in the country. Ravela (2005:36) summarised the philosophy of Uruguay on the purpose and effective use of LSAS thus:

“Quantitative results of assessment are indeed useful in drawing a picture of the education system, but they are not enough to promote an improvement in teaching practices. There is a need for analytical and conceptual work in order to transform numerical results into information and reflection that is useful to teachers”.

Other reported examples of effective utilisation of LSAS results that led to varying degrees of improvements in performance include Chile (Schiefelbein & Schiefelbein, 2003), Namibia and Kenya (Nzomo & Makuwa, 2006). The Chilean study by Schiefelbein & Schiefelbein
targeted schools and teachers; hence, according to Klinger et al. (2008), teachers were given salary incentives in recognition of increases in the test scores of their learners in the LSAS. The Namibian and Kenyan studies (Nzomo & Makuwa, 2006) were part of the SACMEQ studies (Moloi & Strauss, 2005) and had their primary target audiences as policy-makers at national and provincial levels. However, the two education systems took advantage of the detailed level of analysis and reporting on what learners were able or not able to do in the assessments and used their national reports to provide feedback to their regional curriculum and subject advisors. What was common among countries that reported improvements in their performance was reporting data in creative formats that were user-friendly, providing comparative data in the reports, providing guidelines, and appropriate tools and building capacity for use of the data.

1.4 Guidelines for LSAS data use

The guiding principles for this study, derived from literature review on factors that promote effective use of data, suggest that minimum conditions and basic guidelines must be provided for a large-scale uptake of data-driven decision-making in general and use of data to influence teaching and learning particularly in schools (Timperly, 2009:1; Coburn, Honig & Stein, 2009:14). However, Timperly (2009:1) cautioned against assumptions that making assessment data available will, of its own, lead to the use of this information by teachers. Her study on promoting use of assessment data in teaching in 300 schools in New Zealand showed that teachers require a certain minimum level of knowledge about the data itself, how the data is packaged, what the different statistical pieces of information mean, and how the information can be used.

Some of the factors that were found to hinder effective use of assessment data by teachers include data that is not relevant to curriculum needs of teachers, is difficult to access, presented in statistical and complicated formats, and is not mediated with the support of school leadership (Timperly, 2009:21; Taut, Cortés, Sebastian & Preiss, 2009:129; Senghor, 2014). In their study regarding reasons that led to under- or non-use of assessment data by teachers in Chile, Taut et al. (2009:137) recommended that assessment information in Chile be reported in performance standards with three hierarchical levels of performance.
One of the findings made by Taut et al (2009:137) was that teachers did not have sufficient capacity to interpret the assessment results and were thus less likely to use this information to improve their teaching approaches. Similar findings were made by Kanjee and Moloi (2014) in South Africa. These authors argued for “… development of performance descriptors and clear-cut scores to categorise learners into specific performance levels, so that teachers have detailed information about what learners know and can do” (Kanjee & Moloi, 2014:109).

This study was conducted to develop a framework comprising clear guidelines regarding specific and meaningful ways of reporting assessment data from LSAS. The guidelines must facilitate meaningful interpretation of results to assist users who may not have the expected skills of data interpretation. Within the framework there must be provision of support and relevant tools to enhance the use of assessment results in decision-making processes and in classrooms for improving learning and teaching.

1.5 Use of performance standards

A common aspect of those studies that provide clear and easy-to-understand LSAS reports pertains to the use of standards (Hambleton & Rogers, 1991; Goodman & Hambleton, 2004; Ravela, 2005). In educational circles, a distinction is made between “content standards” and “performance standards” (Cizek, 1996; Rodriguez, Rubio, Landsdale, Vukmirovic, Meckes & Gysling, 2011). Rodriguez et al (2011:18) define “content standards” as “… what learners need to learn…”. In the context of South Africa, “content standards” are spelt out in the Curriculum and Assessment Policy Statements (CAPS) by grade and by subject (DBE, 2011c). “Content standards” specify the nature and scope of content knowledge that a learner must acquire in a given grade.

Hambleton (2000:2) 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”.
Cizek and Bunch (2007:13) define the process of standard setting (SS) as establishing one or more cut score(s) 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.

With the increase in the use of performance standards, there has also been a corresponding increase in the variety of standard setting methods (Näström, & Nyström, 2008:10). Näström and Nyström (2008:1-12) compared the Angoff and Bookmark methods of standard setting and concluded that the two methods produced comparable cutscores. In South Africa, Kanjee and Moloi (2013) conducted a standard setting study using the Angoff method (Angoff, 1971) and compared standard-based cutscores with raw score percentage ranges used by the DBE in categorising learners according to performance in ANA tests. The DBE raw score-based reporting framework is prescribed in the CAPS document (DBE, 2011c). The findings showed that the two sets of cutscores identified different categorisations of learner performance. However, in the absence of another standard setting method to compare with, it was not possible to conclude on whether there were real differences among the identified categories of learners or whether the difference was due to the specific standard setting method used.

The Objective Standard Setting (OSS) method is a fairly recent procedure that builds largely on work done by Wright and Grosse (1993) and has been researched widely by Stone (1995, 2001 & 2008) and also compared with other methods by Stone, Kolskey and Sondergeld (2011). In this method specialists study the administered test and, using both prescribed curriculum documents and what they know from practice, categorise each test item into whether the knowledge or skill required to respond correctly is “essential” or non-essential at each of the defined performance levels (Stone, 1995:447). This qualitative categorisation of knowledge and skills is then followed by a confirmatory rigorous measurement process of determining the difficulty scores of the items at each performance level.

In recognition of the inherent subjective nature of standards that are negotiated by panels of specialists in traditional standard setting methods (Cizeck & Bunch, 2007), Shulruf, Poole, Jones and Wilkinson (2014) have developed and advocated a SS method they called the Objective Borderline Method (OBM). The OBM focuses on enhancing the precision of cutscores for borderline learners. It is a mathematical and probabilistic technique, mainly
focusing on ensuring precision of cutscores, especially where stakes are high and the risks of making incorrect judgements must be reduced to the barest minimum (Shulruf et al 2014:16).

Evidently, the emergence of the use of performance standards holds a good promise for reporting results from LSAS generally and in South Africa in particular. However, this study contends that in the absence of a framework that clearly guides the packaging of the results for specific target audiences and how the data should be recorded, reported, disseminated and utilised, the intended impact of LSAS on the quality of education may not be realised.

1.6 The study thesis

The thesis of this study is that the results of LSAS can influence improvements in school and learner performance only when they are recorded, reported, disseminated, interpreted and used effectively by those who make decisions and provide support to schools. In the context of limited expertise, skills, and practical guidelines, there is an urgent need for a comprehensive national framework to guide and support provinces, districts and schools in the effective use of information from LSAS for improving learning and teaching in South Africa.
1.7 Statement of the research problem

While regular LSAS are viewed as a valuable source of information for use in improving learner performance in South Africa (DBE, 2011a), there is limited evidence that results from surveys get communicated to provinces, districts and schools in meaningful ways to influence decision-making for improving learner performance (Kanjee & Sayed, 2013). In a study on how provinces, districts and circuits used ANA data, Govender (2014) reported that provinces and districts relied heavily on ANA results to help them categorise schools for support based on raw percentage correct scores obtained by learners in the tests. Yet the limitations of using raw statistics that compare learners with others to report assessment results in education are well documented (Swanson & Watson, 1982; Montgomery & Connoly, 1987; Bond & Fox, 2007). The key limitation is that raw statistics on their own cannot be used qualitatively to tell what learners were able or not able to do in the tests (Bond & Fox, 2007). It is questionable, therefore, whether support that is given to schools categorised as under-performing on the basis of raw scores only can address real teaching and learning needs.

Furthermore, there is limited information available, as well as lack of clear guidelines that stipulate how data from national assessments should be recorded, analysed and communicated. There is also no guidance on how different levels of the education sector are expected to utilise ANA data for improving learning and teaching, even though most of the goals specified in the Action Plan (DBE, 2011a) are measured in terms of performance in the ANA. In the absence of a national framework with clear guidelines on how ANA data should be presented and utilised, it is highly unlikely that LSAS data will be used for improving learning and teaching in South African schools. This study thus addresses this key challenge of developing a national framework to enhance the use of results of LSAS to improve learner performance in the basic education sector.

1.8 Purpose of the study

The purpose of this study is to develop a National Reporting Framework (NRF) that will provide clear guidelines for reporting information from LSAS, to enhance its use in effective
decision making regarding improvement of learning and teaching in schools. Specifically, the NRF will provide clear guidelines on how information from LSAS:

1. should be recorded and reported;
2. should be communicated and used by national, provincial, district and school communities to help determine what learners know and can do; and,
3. can be used to enhance decision making processes regarding the development of effective interventions to improve school and learner performance.

1.9 Objectives of the study

The key objectives of the study were to:

a) develop a standards-based NRF for meaningful reporting of the results of LSAS in South Africa, with special reference to the results of ANA;

b) investigate and identify the most appropriate method of setting performance standards that will enhance the usefulness of the NRF; and

c) provide exemplars of standards-based reports.

Key principles that underpin the objectives of the study include: the need to ensure that LSAS reports meet the curriculum needs of teachers; reports are easy to read and understand; ensuring that the turnaround times for providing feedback are minimal; data preparation and dissemination are cost-effective; and, data processing involves little to no demand for complicated data decoding skills on the part of teachers and other users.

1.10 Research questions

The key research question that is addressed in this study is:

What should be the key components of a NRF to enhance the quality and optimal utilisation of information from LSAS at national, provincial, district and school levels to improve teaching and learning?

The specific research questions in this study, therefore, are:
1. Which of the three standard setting methods: the Angoff, Objective Standard Setting and Objective Borderline Method, is most appropriate for developing performance standards that are relevant for a National Reporting Framework in the South African context?

2. What guidelines should be provided to national policy-makers, provinces, districts and schools to enhance the utilisation of information from LSAS?

3. What are the views of national and provincial officials regarding comparisons between a standards-based ANA report and a report based on the Curriculum and Assessment Policy Statement (CAPS) rating codes?

1.11 Significance of the study

The benefits of establishing and using an NRF at the different levels of the schooling sector in South Africa include: (i) standardised reporting in a manner that will make the information from the national assessments surveys, in particular the ANAs, easier to understand and use and thus more meaningful to a range of different stakeholders including parents, teachers and education officials; and, (ii) reporting results according to performance levels that provide information of what learners know and can do, and thus enabling teachers and districts to develop relevant interventions to address the needs of specific schools and learners.

As a secondary long-term benefit, establishing the NRF will also benefit the process of developing appropriate test frameworks for LSAS which will, in turn, inform the development of tests that are appropriately targeted to affected learner populations.

1.12 Definition of terms

The following definitions have been provided to ensure uniformity in the understanding of terms used throughout the study.

Large-scale assessment surveys (LSAS): “The process of obtaining relevant information from an education system to monitor and evaluate the performance of learners and other significant role-players as well as the functioning of relevant structures and programs within the system for the purpose of improving learning” (Kanjee, 2007:13).
National Reporting Framework (NRF): A guideline document on the format and content to be included in a report from LSAS to facilitate effective recording, reporting, and use of data to improve teaching and learning.

Performance standards: “... positions on the score scale (cutscores) ... used to indicate whether a learner’s performance meets a certain standard or not” (Näström & Nyström, 2008:1).

Cutscores: “... points of the scale used for representing the test results, each one distinguishing two consecutive levels of competence ...” (Kaftandjieva, 2010:13).

Rating codes: Percentage score ranges specified in the CAPS document for reporting learner performance (DBE, 2011c).

Systemic evaluation: Large-scale assessments of learner cognitive knowledge, skills and competencies which are often accompanied by a collection of information related to factors that might affect teaching and learning in schools (NEPA, 1996).

1.13 Limitations of the study

The central motif in this study was developing an NRF for use in enhancing the effective use of results from LSAS to improve teaching and learning in the basic education sector. The distinguishing feature of the NRF is the use of performance standards to report results from LSAS.

The development of guidelines that must be part of the NRF on how data from LSAS should be recorded, reported, disseminated and utilised was limited to a literature review. This limitation could have been overcome through verification using empirical data from interviews with subject advisors and curriculum officials at national and provincial offices. Due to tight schedules of these senior officials, interviews had to be cancelled on several occasions. However, because the NRF is a policy proposal, once it is accepted by DBE senior management opportunity will be created for public discussions and comments before promulgation.

Although a survey was conducted to test the perceptions of national and provincial officials on the relative strengths and weaknesses of a standards-based and a raw score-based
reporting framework, time constraints on the part of the respondents again could not permit opportunity to verify the findings with detailed interviews to probe what these key role-players perceived were possible challenges to implementing the NRF. The use of the NRF will need to be piloted to establish how provincial and district officials will receive and use it. Feedback from the pilot study will help strengthen the NRF. However, this process itself requires an additional study, the scope of which could not be completed within the time frames of this study, but is nonetheless a key priority to be considered before a rollout of the NRF.

1.14 Organisation of the study

The aim of this study was to develop a NRF that will guide the national, provincial, district and school decision-makers in enhancing their use of information for LSAS to improve learning and teaching. Chapter 1 presented background information briefly noting initiatives that South Africa has taken to address the identified problem of underperformance in the system. One possible explanation for lack of appreciable success in this regard pertains to the limited availability of relevant and timely information to decision makers for developing and implementing relevant interventions to address the key challenge of improving learning. A key reason for this relates to the lack of relevant feedback from LSAS in formats that would help them use information appropriately, and effect change of practice. A preliminary literature review showed that under- or non-utilisation of data from LSAS was not a uniquely South African problem but rather countries that had focused on providing guidance on effective use of LSAS data had realised significant changes in their respective systems. This study aims to develop and propose a NRF that will give clear guidance on how results should be reported and utilised effectively to enhance decision making processes to improve teaching and learning in schools.

Chapter 2 will present research literature to inform and contextualise this study within what has been researched in the areas of reporting and utilising data from LSAS, and thus inform how the development of an NRF can be achieved. The methodology that was followed in collecting and analysing data for the study will be reported in Chapter 3 and the findings and discussions will be presented in Chapter 4. The principles and motivation for the proposed NRF are presented in Chapter 5 and exemplars of standards-based reports, which are part of
the NRF, are presented in Chapter 6. The conclusion and recommendations of the study are in Chapter 7.
2. LITERATURE REVIEW

To position the study within existing relevant research and identify possible gaps in the conceptualisation, implementation, reporting and utilisation of data from large-scale assessment surveys (LSAS), a detailed review of research literature, reported both in South Africa and at international level was conducted. The literature review covers: a) conceptualisation and use of LSAS in education; b) evolution, models and use of LSAS; c) factors affecting use of LSAS data; d) setting and using performance standards in LSAS; e) lessons from case studies on use of LSAS; and, f) an exploration of conceptual frameworks for reporting LSAS. The last section of the chapter proposes options for a national framework for reporting and using LSAS data in South Africa, with specific focus on the Annual National Assessments (ANAs).

A selection of literature was structured with the purpose of the study in mind. To inform the process of developing an NRF for recording, reporting and using results of LSAS in South Africa, which is the main research problem in this study, and to find answers to the specific research questions, a review of research literature from a wide range of relevant sources was conducted. Primary sources that were reviewed included research journals, dissertation abstracts, conference papers and the World Wide Web. Relevant textbooks were also consulted to either firm up or obtain alternatives to the theoretical perspectives that underpinned this study.

The literature search focused mainly on research about the use of data from large-scale assessments to improve teaching and learning, completed studies on the impact of data use to guide decision-making in the education field, and theoretical frameworks that guided the conduct of such studies. Research on the setting and use of performance standards was also an area of interest in the literature review, with special focus on literature reporting studies on different methods of standard setting, the methodologies that were used, and the impact of such standards on the affected education.

2.1 Conceptualisation and use of LSAS in education
The effective use of assessment to identify what learners can or cannot do and what learners know or do not know is an integral part of the teaching and learning process. The DBE (2012:3) defines assessment as:

“a process of collecting, analysing and interpreting information to assist teachers, parents and other stakeholders in making decisions about the progress of learners”.

The Ontario Ministry of Education (2010:6) shares a similar definition but emphasises that the primary purpose of assessment is to improve learning. In practice, teachers in classrooms carry out the assessment of learners on a formative or developmental basis to identify learning gaps and devise more effective teaching strategies that will bridge the gaps and enhance learning (Griffin, 2009). In addition, teachers also administer tests and examinations as forms of assessment that help them make summative decisions on whether learners meet specified levels of knowledge to qualify for progression to the next grade, for awards or certain exemptions.

Unlike classroom assessments, LSAS are managed and conducted at the national (or provincial) level, often in specified periodic cycles, and are designed to provide snapshots of strengths and weaknesses in the education system (Ontario, 2010). The results of LSAS are commonly used to inform decision-makers on how schools perform in relation to one another, areas that require specific interventions, and policies that either need to be reviewed or monitored more closely (Kellaghan, Greaney & Murray, 2009).

Based on the purpose for which the results will be used, assessment bodies have gone further to differentiate among assessment of learning, assessment for learning and assessment as learning (Ontario Ministry of Education, 2010). Although the distinction among these purposes may be ambiguous, LSAS would in the main fall under assessment of learning while assessment for and as learning would be more appropriate in formative settings in individual classrooms. As a measure of learning, LSAS need to provide answers to questions such as: How much (quantity) learning has taken place? What kind (description) of learning has taken place? By how much has learning changed over time? Attempts to answer all these questions, and others, in succinct but clear and meaningful terms, have led to the evolution of various approaches and methods to conducting and using data from LSAS (Greaney & Kellaghan, 2008:85-144).
Another critical issue that impacts upon the use of LSAS relates to whether a norm-referenced or criterion-referenced framework is adopted (Conneley, 2004; De Champlain, 2004; Ontario Ministry of Education, 2010). In norm-referenced assessment a particular test-taker is compared with others who took the same test according to his/her score rank in the overall distribution of test scores (Conneley, 2004:3). Instances of norm-referenced reporting include specifying how many standard deviations an individual test-taker’s score is above or below the mean score of the population of similar test-takers or stating the percentile ranking of the individual test-taker (De Champion, 2004:62). Thus, within a norm-referenced framework, standards are relative.

In a criterion-referenced framework the reported score or standard represents the amount of knowledge and skills that a competent test-taker is expected to demonstrate in the domain that were tested, irrespective of how other test-takers perform (Conneley, 2004; De Champion, 2004). In this paradigm, the assessment generates information that must inform decisions on actions to be taken to remediate the observed situation or outcomes. Similarly, the Ontario Ministry of Education (2010:145) defines criterion-referenced assessment as; “Assessment that focuses on whether a learner’s performance meets a predetermined standard, level, or set of criteria rather than on the learner’s performance measured in relation to the performance of other learners”. Thus, within a criterion-referenced framework, standards are absolute.

The criterion-referenced school of reasoning has led to a proliferation of methods to define, set and operationalize criteria, or standards, which can be used to map learners’ scores to specific repertoires of knowledge and skills in education and other fields (Greaney & Kellaghan, 2008). There is simultaneously a growing body of research literature on guidelines and frameworks on how to set performance standards that are relevant to specific contexts (Cizek & Bunch, 2007).

Target users of information from LSAS include teachers, school principals, district officials, provincial and national policy makers and parents. Given the diversity of potential users, it is vital that feedback from assessments in general, and from LSAS in particular, be organised and presented in formats that meet the needs of specific user groups. For example, teachers may need information that would help them identify levels, knowledge and skills in which
learners demonstrate specific strengths or weaknesses. Parents may want to know how the 
schools that their children attend perform in relation to others that are in similar contexts. 
Provincial and national officials may be interested more in information that points to policies 
which need to be reviewed, and evidence that suggests need for re-allocation of resources, 
whilst districts may need information that will help them identify schools and teachers who 
may need specific kinds of interventions and support to be able to improve learning.

Regardless of the target users, there has been concern that information from LSAS has not 
received due attention, partly because it is often couched in statistical jargon that users cannot 
decipher (Hambleton & Slater, 1997). However, Kellaghan, Greaney & Murray (2009:20-21) 
contend that there is not enough research that has been done in the area of use or non-use of 
LSAS findings. In South Africa, Kanjee (2012) studied trends in reporting the results of 
LSAS and concluded that, although the results had been considered in some policy-related 
decisions, there had been limited focus on using the results to support teaching and learning. 
Given the enormous investments that go into the LSAS, it is critical that those who 
commission or implement LSAS be clear about the interpretability, use, value and impact of 
information obtained from these surveys.

In arguing for the catalytic role that assessment can play in changing practice and policy in an 
education system, Braun and Kanjee (2007) modelled an education system that is defined by 
four key interrelated attributes, viz. access, quality, efficiency, and equity. These authors then 
invoke the concept of “systemic validity” of an assessment system, which they define as the 
ability to generate information that induces change in one or more of the modelled key 
attributes without causing deterioration in the others. Thus, an assessment system is regarded 
as systemically invalid if it does not induce change in at least one of the modelled key 
attributes. For instance, an education system that cannot ensure that all learners are able to 
attend school without being debarred by physical or policy obstacles (access) lacks equity 
because it is effectively denying educational opportunities to the affected learners.

Most LSAS seem to be modelled for a particular short-term purpose rather than guided by a 
long-term conceptual framework. In tracing the evolution of large-scale assessments in ten 
provinces of Canada, Klinger, DeLuca and Miller (2008) identified four categories that were 
modelled according to specific purposes and uses: for accountability, for “gate-keeping”, for
instructional diagnosis or for monitoring learner achievement over time. The authors, however, note that the purposes seemed to change from time to time depending on political, pedagogical and economic dictates in a given period rather than guided by a defined framework.

According to Klinger et al (2008), using the assessments for accountability involved publishing results in the form of school rankings and thus giving parents information for selecting schools for their children. A marked example of how this approach to using test results can be misleading was given by Goldstein (2001) in the context of England and Wales where annual test results at Key Stages 1-4, equivalents of South Africa’s Grades 6, 8 and 10, were used to draw up “league tables” that compared schools on the basis of the common test results. Goldstein (2001) criticised the fact that “league tables” wrongly ignored antecedent variability in the environments in which teaching takes place, different learning styles of individual learners, different school intakes, and the differential impact that schools have on learners. In South Africa, it has been noted that differences in learner performance could be accounted for more by between-school than within-school differences as indicated by corresponding intra-class correlations (Moloi & Strauss, 2005). There is little wonder, therefore, that this model of using assessment was reported to have caused a great deal of disquiet among teacher representatives in Britain and, in the absence of a clearly defined framework to justify its retention, had to be modified (Goldstein, 2001:6).

Modelling large-scale assessments for “gate-keeping” involves using their results to award grades, to determine admission to higher institutions or to offer financial rewards. This model often goes under the name of “public examinations” and in South Africa it occurs only at the grade 12 level. According to Greany and Kellaghan (2008) the model of LSAS for accountability has often been used in systems that do not have traditional public examinations and the large-scale assessments in reality serve as “surrogate examinations”.

Modelling large-scale assessments for instructional diagnosis involves identifying strengths and weaknesses that are reported only to affected stakeholders, for example individual learners and their teachers, for purposes of developing appropriate intervention programmes (Klinger et al, 2008). Examples of such assessments have been reported in Uruguay (Ravela, 2005) and Chile (Schiefelbein & Schiefelbein, 2003). Ravela (2005) describes how
Uruguay used the results from census-based large-scale national assessment to inform training of teachers which in turn led to improved teaching strategies and improvements in the national results. In Chile, teachers were even given salary incentives in recognition of increases in the test scores of their learners in the LSAS (Klinger *et al*., 2008; Ramirez, 2012).

Whilst there has been a growing interest in conducting large-scale assessments among education systems across the world (Greaney & Kellaghan, 2008), evidence indicates a range of reasons for doing so but also indicates that data use for improving teaching and learning is a culture that develops over time, sometimes with unintended consequences (Braun & Kanjee, 2007). In The Gambia Senghor (2014) reported that results of LSAS were, in some instances, used to decide whether learners were going to be promoted to the next class or retained while in other cases they were used to evaluate teacher performance. Consequently, according to Senghor (2014:18), the misuses led to competition among schools and teachers. In this context, Senghor (2014:18) cautioned that these unintended consequences could distort the results of LSAS. There were also instances that Senghor (2014:18) referred to as “narrowing the curriculum” because teachers focused teaching on topics that they saw featured prominently and regularly in the assessments.

In another study, Kanjee and Moloi (2014) conducted a survey on a sample of South African primary school teachers on their perceptions of the usefulness of ANA, their experiences as well as possible challenges that they faced in administering the ANAs and utilizing the resulting data. Kanjee and Moloi (2014:102-103) report that the teachers shared a variety of lessons that they singled out as examples of “best practices” they perceived emanated from ANA. Firstly, rather than expressing views in favour of “teaching to the test”, a prevalent view among both Intermediate and Foundation Phase teachers was that the assessment helped them adapt their teaching methods to suit the identified or diagnosed learning needs of learners. Secondly, teachers were of the view that ANA modelled best practices for developing tests of a high standard at the institution level. Thirdly, teachers indicated that ANA helped them know what curriculum content to cover to prepare learners to do better in subsequent assessments without necessarily teaching to the test.

Kanjee and Moloi (2014:103) observed that, while teachers seemed to be of the view that ANA held a good promise for informing effective teaching and learning, they were, however,
not aware of how to utilise the resulting data for this purpose. According to Kanjee and Moloi (2014:104) and Datnow, Park and Kennedy-Lewis (2012), teachers need to be adequately qualified, have relevant knowledge and expertise to know how to analyse, interpret and use data from assessments to enhance the quality of teaching and learning. Accordingly, Kanjee and Moloi (2014:108-109) argue that teachers required intensive training on how to develop a theory of action that will direct their initiatives for using ANA data; they must be provided with the necessary tools of the trade, e.g. an Excel capturing tool for analysis of assessment item-level data and must be trained on how to categorise learners into performance levels marked by cut-scores so that teaching interventions are targeted to learner strengths and weaknesses.

2.2 Factors affecting utilisation of LSAS

While assessment itself, through the information that it generates, has the potential to lead to identification of factors that are responsible for underperformance, the key issue is that information must be used effectively in decision-making so that relevant interventions can be developed and implemented to address the identified problems and thus improve performance (Kellaghan et al, 2009; Kanjee & Moloi, 2014; Schiefelbein & Schiefelbein, 2003).

Kellaghan et al (2009) observed that one of the shortcomings in the countries that conduct national assessments was the under-use of the data which the assessments make available. According to these authors, the evidence that is available (worldwide) indicates that the use of national assessment findings is not widespread, despite the potential that information derived from an assessment has for sparking reform and despite the expense incurred in obtaining such information. On the one hand, following studies of a number of national assessments and how they had either made an impact or not, Kellaghan et al (2009) proffer a number of reasons for non- or under-utilisation of information from the LSAS. These 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, lack of capacity, and absence of appropriate tools to help teachers use the data. Other researchers (Hambleton & Slater, 1997; Hambleton & Pitoniak, 2006; Underwood, Zapata-Rivera &Van Winkle, 2010) have also blamed reports from national assessments as being complex, difficult to read, and even more difficult to interpret.
In the wake of the Federal No Child Left Behind (NCLB) Act (2002) which required USA states to use assessment data to monitor the extent to which states met their targets in terms of this law, Coburn, Honig and Stein (2009) conducted a comprehensive literature review on whether school districts in the USA routinely and systematically used evidence from learner achievement data to inform their decisions and, if not, what factors discouraged them from doing so. According to these authors, the use of learner achievement data in terms of the NCLB law assumed a high-stakes status because decisions on whether districts received additional funding or benefited from preferential professional development of educators depended on presented empirical evidence from learner achievement data. Some of their findings were quite revealing and, although they may not be generalizable to different contexts, they raise general principles for use of data in decision-making. Specifically, these address the role that data or empirical evidence plays in decision-making as well as key principles underpinning the use of such data.

According to Coburn *et al* (2014:12-18) there are four key roles that evidence plays in decision-making: instrumental, conceptual, symbolic, sanctioning and no role.

**Instrumental role:** In this role data provides the missing knowledge or understanding to solve a known problem. The tacit assumption in this regard is that once appropriate data is made available and used appropriately a solution will be reached through the use of the data. Coburn *et al* (2009) consider that, although somewhat popular, the instrumental role of data use makes oversimplified assumptions about the nature of data and the contexts in which the data must be used. These authors argue that the assumption, evident across a number of studies they reviewed, of a linear relationship between data use and decision-making, i.e. assuming that correct use of appropriate data will logically and directly lead to relevant evidence-based decisions, is somewhat tenuous. The reality was that this relationship was complex and often messy.

Specifically, the complexity of the instrumental role of data use arises out of technical, social and political factors at play at the time of decision-making. Firstly, the data that serves as evidence is often presented in forms that users, in this case district personnel, found too abstract or too technical to interpret. In these circumstances district officials were less likely to consider such evidence in decision-making and subsequent actions (West & Rhoton,
Q. Moloi

1994). The more concrete and user-friendly the data, the more it was likely to be used in decision-making processes.

Kanjee (2007) traced trends in the introduction, growth and changes in purposes of LSAS in South Africa since mid-1990s and observed that most of the surveys were designed to provide information for policy makers and very few aimed at providing data for teachers and teaching. While the provision of assessment data was instrumental in showing the extent of under-performance and inequalities in performance in the education system, Kanjee (2007:531) observed that effective use of the data, especially at classroom level, was compromised by three factors, viz. absence of policies and guidelines on how the data should be used, lack of necessary capacity and expertise among officials and teachers on how to use data and absence of relevant resources and tools for data use. As a result, Kanjee (2007:531) proposed that development and implementation of a national framework with clear guidelines on how assessment data should be analysed and reported to various users was one solution to ensuring that data use made an impact on performance in the education system.

Secondly, it was found that, rather than data evidence leading directly to decision-making, social and political worldviews, assumptions and prior knowledge of the users often influenced whether they will or not use available data (Coburn et al, 2009:7). The inference was that, unless properly mediated with affected groups of users, data that was presented to district officials and educators who held divergent views about the contemporary social and political contexts was more likely to be superseded by contextual factors in guiding decision-making processes. By extension, achievement data that is reported in forms that do not match the epistemological belief systems of educators, i.e. ways in which educators believe that knowledge is acquired or constructed by learners, could be either ignored or rejected (Timperly, 2009:21). Data do not speak for themselves but are often interpreted and weighed against other factors such as budgets, administrative protocols and political dynamics even where evidence would seem to play an instrumental role (Coburn et al, 2009:13).

Conceptual role: In this role data evidence is used either as background information, to increase awareness, to provide new individual or shared knowledge, but is not used to influence or take any particular decision (Weiss, Murphy-Graham & Birkenhead, 2005). Coburn et al (2009:14) argue that, although conceptual use of data has no immediate and
direct influence on specific decision-making, it has powerful potential to become part of on-
going institutional interactions, attitudes and behaviours which may be either positive or
negative. A classic illustration of the conceptual role of data or evidence in non-educational
circles was noted by Weiss et al (2005) who reported on how intensive and repeated media
coverage of the inefficacy of a large drug prevention programme in Los Angeles increased
knowledge about drugs among young people but had little or no effect on their drug abuse
behaviour. Applying this concept in South Africa, there is widespread awareness about, and
acceptance of the ANA (Kanje & Moloi, 2014), both in the education system and in the
public media. But this knowledge or awareness may not be assumed to be prompting action
from those who are supposed to use the data.

**Symbolic role:** In its symbolic role, data evidence is used to justify decisions or actions
that have already been taken instead of prompting new decisions or actions. This could be for
various reasons which include political, technical and social pressures that often characterise
operations at district level (and other levels of the system) and prompt officials to totally
discard evidence in favour of what they anecdotally believe to be desirable in the prevailing
circumstances (Coburn & Talbert, 2005; Coburn et al, 2009). Needless to argue, collecting
data at huge costs and then only using it for symbolic purposes may not be a sustainable
option for education systems that need limited resources for competing priorities. Education
officials in many of South Africa’s provinces, especially large and predominantly rural
provinces, are often confronted with dealing with many teachers in a Circuit or district. In
these circumstances, workshops that are intended to be interactive sessions that are ideal for
building capacity for data use may easily degenerate into plenary sessions for talking about
data and may not contribute to data use to improve teaching.

**Sanctioning role:** According to Coburn et al (2009:16-17) evidence plays a “sanctioning
role” when its use or presentation influences choices that can be made in order to secure
approval from a higher authority. Evidence is used to impose compliance. For instance, while
all public schools are compelled to participate in the ANAs, the DBE allows independent
schools to choose if they want or do not want to participate but those that choose to
participate and show acceptable performance stand to secure state subsidies. Independent
schools may, therefore, participate in ANA not because they are convinced about the
evidence that ANA produces but because compliance helps them stand a good chance to have their applications for subsidy approved.

Coburn et al (2009:16-17) reported that when the USA federal government made evidence-based decisions a requirement through legislation like the NCLB imperatives, districts, apparently in the absence of a common regional or national curriculum, selected school programs that they knew were popularly known for evidence-based claims in order to secure easy approval even though they had different views. To the extent that different real and potential user groups perceive and agree on the importance of data use in decision-making, initiatives that are primarily driven by sheer compliance with requirements for data have the potential to influence certain sectors of the system to use data to inform decision-making. In other words, the politics at play may result in parts of the same system or institution using the same data for diametrically opposed purposes.

The use of ANA to hold teachers and/or schools accountable will fall into the category of the sanctioning role of data use. Unless schools and teachers are convinced about the value that the use of ANA data can add to the effectiveness of teaching and learning, it is possible that their participation in the assessment may be sheer compliance with national policy.

No role: There was literature which, according to Coburn et al (2009:17), showed that a number of districts did not use any data to inform their decisions. All decisions were based either on anecdotal, political or financial considerations. This could be typical of many schools and districts in South Africa. Because ANA results are not included in assessment for promotion, for instance, teachers may administer the assessment just for compliance because they perceive it as having no consequences and erroneously believe that the data has no practical use or value. In the study by Kanjee and Moloi (2014:103), the majority of the South African teachers who were surveyed concurred that ANA had an important role to play in effective teaching and learning even though their views on how the assessment data was to be used varied widely.

Key factors which, according to Coburn et al (2009), seemed to affect the role that data played in decision-making by districts included policy pressures, political imperatives, and organisational capacity constraints. In separate studies, Marsh (2012) and Parke (2012) synthesised research literature on factors that either promote or inhibit the use of data by
schools and districts in the USA. Some of the factors that they reported were likely to promote data use included leadership capability in data use, accessibility of relevant databases, relevant skills and knowledge of targeted data users and the user-friendliness of the data. Both studies showed that lack of skills and knowledge of statistical data inhibited educators and policy makers from making use of data for fear that they might misinterpret results and arrive at misleading conclusions (Parke, 2012:3; Marsh, 2012). Similar findings have been reported in other studies. Govender (2014) conducted a survey of South African education officials at district and circuit levels to establish the extent to which they understood and used ANA data in supporting and managing schools. Some of the key and pervasive findings of the survey were that lack of proper leadership in data use, the absence of a culture of data use and dearth of appropriate skills for analysing and presenting data in meaningful formats were some of the factors that inhibited effective use of data to improve teaching and learning in schools.

Although intrinsic motivation was reported to be the key driver in cases where data was used to inform practice, capacity building and support were found to be necessary elements to sustained use of assessment data by teachers (Marsh, 2012:19). One factor that was found to contribute to motivating educators to use data was the way the data was presented, that is in usable and easy-to-understand formats (Marsh, 2012:12). There were, however, cases where even though data was simplified and made comprehensible, educators were still reluctant to use the data because they suspected that the results were going to be used beyond identifying learning deficiencies but were actually going to be used to evaluate the teachers (Goldstein, 2001). Under these conditions and fearing what they perceived as possible victimisation, educators resorted to use data only for either symbolic purposes or no role at all.

Generally, researchers seem to concur that use of data for decision-making in educational circles is influenced by a plethora of factors and that focusing on only one factor may not necessarily lead to the expected behaviour. Some of the case studies that were reviewed for this study had revealing findings and guiding lessons that education systems can use to enhance the effective utilisation of LSAS data to improve the quality of teaching and learning in their schools.

2.3 Case studies of effective utilisation of LSAS
The review of case studies was framed by one broad question: Under what conditions has effective utilisation of LSAS data to improve teaching and learning been successful? The case studies were selected from both developing or under-resourced and developed or affluent education systems, and include Kenya, Namibia, New Zealand, Uruguay, and the USA. Some of the findings on what worked and what did not work have been summarised in this section with a view to lifting applicable experiences for the purposes of this study.

**New Zealand**

Considerable work has been reported from New Zealand, not only on the LSAS that were conducted, but also on intensive research initiatives that were undertaken to understand how feedback from the surveys could be utilised optimally to promote effective teaching and learning (Schägen, Hutchinson & Hammond, 2006; Hattie & Timperly, 2007; Timperly 2009). Specifically, the research findings showed that where relevant feedback on performance in the LSAS was provided, both teachers and learners were motivated to identify and engage in more cognitively challenging learning activities (Hattie & Timperly, 2007:86).

Timperly (2009) gave an account of how teachers in New Zealand had, after extensive experiences, identified some of the conditions that they considered necessary to promote a culture of data use for improving teaching and learning in schools. Key conditions that Timperly (2009:21) identified can be summarised as follows:

a. teachers need a certain minimum level of knowledge about assessment data for them to use this information effectively in teaching;

b. developing a culture of data use requires considerable time;

c. teachers need assessment data that is relevant to the curriculum that they follow;

d. use of assessment data to improve teaching and learning requires a certain minimum level of pedagogical and content knowledge by the teacher; and,

e. instructional leadership plays a critical role in promoting a culture of data use in schools.
What made the New Zealand experience, as reported by Timperly (2009), unique was that it was a large-scale initiative that involved 300 schools, as well as learners and teachers from the schools, school leaders and parents. Often initiatives of this kind are implemented on a small-project scale and then subsequently experience new challenges when they are taken to scale. It can be expected, therefore, that lessons learnt in studies of the New Zealand scale and larger will be enriched with wide-ranging experiences and are likely to cover a wide range of contextual idiosyncrasies. The studies may not necessarily be replicable, but they provide principles that could guide implementation in different contexts.

**Seattle (USA)**

An experience that could complement the New Zealand school experiences with data use was reported by the American Association of School Administrators (AASA: 2002:14) with use of data by districts to manage, monitor and support schools. Some of the lessons learnt from one district in Seattle could guide implementation and initiatives in other contexts. In a district that had consistently adopted a data-driven decision-making approach, the District Superintend shared the following guidelines that they had followed on how to make the best use of data at district level and lead schools to do the same (Timperly, 2009):

a. The district started “small” and avoided overwhelming school staff with huge ‘data dumps’ beginning with learner achievement in reading and mathematics as core issues;

b. They learnt to listen to what the data told about the big picture and avoided getting lost in too many details;

c. To avoid recrimination by staff, they worked hard and patiently to create trust and build support to promote use of data without fear; and,

d. A wide range of training opportunities were created for staff to learn how to use data.

All available evidence shows that, making data available will not, on its own lead to its use by teachers (Ingram, Louis & Schroeder, 2004). Guidance, capacity building, support and provision of relevant tools all make for a conducive atmosphere for optimal utilisation of data to make observable positive impact on teaching and learning (Timperly, 2009).
Research literature suggests that, although most national assessments seem to be initiated with good intentions, often the assessments are not guided by a clearly defined framework for their development and utilisation (Goldstein, 2001). Consequently, the results of the assessments end up being either not utilised, or under-utilised or utilised for unintended purposes. In these circumstances assessment may not make the intended impact. This study, supported by emerging evidence and wide-ranging experiences with data use, postulates that a framework that provides clear and relevant guidelines and examples of what can be done with the data has the potential to enhance data use to improve teaching in schools.

**Namibia**

In Namibia, the results from SACMEQ II (Makuwa, 2003) showed that in seven of the 13 regions, between 80% and 90% of grade 6 learners were performing at “non-numerate” levels. A closer investigation revealed that in four of the seven under-performing regions, only 20% of the grade 6 teachers achieved acceptable mastery levels on the SACMEQ mathematics “described scale” and most of them had never been trained to teach mathematics. This level of granularity in reporting prompted specific actions in the education system. The Ministry established what they called the Management Policy Co-ordination Committee (MPCC) to initiate and drive programmes that would address equity and quality bottlenecks in the system. Districts developed checklists of minimum expectations on what teachers should do in the classrooms, what actions were expected from principals and inspectors were given specific standards by which to evaluate schools. Makuwa (2003) reported that the effect of these initiatives brought about significant improvements in performance in Namibian schools.

**Kenya**

In Kenya, comparisons between the results of national examinations at grade 6 level and the SACMEQ II results showed not only that learner performance was lower in both Reading and Mathematics scores, but, because the results were reported in performance standards with specified cut-scores, the also helped identify specific deficiencies in curriculum mastery (Nzomo & Makuwa, 2006). In addition, the SACMEQ results showed hitherto unknown regional variances in performance and also variances between expected resource levels, for example, the variances between the number of textbooks per learner according to policy and
the actual textbooks that were available in schools. The Ministry’s response included a national curriculum review process, reduction of the number of prescribed subjects at grade 6 level as well as in the number of examinable subjects in order to increase efficiency and enhance performance.

It is worth noting that in both Namibia and Kenya overall learner performance in SACMEQ III, reported in terms of knowledge and skills that learners displayed at each performance level, was appreciably higher than in SACMEQ II (Nzomo & Makuwa, 2006). Although a direct causal relationship between the interventions and the improved results cannot be claimed, given the known complex interrelations among various factors in education, the effective reporting, dissemination and utilisation of the assessment results in these education systems justifies the investment (Nzomo & Makuwa, 2006).

Uruguay

Ravela (2005) notes that the Ministry of Education in Uruguay, which also conduct a census assessment similar to that of the ANAs, effectively used assessment information to improve learner performance, with largest gains reported in schools serving low socio-economic learners. The Ministry instituted a programme to improve learning by combining the use of assessment information with teacher development and school-based interventions. The programme was conducted over a number of months, and comprised: the use of scores obtained from a census assessment survey to identify poor performing schools; bi-weekly teacher meetings, held during weekends, to discuss school assessment results; providing teachers with the necessary training and to develop interventions to address specific needs of learners; and, follow up sessions to discuss results of previous week’s intervention and to plan for the next two weeks (Ravela, 2005).

With the caveat that solutions to general problems in education need to fit local contexts, the education systems in the literature reviewed for this study could provide key lessons for South Africa regarding the use of LSAS to influence school and learner performance in the desired direction. Braun and Kanjee (2006) reported some of the specific strategies that were employed by the Chileans in effectively reporting and successfully utilising results from national assessments that influenced school and learner performance. To a lesser degree of detail, Braun and Kanjee (2006:29) reported how Brazil defined the mechanism for
conducting national assessments and how, subsequently, the results of the national assessments were accepted by all role-players in education and used for monitoring policy formulation and quality of education.

Successful accounts have been reported of observable changes in policy, practice and learner performance that have been linked directly to effective utilisation of feedback from national assessments (Nzomo & Makuwa, 2006; Ravela, 2005; Bernard & Michaelowa, 2006). Nzomo and Makuwa (2006) attribute the effectiveness of the utilisation of the SACMEQ results in their respective countries to the specificity of reporting. The results were reported using performance standards, i.e. according to levels of competency on “described scales” derived from the item response modelling (item maps) and sub-divided by cut-scores. A detailed description of the skills required to answer items related to a particular level was also given.

Taken together, the key mechanisms to which the impact of national assessments in Namibia, Kenya and Uruguay could be ascribed include:

a) reporting performance in well-defined standards that are characterised by clear cut-scores and definite performance level descriptors;

b) using a differentiated approach to communicate the results of LSAS to specific target audiences, for example teachers, principals, district officials, teacher trainers, parents and high level policy makers;

c) tracking improvement in performance in LSAS over time;

d) engaging competent professionals to facilitate LSAS information dissemination sessions at various levels of the education system;

e) designing and releasing consistent key messages about the results of the LSAS in the media; and,

f) identifying and rewarding disadvantaged schools that make efforts and show evidence of improvement in the LSAS.
The evidence exemplified from the case studies provides a useful backdrop against which current practices of LSAS in South Africa can be studied and appropriate suggestions for improving the utilisation of the assessment data to improve the quality of teaching and learning within the relevant context. In particular, the technique of reporting assessment results in well-defined performance standards that are embedded within a national framework for reporting was one of the common features among the reviewed case studies where data use was reported to have impacted positively on teaching and learning (Makuwa, 2003; Ravela, 2005; Nzomo & Makuwa, 2006).

Some of the common features among the countries that have reported significant improvements following application of assessment results were: a) specificity in reporting what learners could or could not do in the LSAS tests using performance levels with clear cut-scores; b) a simple and clear communication strategy which outlines the purpose of the assessment, the results and their implications at each level of the system; c) sustained activities by a stable team to disseminate results to all stakeholders; and, d) broad participation and ownership of results by staff at all levels of decision-making.

2.4 The value and challenges linked to feedback from LSAS

Mirazchiyiski (2013) traced the history and recounted the achievements of the International Association for the Evaluation of Educational Achievement (IEA) in conducting LSAS that involved countries across continents and cultural boundaries but added that the IEA studies were fraught with challenges when it comes to providing timely, meaningful and useful feedback to participating schools and learners (Mirazchiyiski, 2013:7-11). In particular, Mirazchiyiski (2013) pointed to the limitations set by the sampling and instrument designs of the IEA studies which resulted in limited usefulness of feedback from these studies at class and individual learner levels.

One of the recommendations that Mirazchiyiski (2013:52) made towards enhancing the usefulness of IEA reports at school level was to include in school reports information on the performance of learners from schools in similar and comparable contexts. While this recommendation may sound admirable, it is likely to be limited to a norm-referenced paradigm within which learners are compared with others who took the same test (De Champion, 2004) rather than comparing their performance with set standards. Given the
remarkable developments that have taken place in the area of development and use of performance standards since the genesis of the IEA studies, this study argues that Mirazchiyiski’s recommendation could be enriched significantly if school and learner reports from LSAS could be steeped in a criterion-referenced paradigm through reporting against objective performance standards.

2.5 The quest for standards

Perceptions that link assessments with “standards” are common in society. Reference to “standards” is generally associated with real or perceived quality of a product or a service, hence one often hears reference to ‘rising or falling standards’ without the users attempting to define the standard. This is because there is a(n) (tenuous) assumption that the users and their audience(s) understand the same meaning. Although defined differently from one context to another or defined variously even within the same context, the notions of “quality”, “satisfaction”, “fitness for purpose”, “acceptable requirements”, etc., often seem implied. In educational circles, researchers (Cizek, 1996; Rodriguez et al, 2011) distinguish between “content standards” and “performance standards”. Although there could be overlaps in these concepts, each one seems to have a specific focus.

2.5.1 Content standards

In specifying the focus of ‘content standards’, Rodriguez et al (2011:18) define them as “… what learners need to learn…” In the context of South Africa, content standards are spelt out in the Curriculum and Assessment Policy Statements (CAPS) per grade and per subject (DBE, 2011c). Content standards specify the nature and scope of content knowledge that a learner must acquire in a given grade. If it is accepted that content knowledge in any field is ever growing, then the need to define specific quanta of knowledge and skills becomes obvious.

The post-apartheid South Africa has come a long way is establishing and, on a number of occasions, revising the national curriculum to reflect and emphasise the content that must be taught in the classrooms, without necessarily downplaying other dimensions that are important in a balanced curriculum (DBE, 2002). Not only should the content be clear and amenable to instruction, but clear procedures on how assessment will be conducted and
reported are the subject of CAPS (DBE, 2011c). CAPS specifies the content to be taught, and how often assessment should be conducted and reported. What CAPS does not specify is how much of the content in the curriculum should be mastered by learners in a grade and at what level. CAPS does not define “acceptable” levels of performance, although there seems to be loose reference to, and expectations about “acceptable levels of performance” in the sector as can be observed from the Presidential injunction of 2010 (South African Parliament, 2010).

Apparently referring to “content standards”, a Ministerial Task Team that was commissioned to review what was then a principally outcomes-based national curriculum in South Africa fore-grounded their findings by observing that:

“A nation’s national curriculum is at the heart of its education system. It is a primary source of support and direction for learning and teaching in the education system, and plays the role of equalizer in terms of ‘educational standards’” (DBE, 2009:11).

The Ministerial Task Team (DBE, 2009:34) then proceeded to include copious notes on how and why more content needed to be included in the national curriculum, adding that “… learners need to have a store of (content) knowledge on which they base their thinking”. Incidentally, virtually nothing was said about “performance standards” in the Ministerial Task Team report, an indication that perhaps the concept and value of “performance standards” may neither have been considered nor taken as a priority in South Africa.

2.5.2 Performance standards

Rodriguez et al (2011:18) define ‘performance standards’ operationally as “… a framework for evaluating how much of these knowledge and skills learners should have mastered”. So, while content standards answer the “what”, performance standards answer the “how much” question. An apt description of the purpose of performance standards is offered by Hamilton (2000:2):

“Well-defined domains of content and skills and performance categories for test score interpretation … aimed at describing what examinees know and can do”.

36
Hambleton (2000) notes that the primary purpose of using performance standards in reporting LSAS results is to place examinees into a set of ordered performance categories. The definition, description and setting of performance standards and their use constitute the greater part of the proposed NRF in this study.

### 2.5.3 Setting and reporting performance standards

Setting performance standards involves determining a cut score that separates test-takers into identifiable categories on the basis of what they know and can do or do not know and cannot do in a test. Along the continuum from not knowing or not being able to do anything at all in a given test, on the one end, to the other extreme of knowing or being able to do everything, what score is necessary for performance to be judged as “acceptable”? According to Barton (1999:19), “A set of content standards and a set of test questions intended to reflect that content lead directly to setting performance standards”.

Seminal work on setting performance standards is traceable to Angoff (1971). Cizek and Bunch (2007:13) define standard setting as “a process of establishing one or more cut scores 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”. The standard setting process involves technical analysis of learner responses as well as expert inputs from teams of professionals and members from relevant stake-holder groups who serve to validate the technical results (Tiratira, 2009).

Performance standards specify the amount of knowledge and/or skills that a test-taker must demonstrate, on a given continuum divided by cut scores, to satisfy the demands of a particular outcome. An important feature of performance standards is performance levels (PLs) and performance level descriptors (PLDs). Zieky and Perie (2006:3) 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. PLDs are then defined as detailed descriptions of “the knowledge, skills, and abilities to be demonstrated by learners who have achieved a particular performance level within a particular subject area” (Zieky & Perie, 2006:4). Morgan and Perie (2004:5) affirm that PLDs are “… working definitions of each of the performance levels (that) … define the rigor associated with the performance levels”.

37
While different methods of developing performance standards and corresponding PLDs have been explored, a common conclusion from research has been that each method has its own strengths and weaknesses and, therefore, the “best” standard setting method may not exist (Tiratira, 2009). Notwithstanding possible pitfalls related to the complexities and subjectivity inherent in different methods of setting performance standards, Cizek and Bunch (2007:36) and Bejar (2008:3) emphasise the critical role of communicating and using clear standards to enhance decision-making at both learner- and policy-level.

At individual learner level, teachers need to identify learners who meet or do not meet expected levels of knowledge or skill, categorise them according to need, and decide on appropriate interventions for each category. Not only do standards help focus interventions or support to address specific individual needs, but they also ensure that learners and teachers share a common understanding of what is expected in learner performance (California Department of Education, 2007:2). Clear knowledge of what is expected spurs individuals, learners and teachers, to exert more effort to achieve it. Standards also help improve the precision with which reporting to parents can be made. Clearly and precisely communicated learner progress reports which are based on fair, valid, and easy-to-understand standards demystify expectations and increase trust between the concerned institution and its clients – parents in this case (California Department of Education, 2007:2).

At policy level, the USA’s mandatory No Child Left Behind Act (NCLB, 2001) is a classic example of the importance of using clear standards but also of the need to contextualise standards. Cizek and Bunch (2007:5-11) report that the mandate requires that achievement for all learners in specified school grade levels and subjects be reported using the performance categories Not Achieved, Achieved, and Advanced, based on clear cutscores that define the borders between any two of the categories. Notwithstanding ensuing legal battles and critical social debates which were mainly related to contextualisation of standards, the use of standards enhanced the use of evidence to make or dispute informed decisions, specifically in addressing matters of equity in educational provision and performance (Cizek & Bunch, 2007:11).

Use of standards also increases the accuracy of determining the percentage of learners who perform at a given level and, more importantly, identifying the social, economic and
geographical profiles of the learners in each category (Kanjee & Moloi, 2012). In South Africa, where social inequality is known to be among the highest in the region (Van den Berg, 2007:1), the use of standards in reporting LSAS results could provide more accurate data to inform policy decisions on the levels of resources such as funding, professional staff and learning materials, to be deployed in specific categories and profiles learner populations.

Unlike reported practice in the Western education systems, for example the National Assessment of Educational Performance (NAEP) in the United States of America, in South Africa the increase in the frequency of conducting LSAS has not been accompanied by a corresponding increase in efforts to effectively use assessment information for improving learning (Kanjee & Sayed, 2013; Kanjee & Moloi, 2014) and the setting and use of performance standards has certainly not taken root (Pitoniak & Yeld, 2013). Instead, learner performance is generally reported according to prescribed levels of performance and in raw scores that do not specify the knowledge and skills that learners at each level should be able to demonstrate (DoE, undated). The limitations and possible misinterpretations attendant to reporting learner performance in single raw scores, such as averages, have been well documented (Crundwell, 2005).

Literature review from education systems that have a longer history of conducting LSAS reveals clear trends in the development of information-rich formats of reporting to enhance the meaningfulness, interpretability and utility of assessment results (Boulet, De Champlain & McKinley, 2003; Greaney & Kellaghan, 2008; Jaeger, 1995; Norcini, 1988; Swanson & Watson, 1982). Further, the literature also shows how the use of performance standards to report assessment results has evolved as have the challenges in using and validating different methods of standard setting (Jaeger, 1989; Kane, 2001; Cizek & Bunch, 2007; Zieky, Perie & Livingston, 2008). However, there have also been divergent views on the ‘best method’ of setting standards (Cizek & Bunch, 2007), on the best way to use different methods of standard setting, or to compare results and allow policy makers to make informed decisions on the final standards (Davis-Becker & Buckendahl, 2011) considering the validity of the different methods as well as the context in which the standards will be applied.

Pitoniak and Yeld (2013) reported serious constraints that the unique South African apartheid legacy placed on feasibility of desirable consensus in the standard setting process compared
to education systems like in the US. From their interactions with South African educators in initiatives to set performance standards, Pitoniak and Yeld (2013:23) made observations that may have limiting implications to criterion-referenced assessment and use of performance standards in South Africa:

“There are pervasive and lingering suspicions about assessment in South Africa. The society is highly sensitive to judgments, which are seen as a means of division and discrimination. While this is of course true in that all assessment aims to make judgments about performance, the country’s past was dominated by bias and prejudice, and this has engendered lasting suspicion”.

Pitoniak and Yeld (2013:27) expatiate on their observations of the attitudes of the South Africans of different racial groupings who participated in their apparently unsuccessful standard setting endeavours:

“The enduring injustice of unequal opportunities to learn has made educators feel that failing learners (or categorizing them as Basic or Intermediate) is simply a case of blaming the victims and further disadvantaging them. There are fears that “educational underpreparedness” will be conflated with “lack of intelligence” and further confirm the stereotype, so assiduously cultivated by apartheid, that Black people are less capable of higher order thinking than others”.

While the situation may not be as dire as Pitoniak and Yeld (2013) report, the fact is that the need for inclusivity and representativity in diverse societies like South Africa is arguably more acute than ordinarily recommended in research literature which includes studies by Hambleton (2000) and Skorupski & Hambleton (2005). While the study by Pitoniak and Yeld (2013) attempted to set standards for university entry candidates, attempts to set performance standards for the basic education sector in South Africa have been few (Kanjee & Moloi, 2012) and there is no evidence that their recommendations to implement standard setting have been heeded. Yet, by the admission of Pitoniak and Yeld (2013:23), the introduction of Annual National Assessments in South Africa ushered an era with good prospects for promoting criterion-referenced modes of assessment and standard setting towards enhancing the use of assessment data to improve teaching and learning in schools.
2.5.4 Approaches and methodologies to standard setting

Strikingly, the magnitude of available research on setting and use of performance standards in education stands in direct contradiction of the fact that standard setting has been around for hardly fifty years. Seminal work on setting performance standards began around the sixties and the earliest initiatives are commonly associated with the pioneering works of Nedelsky (1954), Angoff (1971) and Ebel (1972). Since then, there has been a phenomenal two-pronged growth in research into standard setting in education.

On the one hand, there has been increasing research into comparing different methods of standard setting for efficacy and accuracy and providing guidelines on how to ensure that standards are valid and defensible, especially in systems where stakes in assessment are high and application of questionable standards may lead to costly legal suits (Cizek, 1996; Hambleton, 2000; Zieky & Perie, 2006).

On the other hand, there seems to be an emergence of critics of not only the defensibility of the (traditional) SS methods but also of the theoretical and philosophical assumptions that underpin the processes from among proponents of the latent-theory model, particularly the Rasch model, and related models of measurement in the social sciences (Stone, 1995, MacCann & Stanley, 2006).

There have been laudable initiatives to develop what so far seem to be ‘crude taxonomies’ for organising the increasing volume of research literature on standard setting (Hambleton, 2000; Morgan & Perie, 2004). As would be expected, there are differing views as to whether it is possible to have water-tight groupings of standard setting methods (Cizek & Bunch, 2007) but two main classifications seem to be popular, viz. “examinee-based” and “test-based” methods. Arguably, the classification could be a matter of focus rather than real distinctions. Examinee-based methods focus on assessing an individual on the basis of work they have completed while test-centred methods assess the individual’s performance on each item in a test (De Champion, 2004), hence there are strong views from some quarters that the classifications should rather be seen as “holistic” against “analytic” methods of standard setting (Cizek & Bunch, 2007:9).
What seems to come through is that the proliferation of standard setting methods seems to proceed in response to debates about the role to be played by panels of professionals or judges in standard setting and the statistical techniques to be employed in generating cut scores. Accordingly, three main streams of standard setting methods; the traditional Angoff method, Rasch-based methods, specifically the Objective Standard Setting (OSS) method and the Objective Borderline Method (OBM), were identified and summarised in this chapter. Because standard setting is a highly regimented process, for each method step-by-step procedures that must be followed have been provided in reasonable detail.

2.6 The traditional Angoff method

Although the Angoff method of standard setting has been modified over the years and has a wide range of variations, its essence remains the involvement of panels of selected professionals in estimating the competence of a “minimally competent” hypothetical borderline learner in each item in a given test (Hambleton, 2000; Näström & Nyström, 2008). Ideally the panel should be large enough to represent all possible views in the affected field or subject and thus minimise likely rater bias. Views on the ideal size of the panel vary from a minimum of 15 to as high as 45 (Hambleton, 2000) although numbers as low as six panellists have been used with negligible negative effects in other studies (MacCann & Stanley, 2006). Inherently this makes the Angoff method one of the more costly standard setting activities, given the relatively large number of panellist as well as the time required, a criticism that has been observed in all studies that employed the method.

2.6.1 Sequence of steps in the Angoff method

Individual ratings of minimally competent candidates get summed up for each panellist and finally averaged across panellists to obtain a cut score. Measures that are recommended to minimise inter- and intra-rater variances in the Angoff method have been presented in fair detail by Hambleton (2000), and include: (i) Intensive training of panellists before starting with ratings to ensure a common understanding of concepts and techniques; (ii) Structuring the rating process in rounds that are intercepted by discussions to maximise reflection by individual panellists and thus minimise variances in ratings, some of which may be attributed to ‘conceptual drift’ or loss of focus in the panellists’ minds on how a “minimally competent
candidate” is defined (Ricker, 2006); and, (iii) Presenting data that enables panellists to get a sense of what the impact of their cutscores could be if applied in the real ‘world’.

Criticisms against the Angoff method have been growing over the years. A number of reviews of the method have shown that the final cutscores or standards resulting from this method invariably are modifications that are arrived at more by consensus than by empirical methods (Ricker, 2006). Reviewers were of the view that this practice reduces the value of and calls into question the need to have an apparently rigorous method of ostensibly generating empirical data if eventually the final result is arrived at largely by sheer consensus. Consequently, the critics argue that the Angoff method fails to produce standards that are reproducible as evidenced by numerous modifications that seek “to correct largely unusable outcomes” of the method, (Stone 1995:447). However, Ricker (2006) notes that in spite of criticisms, the Angoff method continues to receive application in standard setting apparently because, unlike most methods, it does not involve complicated statistical procedures

2.6.2 Rasch-based Angoff SS methods

The use of the Item Response Theory\(^1\) (IRT) model in general, and the Rasch model in particular, in analysing assessment data has been gaining popularity in many countries of the world that conduct large-scale assessments or among social scientists who are interested in developing scientific measures for social contexts (Bond & Fox, 2007). In South Africa Dunne, Long, Craig and Venter (2012), using assessment in mathematics as an example, reported a fairly detailed illustrative initiative on how the use of the Rasch model in analysing assessment data could help bridge apparent gaps between school-based and externally administered systemic assessments. Arguing for assessment data to be of good and quantitatively rigorous quality in order to be useful, Dunne et al (2012:13) aver that:

“*The Rasch model is essentially a single complex hypothesis built from several requirements about a context, about a test instrument and its constituent items, and about the way in which the context and instrument interact to produce special forms of measurement-like data.*”

\(^1\) Modern theory of testing which associates scores on test items with latent abilities of test takers.
But there have also been dissenting views on the suitability of the Rasch model, in particular, for analysing large-scale assessment data (Goldstein & Blinkhorn, 2008). Although the Rasch family of models has been assumed to be a subset of IRT models, there have again been dissenting views on this matter as well, hence it will not be pursued any further in this study; instead, focus will be on the Rasch family of models.

The simplest mathematical definition of the Rasch model (dichotomous model) estimates that, a person “n” of ability $\beta_n$, faced with a dichotomous (right or wrong) item “i” of difficulty $\delta_i$, has the probability $P_{ni}$ of giving a right answer ($x_{ni}=1$ and not 0), and that probability is represented by Equation 1:

$$P_{ni}(x_{ni}=1/\beta_n, \delta_i) = \frac{e^{(\beta_n - \delta_i)}}{1+e^{(\beta_n - \delta_i)}}$$

(1)

In Equation 1 the placeholder “e” is a constant whose value is 2.7183. It is the inclusion of this constant, 2.7183, that makes Rasch estimates (e.g. person abilities and item difficulties) true measures that can be added in such a way that adding “one more” of one quantity to another, results in “one more” of what was in the original quantity (Bond & Fox, 2007). For example, with raw scores (percent correct), a person who gets a score of 11% cannot be said to be “one more” percentage point able than the person who got a score of 10%. This difference in raw scores does not make this conclusion logical. With Rasch measures “one more” measure of ability means the person is “one more” able than the other person.

The use of the Rasch model in assessment has been gaining momentum since its inception in the work of the Danish psychometrician, Georg Rasch (1960). In simple terms, the essence of the Rasch model is in estimating the ability of a test taker and the difficulty of an item in a test in such a way that a probabilistic comparison can be made between these two, commonly referred to as ‘parameters’ – person ability and item difficulty. For instance, a person whose ability is equal to the difficulty of an item that they must respond to has a 50:50 probability of answering the item correctly. This is easy to see if, in Equation 1, $\beta_n = \delta_i$ then $\beta_n - \delta_i = 0$ and the probability becomes: $P_{ni}(x_{ni}=1/\beta_n, \delta_i) = \frac{e^0}{1+e^0} = \frac{1}{1+1} = 0.5$ or 50%

---

If the person ability is higher than the item difficulty, say the two exist in the ratio of 75:25, then the probability of answering the item correctly is higher. Conversely, if the person ability is much lower than the difficulty of the item, say they exist in the ratio 25:75, then the probability of the person answering the item correctly is lower. Rasch (1960:117) sums up the model as follows:

“A person having greater ability than another person should have the greater probability of solving any item of the type in question, and similarly, one item being more difficult than another means that for any person the probability of solving the second problem is the greater one.”

The use of Rasch measures, therefore, introduces robustness that enables users of the assessment results to be confident that a higher (Rasch-based) score indicates a learner who has a higher probability to perform better than another who obtains a lower Rasch-based score. In the use of performance standards that are Rasch-based, one can be more confident that a learner who is categorised at a higher performance level is likely (probabilistic) to have knowledge and skills that exceed those of a learner at a lower performance level provided the measurement is valid, i.e. for instance, a mathematics test is measuring mathematical knowledge and not the ability to read (Bond & Fox, 2007:268-270; Andrich & Marais, 2011).

Initial use of Rasch measurement in setting performance standards was purported to replace ratings of panellists with Rasch-estimated measures in the Angoff method (MacCann & Stanley, 2006). Rasch measures were found to be more stable (less variable) than rating scores allocated by panellists and this was viewed as a way of improving the quality (precision) of Angoff-generated cutscores. Subsequent initiatives led to the development of the Bookmark method of standard setting. In the Bookmark method the abilities of the test takers and the difficulty values of items (both measured in logits) that are at 50:50 probability are placed alongside each other in an increasing order with each item on a separate sheet or booklet, hence the name. The role of panellists in the Bookmark method is to go through the increasing order of measures in the booklets and identify an item that they think two-thirds, or any specified proportion, of borderline learners would answer correctly. The difficulty value of that item, often expressed as a percentage, becomes the cut score (MacCann & Stanley, 2006).


2.7 The OSS method

The Objective Standard Setting (OSS) is a test-centred Rasch-based SS method that radically differs from methods that are essentially designed to improve the traditional Angoff method. The OSS defines the standard expressly by getting subject specialists to identify “essential” content and skills that are required for a learner to be categorised as achieving the standard whilst also providing the necessary quantitative tools (Stone, 1995: 447). Using the scores obtained by learners on the qualitative “essential” items, interval measures are then created, using Rasch techniques, to convert the qualitative information into estimates of the underlying latent knowledge of the candidates.

The essence of the OSS method lies in the determination and use of Rasch measures, viz. person ability and item difficulty, that are considered to be “objective” because they are independent of the test that the person wrote and independent of which person wrote the test, respectively (Bond & Fox, 2007). The OSS method can, therefore, be used only where interval measures are involved. It cannot be used with raw scores, for instance, because of the limitations of raw scores that were discussed earlier in this research.

Stone (1995) argues that, while both the Angoff and OSS methods engage teams of specialists to represent possible variety of viewpoints in the standard setting process, the operationalization of the teams proceeds in different directions. On the one hand, the Angoff method averages the various viewpoints, apparently to minimise the error caused by variance or disagreement among the panellists. On the other hand, the OSS recognises error of measurement as ubiquitous and realistically builds it into the definition of the standard as shown in how the OSS final standard is calculated (Stone, 2001:192-193). The purpose and key concepts that underlie the OSS were embedded in the seminal work by Wright & Grosse (1993) and have subsequently been articulated by Stone (1995).

Two key concepts are important in the OSS: (i) Selectable pool which refers to all the items in the test as a representation of curriculum requirements; and, (ii) Essential items which refers to the sub-set of the “selectable pool” that comprises items which a learner must be able to answer correctly at a particular performance level. The key activity in the OSS method involves experts identifying ‘essential items’ that reasonably cover all the domains in the given test. This is the most important role played by experts in this method. The activity
Q. Moloi

directly addresses the critical ‘content validity’ that most methods that focus on “performance” tend to miss (Stone, 1995). Stone (2001:191) contends that identification of essential content by experts is an essential evaluative rather than a measurement process. The process and the activity serve to identify content and skills that must characterise a standard in a particular subject and grade.

For each panellist the average difficulty of the identified “essential items” is calculated and the average measure of difficulty of all “essential items” across the panellists constitutes what Stone (2001:191) refers to as the “Criterion Point” or the transition point from one performance level to the next. Stone (2001:191) motivates for the technique of including individual judge criterion points by emphasising that judge difference is important and must be maintained rather than attempting to reduce it to some kind of homogeneity.

Calculation of OSS cutscores includes two other important concepts, viz. determination of acceptable “Mastery Level” and an informed consensus on the magnitude of the “measurement error” that should be tolerated in the standard. “Mastery Level” refers to comprehensive competence, often expressed as a percentage, required to achieve the expected performance level. Determining the “Mastery Level” answers the question: How much of the expected or prescribed content should learners master to be considered competent at the specified performance level? For instance, the affected institution may require a score of 60%, 70%, 85%, etc. in a test as an indicator of content mastery (Stone, 2001:191). Determining the “Mastery Level” is a policy decision but the experts who participate in the standard setting process contribute, from their knowledge of the implementation of the curriculum, content-related inputs that help validate the decision.

Calculation of the measurement error is done in recognition of the ubiquitous presence of error in measurement (Stone, 2001:192). Whilst measurement error may not be significant in extreme scores, i.e. extreme high and low performers, because one can be almost certain that the true score is close to the observed score, the same cannot be said of non-extreme scores where one has to decide exactly where to place the cut score within the error band that surrounds the observed score. The level of precision that is required on the cutscores needs to be decided upon in a transparent discussion (Stone, 2001).
Ordinarily the process of determining the level of precision that is desired in the standard begins as a group activity to agree on the acceptable error band around the cut score (i.e. Confidence Level) (Stone, 2001:192). The group discusses and recommends an acceptable Confidence Level for the standard. Panellists must decide on whether the test is perfect and, therefore, the standard must protect the public by allowing a relatively narrow Confidence Level or high level of strictness or the test is prone to error so the standard must protect the individual by allowing a relatively wide Confidence Level or low level of strictness. The cut score at each performance level is then calculated as:

$$\text{Cut Score} = \text{"Criterion Point"} + \text{Mastery Level} \pm \text{Confidence Level}$$  

(2)

In Equation 2, all the variables must be expressed in logits and then the final answer can be converted to a percentage, if required.

An illustration of the steps and key concepts that are involved in calculating cut scores in the OSS method based on Equation 2 is presented in Table 1 from some made-up data.

### Table 1: Illustration of steps and key concepts in calculating OSS cut scores

<table>
<thead>
<tr>
<th>Steps</th>
<th>Key Concepts</th>
<th>Not Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Criterion Point</td>
<td>-0.9</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>2</td>
<td>Mastery Level</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>3</td>
<td>Confidence Interval</td>
<td>1.96</td>
<td>1.96</td>
<td>1.96</td>
</tr>
<tr>
<td>4</td>
<td>SEM</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>5</td>
<td>(Confidence Interval) x SEM</td>
<td>0.6</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>6</td>
<td>Final cut score ((1+2-5)) in logits</td>
<td>-0.6</td>
<td>0.3</td>
<td>0.6</td>
</tr>
<tr>
<td>7</td>
<td>Final cutscores ((%)^3)</td>
<td>35.5</td>
<td>57.4</td>
<td>64.7</td>
</tr>
</tbody>
</table>

Adapted from Stone (2001:191-195)

The first column in Table 1 lists the sequential steps that are followed in calculating the cut scores. The key concepts involved in each step are listed in the second column. In each of the third to fifth columns the values that are either the result from relevant calculations or are read off from relevant sources are shown for each performance level.

---

\(^3\) Logits converted to percent = “100/(1+exponent(-logits))”, derived from conversion of percent to logits which is: Percent to logits =ln(percent/(100-percent)), (Bond & Fox, 2007:24-26).
In the specific example shown in Table 1 the “Criterion Point” in Step 1 is the mean difficulty of “essential items” (in logits) for the relevant performance level; e.g. the “Criterion Point” at the “Achieved Level” is 0.1 logits. The “Mastery Level” in Step 2 was taken as 70% (0.85 logits) and for the same test the same value applies across all the performance levels. In Step 3 the “Confidence Interval” was taken as 95% or 1.96 of the standard deviation, also expressed in logits. The “SEM” in Step 4 is the Standard Error of Measurement of the test and its value was read directly from the output in Table 3.1 of the Winsteps Programme, where it is reflected as the “Model RMSE” and defined as the “statistical ‘root-mean-square’ average of the standard errors” (Linacre, 2014:307). For the same test, the same value of SEM applies across all the performance levels. In Step 7 the final score is calculated from Equation 2.

The “Confidence Level” in Equation 2 is calculated as the product of the “Confidence Interval” (Step 3) x “SEM” (Step 4). In high stakes examinations the value of the “Confidence Level” shown in Step 5 is usually added in the calculation of the cut score in Equation 2 but for low stakes assessments such as the ANAs the value is often subtracted as shown in Step 6. The resulting cut score of 0.3 logits in Step 7 is then expressed as a percentage (57.4%) for the “Achieved Level” in Table 1.

It is important to note that where more than one cut score is to be calculated, the steps that are listed in Table 1 are repeated separately to get each cut score related to the specific performance level. For instance, in this study three cutscores were calculated, one at the “Not Achieved”, the second at the “Partly Achieved” and the third at the “Achieved” levels, respectively.

2.8 The OBM

The Objective Borderline Method (OBM) is ascribed to work done by Shulruf, Poole, Jones and Wilkinson (2014). Unlike the Angoff and OSS methods, the OBM is a purely statistical technique. As the name indicates, the OBM of standard setting focuses on enhancing the precision of cutscores for borderline learners. It is a mathematical and probabilistic technique, mainly focusing on ensuring precision of cutscores, especially where stakes are high and the risks of making incorrect judgements must be reduced to the barest minimum (Shulruf et al 2014:16). The theoretical framework of OBM, like the IRT, is based on the recognition that
there is need to distinguish between learner’s true latent abilities and the observed raw scores that they achieve in a test.

The raw score only approximates the true ability score and is always associated with some measurement error that must be accounted for in the process of setting cut-scores. The relationship between the test score and the true ability score can be represented by the equation:

\[
S (\text{raw score}) = T (\text{true score}) + \epsilon (\text{measurement error})
\] (3)

The OBM addresses the probability that, due to the ubiquitous error (\(\epsilon\)) around any raw score, learners in the borderline zone, i.e. between clear passes and clear failures, could be erroneously mis-classified as either failures or passes. For instance, if a policy specifies a cut score of 50% for making judgements on whether learners pass or fail an examination, there is a probability that learners whose scores lie in the neighbourhood of the 50% cut score could be made to fail when they actually were supposed to be awarded a pass, or be made to pass when they actually were supposed to fail. In other words, as a result of the existence of an error of measurement around any true ability scores, there will always be learners whose scores fall within a “Range of uncertainty” that includes true borderline cases (b), cases that are potential failures but overlap the borderline zone (f) and cases that are potential passes but overlap the borderline zone (p). The distribution of candidates’ scores in any test or examination is illustrated in Figure 1

![Figure 1: Distribution of ability scores in a test](Adapted from Shulruf et al, 2014:5)

On the extreme end of Figure 1 are scores of candidates who are certainly incompetent and include candidates who score the minimum score in the test (left) and candidates who are certainly competent and include candidates who score the maximum score in the test (right). Candidates whose scores fall within the “Certainly incompetent”, and the “Certainly competent” zones, clear failures and clear passes, respectively, are not affected in determining the cut score using the OBM. Only candidates whose scores fall within the
“Range of uncertainty” zone, potential failures \((f)\), borderline cases \((b)\) and potential passes \((p)\), are implicated. The width of the “Range of uncertainty” depends on the range of error (RE) that is permitted. For instance, if the cut score is assumed to be 50% and RE is assumed to be 10%, then the “Range of uncertainty” will stretch from 40% (50%−10%) and will be bounded by 60% (50%+10%) on the higher end. Although the OBM is most applicable in high-stakes assessments where mis-classification of candidates to either side of the cut score could have serious implications, it is a useful technique for validating cutscores where two SS methods result in different cutscores to help determine the most precise value.

Shulruf’s \textit{et al} (2014) technique involves identifying numbers of learners who are potential passes \((p)\), according to the cut score that is prescribed in the policy or agreed upon by examiners, number of those who are potential failures \((f)\) and the number of those who are borderline cases with the probability of being either passes or failures \((b)\). From Figure 1 it is evident that the candidates whose scores lie in the potential pass zone \((p)\) have no probability of being clear failures but do have a probability of being in the borderline zone. Similarly, candidates whose scores lie in the potential fail zone \((f)\), have no probability of being clear passes but do have a probability of “not failing” if the fall within the borderline zone. Using the statistical definition of the probabilities of independent events, Shulruf (2014) derives what he calls the “Pass Index” such that:

\[
\text{Pass Index} = \frac{b}{b+p} \times \frac{p}{p+f}
\]  

Shulruf (2014:7) describes the Pass Index as “… a close approximation to the proportion of learners with a borderline score who are competent”. By multiplying the Pass Index by the number of borderline learners \((b)\) one calculates the number of border line learners who qualify to be classified as passes. The minimum test score that is associated with the number of borderline learners who can be awarded a pass is the cut score.

An example to demonstrate how the OBM steps are followed has been summarised in Figure 2. In a hypothetical sample of 100, candidates’ scores range from a minimum of 0 to a maximum of 100. The cut score that is prescribed either by policy or as an agreement among examiners is 50. But this score is surrounded by an error of measurement as was shown in Equation 4. To avoid the errors of imprecision leading to either failing candidates whose true
ability scores qualify them as passes or passing candidates whose true ability scores qualify them as failures, the OBM is applied.

Firstly, an error range (RE) must be assumed around the prescribed cut score of 50, say RE=10, so that the “Range of uncertainty” stretches from 40 to 60. Secondly, all scores below 40 and above 60, clear failures and clear passes, respectively, will not be relevant for the determination of a true cut score. In this example, of the 100 candidates who wrote the examination, 68 had scores in the “clear failures” and “clear passes” zones and were considered not relevant in calculating the true cut score. Thirty-two candidates had their ability scores in the “Range of uncertainty” zone. Their scores and score frequencies have been shown in Figure 2.

![Figure 2: Example of calculating a cut score in OBM](Adapted from Shulruf et al., 2014:15)

From Figure 2 the candidates’ scores in the “Range of uncertainty” ranged from 42 to 57. Six ($f=6$) of the scores fell in the “potential fail” zone, thirteen ($b=13$) fell in the borderline zone and fifteen ($p=15$) fell in the “potential pass” zone. For each score the frequency or number of candidates who obtained the score is shown in brackets. For example, two candidates obtained a score of 44 in the “potential fail” zone and four obtained a score of 52 in the borderline zone.

The Pass Index is calculated as follows:

$$\text{Pass Index} = \frac{b}{(b+f)} \times \frac{p}{(b+p)} = \frac{13}{(13+6)} \times \frac{15}{(13+15)} = \frac{13}{19} \times \frac{15}{28} = 0.37.$$ 

The proportion (0.37) is used to calculate the number of borderline cases who qualify to be awarded a pass. Thus $0.37 \times 13 = 4.8$ OR 37% of 13 = 4.8. Round off to the nearest whole...
number = 5. Then identify the Top 5 borderline cases and the scores that they obtained. From Figure 2, the highest score in the borderline zone was 54 and it was obtained by four candidates. The next score lower than 54 was 51 also obtained by four candidates. So the Top 5 scores are four 54s and one 51. The true cut score is the lowest of the Top 5 which is 51, slightly higher than the 50 that was prescribed in policy.

2.9 Historical and current practices of LSAS in South Africa

Similar to the education systems reported in the literature review, the current design and utilisation of LSAS in South Africa has evolved and undergone major changes since 1998. Largely, the changes have tended to coincide with changes in political leadership that accompany national election cycles. Over a period of approximately fifteen years between 1995 and 2010 interspersed by three five-year election and administration cycles, LSAS in South Africa have undergone three major shifts.

2.10 Systemic evaluations of 1998 - 2003

During the first cycle South Africa developed and adopted periodic sample-based “systemic evaluations”, designed to be conducted on representative samples of schools and learners at Grades 3, 6 and 9 as a measure of monitoring how the post-apartheid national education system was achieving the goals of access, equity and quality (DoE, 1998).

The then DoE developed a “Framework for Systemic Evaluation”, unpublished, (DoE, 2003) which specified, among others, that systemic evaluation was to give effect to the mandate of the Minister of Education to monitor and evaluate the levels and quality of provision, delivery and performance of education standards across all the provinces. The objectives of systemic evaluation included determining “levels of learner achievement” and drawing conclusions about suitable education interventions, while additional information was provided regarding the indicators of access, equity and quality to be monitored, and the frequency of conducting the assessments.

These assessments involved testing learners in literacy and numeracy and administering self-completed questionnaires on school principals, teachers, learners and parents to collect contextual information on the conditions that might impact on teaching and learning in the schools (DoE, 2003b).
2.11 Systemic evaluations in nodal zones in 2004 - 2007

By 2007, although the concept of systemic evaluations was still retained, the practice of conducting the evaluations had taken other dimensions. In addition to participating in national systemic evaluations, individual provinces conducted local versions of assessments, also called “systemic evaluations”. The national DoE also started a practice of conducting assessments in nodal zones which were politically designated geographical areas targeted for accelerated socio-economic redress. In these areas “systemic evaluation” was used to establish baselines and then track learner performance over time to monitor if focused interventions resulted in change in learner performance (Centre for Evaluation and Assessment, 2004).

2.12 Foundations for Learning Campaign and ANA in 2008 - 2009

In 2008, following the release of the results of the second systemic evaluation at the Grade 3 level which showed very insignificant changes in learner raw scores in literacy and numeracy, the then DoE (DoE, 2008) introduced the Foundations for Learning Campaign (FFL) as a redress initiative which spelt out a detailed intervention strategy to improve teaching, school management and resourcing of schools on the lower end of the poverty continuum (poverty Quintiles 1-3).

Part of the strategy was to suspend systemic evaluations and participation in international studies for five years until 2011 and focus on internal improvement of the conditions of teaching and learning. To monitor if these interventions were making any impact on learner performance, the DoE mandated the administration of annual national assessments in literacy and numeracy on all learners in Grades 3, 6 and 9. No contextual information was collected against which to interpret the results of learner performance in ANA (DoE, 2008).

2.13 ANA from 2010 onwards

In 2009 the elected Government announced an outcomes-based approach to delivery of services and identified improvement of the quality of basic education as its top priority (South Africa, Presidency, 2009). To give effect to the Government priority, the DBE, which had taken over the function of the DoE at the basic education level, identified annual national assessments (ANA) as the key mechanism for monitoring the achievement of set educational
targets in terms of learner performance. Thus ANA would now be administered on all Grade 1 – 6 and Grade 9 learners in all the approximately 23 000 schools in South Africa in what was to be known as “Universal ANA” to emphasise participation by all learners in all schools. The first ANA cycle was conducted in February 2011 (DBE, 2011b) and the assessments have been conducted every year since then (DBE, 2014). Grades 7 and 8 were planned to be included in the assessment with effect from 2014.

The implementation of the Universal ANA involved both the DBE and provinces in various administrative responsibilities (DBE, 2013a:17-28). The DBE, utilising appropriate professional and technical expertise, set the tests, printed and distributed them to schools. Following a national common timetable, principals of schools assigned teachers who did not teach the affected grades to invigilate or monitor the administration of tests in the schools. The DBE, provinces and districts deployed officials to monitor the test administration process overall but also randomly selecting schools for direct observation of how they managed the process.

Universal ANA tests were marked by teachers who taught the affected grades at school level and the results were reported to parents. Schools were expected to use the results to inform and develop appropriate intervention strategies to improve teaching and learning. To guide schools on how to interpret and utilise ANA results to develop appropriate intervention strategies for improvement, the DBE compiled guidelines that were supposed to be distributed to all schools after each cycle of assessment (DBE, 2011d, 2012d, 2013b).

The data from all the schools was uploaded onto a national electronic database. From the database the DBE analysed the data and produced a national report that was disaggregated to the nine provinces and was used to inform the public about the state of achievement in basic education in the country (DBE, 2011b, 2012c, 2013a).

Alongside Universal ANA, a representative sample of schools with learners in grades 3, 6 and 9 were selected and monitored by an independent agency (DBE, 2013a:26-28). The independent agency deployed fieldworkers in all the sampled schools. Their responsibility was to ensure that tests were administered under controlled conditions and that the schools did not interfere with the independence of learners in responding to test questions. At the end of the test administration the fieldworkers selected representative samples of learner scripts.
from grades 3, 6 and 9 in each of the two subjects and removed these for independent marking by teachers who were appointed and supervised by the independent agency. The independent agency then analysed the data and reported learner performance from the independent verification process.

A number of key observations can be made from the evolution that LSAS have undergone in South Africa. First, LSAS have changed purpose, frequency and design over the years, mainly driven by political changes. Second, besides the original “Framework for Systemic Evaluation” which has ceased to operate, the changes have not been led or accompanied by corresponding developments in a national framework pertaining to the administration nor to the reporting and use of results. A common feature of all the phases in the changes has been reporting learner performance in “raw scores” based on a scale that categorises learners according to the total score they obtain in the test. Over this period, however, the reporting scale has changed from a four-point scale (DoE, 1996) to a seven-point scale (DBE, 2011c) (discussed in the next section).

In the evolution of national assessments in South Africa, the absence of a national framework that regulates the implementation of the assessments, and guides the reporting and utilisation of results left room for each phase of the evolution to appear disjointed from the previous phases. Partly this could account for the apparently arbitrary changes to the purpose of the LSAS as has been reported in this review. A national framework, adopted and supported at all levels of the education system, will not only stabilise the use of LSAS in general, and the ANAs in particular, but will also specify what to measure, and when, how to measure and utilise the assessment results to improve teaching in schools.

2.14 Use of various reporting formats

With the introduction of the Outcomes-Based Education (OBE) curriculum in 1997 (DoE, 2009:12), the DoE also introduced a format of reporting learner performance on a four-level scale with each level defined by a range of raw score percentages (DoE, 1998). The four-scale reporting structure, the range of raw scores spanning each level and the labels of the levels have been presented in Table 2.

Table 2: Levels and score ranges in the DoE reporting framework
Following the Curriculum Review of 2009 (DoE, 2009) and the implementation of the Review recommendations, the DBE continues to report assessment results in raw scores that are defined in the scale that is provided in the CAPS. In the revised scale the levels or performance have been increased from four (4) to seven (7) with corresponding revision of the score ranges and labels for each level. The revised levels of performance, their descriptions or labels and score ranges, also referred to as “Rating Codes” in the CAPS document (DBE, 2011c), have been presented in Table 3.

**Table 3: DBE performance levels**

<table>
<thead>
<tr>
<th>Level</th>
<th>Level description</th>
<th>Score range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Outstanding Achievement</td>
<td>80–100</td>
</tr>
<tr>
<td>6</td>
<td>Meritorious Achievement</td>
<td>70–79</td>
</tr>
<tr>
<td>5</td>
<td>Substantial Achievement</td>
<td>60–69</td>
</tr>
<tr>
<td>4</td>
<td>Adequate achievement</td>
<td>50–59</td>
</tr>
<tr>
<td>3</td>
<td>Moderate achievement</td>
<td>40–49</td>
</tr>
<tr>
<td>2</td>
<td>Elementary achievement</td>
<td>30–39</td>
</tr>
<tr>
<td>1</td>
<td>Not achieved</td>
<td>0–29</td>
</tr>
</tbody>
</table>

(Source: DBE, 2012c)

From Table 3 one observes that there were minimal changes in the manner in which assessment results are reported since the curriculum review that was instituted by the Minister in 2009 (DoE, 2009). A visible change was the increase of the performance levels from four (4) to seven (7). One criticism against this re-arrangement is that, with the increased number of performance levels, reported performance at each level may suffer decreased accuracy since, from an ordinary test of a one-hour duration, the test items would either be clustered exclusively in fewer levels or would be stretched too thinly across all the levels. The other criticism is that both the four- and the seven-level scales suffer the same weakness: neither scale enables detailed reporting on what learners at each performance level can or cannot do nor the knowledge that they command or do not have.
Although there may be generic features, the format and content of reporting learner performance vary according to the purpose of the report, the target audience and time of reporting (Queensland Government, 2008:8). The purposes of ANA, the target audiences and the types of envisaged reports have been mentioned in the education sector *Action Plan* (DBE, 2011a:3). The overall purpose of the assessment has been specified as measuring learner achievement levels in literacy (language) and numeracy (mathematics), establish benchmarks and make comparisons of performance across schools and districts. The target audiences and potential users of the assessment data include have been identified to be, among others, national, provincial and district staff, entire schools, teachers, learners and parents. Each audience is expected to receive a relevant report, interpret the results and take appropriate action in response.

However, at least in two consecutive years of the implementation of ANA (DBE, 2012b & 2013a) the reports that were published contained information at national, provincial and district levels. The results were presented in tables and graphs with explanatory texts. Extracts from the mathematics reports at national level, tabulated data and a bar graph, have been shown in Table 4 and Figure 3, respectively.

**Table 4: ANA 2013 national results**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Average percentage - Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2012</td>
</tr>
<tr>
<td>1</td>
<td>68</td>
</tr>
<tr>
<td>2</td>
<td>57</td>
</tr>
<tr>
<td>3</td>
<td>41</td>
</tr>
<tr>
<td>4</td>
<td>37</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>9</td>
<td>13</td>
</tr>
</tbody>
</table>

(Source: DBE, 2013a:2013:31)
From both Table 4 and Figure 3 it is evident that limited information is communicated in these reports that can be of use to teachers or schools for addressing specific learner weaknesses and strengths. The average percentage points per grade in Table 4 do not convey any information about what learners can or cannot do, what they know or do not know. Similarly, the seven levels into which learners have been categorised from “Not achieved” through “Outstanding achievement” in Figure 6 are mute in terms of what is implied by these labels, i.e. what knowledge and skills were demonstrated or not demonstrated by learners categorised into anyone of these levels. In particular, teachers would find very little, if any, information to help them carry out their core mandate of supporting learners improve learning.

In arguing for reporting formats and contents of school reports that are meaningful, Mead (2009), cited in Smith and Stone, (2009:489) notes that: “In educational testing, no assessment is worth the instructional time lost if the results are not communicated back to the school, the teacher, the parents, and the learner in a manner that can improve the educational outcome for that learner and that school”. In a survey that involved a mix of district officials, lead teachers and ordinary teachers and parents, Taut et al (2009:129-137) evaluated the extent to which results from Chilean national assessments were accessible, easy to understand and use by the participants in their various roles regarding schools and schooling in Chile.
their findings they reported that interpretation of basic information from the assessment reports by the majority of teachers and parent was incorrect.

Meanwhile there have been significant developments and innovations in improving the meaning and quality of information in LSAS, mainly linked to the growing use of IRT in general and the Rasch model in particular. By contrasting some of these innovative reporting formats with the formats used in South Africa it becomes evident that a new approach to reporting results of LSAS in South Africa could be considered. An example of innovative reporting formats, how they compare to South African report formats and their potential to lead to improved teaching were investigated from other studies including SACMEQ.

**Figure 4: SACMEQ levels for grade 6 mathematics**

(Source: Moloi & Chetty, 2007)

Figure 4 shows distribution of grade 6 learners in a SACMEQ mathematics test according to levels of mathematical skills and knowledge that they demonstrated in the test. This reporting is structured to reflect:

a. A scale score that the learner obtained in the test;
b. The level associated with the scale score that the learner obtained in the test. Each level is described appropriately to reflect a summary of the mathematics content knowledge that the learner displays according to the set of items that they answered correctly in the test. For instance, Level 2 is the “Emergent Numeracy” level;

c. Labels are then followed by “level descriptors” which describe in words what the learner at this level is expected to know or to be able to do. For instance, a learner at the Emergent Numeracy level is expected to be able to, among others, apply two-step mathematical operations that involve either addition or subtraction and ‘carrying’; and,

d. Examples of items which, commensurate to their cognitive demands, are pegged at the given level.

Extracts of examples of items at Level 2 and Level 5 are shown in Figure 5 and Figure 6, respectively.

Emergent numeracy (Level 2) skills:
Applies a two-step addition or subtraction operation involving carrying.

<table>
<thead>
<tr>
<th>Subtract …</th>
</tr>
</thead>
<tbody>
<tr>
<td>6,000</td>
</tr>
<tr>
<td>−2,369</td>
</tr>
<tr>
<td>A. 3,531</td>
</tr>
<tr>
<td>B. 3,631</td>
</tr>
<tr>
<td>C. 3,742</td>
</tr>
<tr>
<td>D. 4,369</td>
</tr>
</tbody>
</table>

Figure 5: Level 2-emergent numeracy item
Level 5: Competent numerate

45. Use the price list to find the total cost of 5 litres of milk, 3 kg of tomatoes and 2 chickens?

<table>
<thead>
<tr>
<th>Item</th>
<th>Unit Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk</td>
<td>1 litre R0.60</td>
</tr>
<tr>
<td>Tomatoes</td>
<td>1 kg R0.50</td>
</tr>
<tr>
<td>Chicken</td>
<td>each R6.70</td>
</tr>
</tbody>
</table>

A. R 7.80  
B. R 16.90
C. R 17.90
D. R 18.90

**Figure 6: Level 5- competent numerate**

The two examples in Figure 5 and Figure 6, respectively, demonstrate the distinct levels of cognitive demand at the two levels. The example in Figure 5, the emergent numeracy level item, demands two steps to complete whereas in Figure 6, the competent numeracy level, it takes four steps to complete the answer correctly.

Mead (2009), cited in Smith and Stone (2009:489-512) proposed a reporting format that also uses Rasch-based scale scores but presents information in visual as well as text forms. A highly modified format of a report based on Mead’s work has been shown in Figure 7.
In Figure 7 an example of a report based on the use of Scale Scores is presented. Besides reporting on Scale Scores, other key features of this report include:

- The learner is compared with him-her-self and also with others;
- He functions around the 80\textsuperscript{th} percentile, meaning that 80\% of his class operate at levels lower than him;
- The report shows his/her position in terms of the performance standards, e.g. he is currently functioning on the upper end of the Achieved Level; and,
- An example of the skills and content that he/she demonstrates “comfortably” at this level is given (Numbers – multiplying and dividing mixed numbers).
Q. Moloi

His challenge is to acquire necessary skills to tackle problems of probability and solving geometric problems. To do that, his overall Scale Score will have to increase for the current range between 180 and 210 between 276 and 278. To know the big picture of where he is and what he needs to do, the learner will be referred to the performance level descriptors. This report does not only inform the learner. It also informs the teacher (necessary interventions) and the parent (necessary support). This kind of report can be produced for a district as well where the place of the candidate is taken by a school. The district official knows where each school is in relation to others, but also where it is in terms of specific content knowledge of its learners. The SACMEQ reporting format is also based on principles of Rasch measurement. It also identifies levels of performance with specific descriptions of skills and knowledge that a learner at a given level is expected to be able to demonstrate.

A notable feature of the DBE reporting scale (DoE, 2012c) is that the categories seem to be arbitrary when compared to standard-based categories that are defined in terms of PLDs. Only a total raw score is used to place a learner in a particular category or level. Furthermore, raw scores are sample- and test-dependent. On the one hand, on the same test, a sample of bright learners will achieve higher raw scores and levels than their less-gifted counterparts. On the other hand, the same sample of learners will achieve lower scores on a more difficult test than on a relatively easier test. The use of raw scores to report results of LSAS has been found to be unreliable, hence the increasing use of more stable performance standards (Glass & Hopkins, 1984).

To investigate whether using raw scores and using performance standards – developed through a rigorous procedure - to report the results of LSAS gave similar results, Kanjee and Moloi (2012) used the two methods to analyse and report learner performance from the Verification ANA of 2011b. The results of the investigation showed that, in both literacy and numeracy and at both Grade 3 and Grade 6 levels, compared to cutscores from the standard setting process, the cutscores from the DBE scale consistently tended to over-estimate learners performing at the “Not Achieved” level and to under-estimate learners performing at the “Partially Achieved” level.

These findings can have serious implications. Interventions may be directed to either where they are not needed or to where they will make the least impact over a long time. For
example, if more learners are mis-diagnosed to be at the “Not Achieved” level than at the “Partially Achieved” level and resources are subsequently deployed accordingly, it may cost unnecessarily more resources over a longer time to redress the problems, whereas it would be relatively easier and less costly to redress problems of learners who are already performing at the “Partially Achieved” level.

Bond and Fox (2007), supported by Wright (1992), Magno (2009), and Sick (2008), argue that raw scores are quantifications of correct responses, and not measures in the sense that measurement is conceived of and communicated in physics, for instance. The values of raw scores depend on the difficulty of the test and the abilities of the test takers. On an easy test, percentage correct raw scores tend to be higher than in a difficult test while in the same test, learners of higher ability score higher than their counterparts of lower ability (Bond & Fox, 2007). The inaccuracies that are inherent in the use of raw scores are, therefore, likely to send conflicting messages, especially when tests are intended to track performance of different samples of learners over time. Secondly, raw scores (on their own) communicate very little, if any, specific content knowledge or skills which learners are able or not able to demonstrate in a test. Thirdly, the use of raw scores is predicated on the assumption that the effort of improving one’s score on the lower end of the ability continuum is the same on the upper end. Yet, studies based on the latent-theory model show that it is more difficult to improve from a score of, say, 90% to 95% than from 20% to 25% even though the interval (5%) is the same in both cases (Bond & Fox, 2007).

In view of the identified mismatches between the results achieved through a standard setting process and the results achieved using the DBE raw score scale, use of the standard setting procedures appears to give more accurate results, in addition to providing information on what learners know/don’t know and can/cannot do (Kanjee & Moloi, 2012). The procedure has a theoretically sound basis, its procedures are well-documented and its results command wide respect.

2.15 Frameworks for data-driven decision-making

The literature on data-driven decision-making was to determine whether South Africa’s initiatives to entrench a culture of assessment and evidence-based monitoring of educational outcomes were happening in isolation or there were lessons to learn from elsewhere. The
review showed that the last decade of the 20th century saw phenomenal increase in research into assessment-data-driven decision-making in education systems across the world (Hamilton, 2003; Marsh, 2012). In the United States, this change was spurred by the introduction of legislation that compelled states to use standards-based monitoring to ensure that all learners did not only gain access into education but also achieved proficient levels of performance (Mandinach, Honey & Light, 2006:2). The consequence was that teachers and district officials found themselves faced with the responsibility of having to generate, analyse and report on assessment data – a task for which they were not equipped for, nor did they have the requisite resources (Mandinach, Honey & Light, 2006:15).

Large-scale assessments are conducted to answer identified questions, policy concerns and questions relating to learner achievement in the affected countries. Given their magnitude and resource-intensity, one could argue that these studies are never conducted in a void. A theory of action, implicit or explicit, informs the actions and steps that are taken. One example of a comprehensive theoretical framework that drives a large-scale assessment was developed by SACMEQ (Makuwa, 2009) and provides a model that is predicated on the essence of feedback from the assessments and how the feedback should influence data-driven decision-making, at the least at the policy level. An iconic representation of the SACMEQ Framework for conducting studies that evaluate conditions of learning and teaching in the SACMEQ network has been shown in Figure 8.
From Figure 8, it can be observed that SACMEQ works on a cyclic framework with three major phases, viz.: 1) policy research phase in which policy concerns of member countries are identified, relevant data to answer policy-related questions is collected and analysed, including answering questions about learner achievement; 2) policy development phase in which research and assessment data is fed back to the affected systems with policy suggestions in the form of policy briefs, national reports, conferences and workshops; and, 3) the policy evaluation phase in which the participating countries are expected to implement what was learnt in the previous assessment.

Although the SACMEQ framework provides a model that can be adapted, the argument in this study is that the framework is crafted in very broad terms, understandably because it is meant for countries comparisons in very diverse contexts. For instance, the framework may not apply adequately well in the Seychelles, a country with less than thirty primary schools, and in South Africa, a system that operates in more than 27 000 schools, but both countries are members of SACMEQ. But also, unlike ANA, SACMEQ has turnaround times of up to four years. This study maintains that South Africa needs an NRF that is focused on and is...
responsive to her needs to guide effective utilisation of LSAS data to improve the quality of teaching and learning in schools and thus enable the system to improve performance.

Reflecting on developments in the area of using data to make decisions in education districts, Breiter and Light (2006:207) noted that most of the research they had reviewed was largely on case studies and lacked a theoretical framework to back the initiatives up. Consequently, Breiter and Light (2006:208) have proposed a theoretical model or framework for data-driven decision-making that they argue has since been adopted widely in relevant subsequent research. 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 (2006:208). In foregrounding their proposal, Breiter and Light (2006:208) refute earlier misconceptions that “decision-making” requires one to work with innumerable pieces of data and suggests 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 synthesizing). A modified pictorial representation of Breiter’s and Light’s conceptual framework of data-driven decision-making has been shown in Figure 9.

![Figure 9: Conceptual Model of data-driven decision-making](image)

(Adapted from Breiter & Light, 2006)

Figure 9 shows that data-driven decision-making has three main elements: (i) data which includes unprocessed statistical inputs; (ii) information which results from an analysis and
summaries of the processed data; and, (iii) knowledge (information that is steeped in a particular context) and processes of systematically transforming the data into decisions. The model assumes a feedback loop that links implementation and impact back to data and begins the cycle again. It is important to note that Breiter and Light (2006) argue that decision-making does not begin with data but with knowledge of the needs of learners, teachers or even the officials. It is this 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.

In a separate study, Ronka, Geier and Marciniak (2010) conducted case studies of education districts that were committed to using data to inform decisions and actions. They subsequently developed a ‘theory of action’ that is underpinned by three major features that characterise effective data use that leads to improved learner achievement, viz. quality of data as a condition for effective use, activities that are driven through data and the results of effective data use (Ronka et al, 2010:2-3). According to these authors, the ‘data use theory of action’ that they advocate operates optimally in an educational environment where there are on-going and conscious efforts to sustain data use, data use is systemically adopted at all levels of the organisation and the focus of activities is on learner-centred data (Ronka et al, 2010:2-3).

The first feature of Ronka et al’s (2010:2-3) ‘theory of action’ is necessary “conditions for (effective) data use” which they specified as the quality of data, the capacity of practitioners to use the data and a pervasive culture for data use in the organisation as a whole. Marsh (2012) also maintains that necessary conditions for effective data use in educational circles are the characteristics and quality of the data that forms the basis for the ensuing actions and interventions. According to the reviewed case studies, data of a good quality must, among others, be standardised, presented in ways or formats that make it easy to understand and must be available at time of need (Ronka et al, 2010:3). Breiter & Light (2006:208) share similar views. Similarly, in a report from focus groups of various users of assessment reports, Ryan (2003) submitted details of the quality features and formats that the participants wanted to see in score reports if the reports were to be considered useful. Specific examples of data quality that users preferred included, among others, grouping data in meaningful ways, using
both numeric and graphic formats and avoiding statistical jargon in reports (Ryan, 2003:21-25).

Some of the observations that Ryan (2003:4-6) made from schools that created, maintained and consistently adhered to the necessary conditions for data use were that teachers progressively espoused an enquiry-based approach on what kinds of teaching led to best learning. Schools were able to identify learners, early enough, who were at risk of dropping out of school because they were not coping in specific areas of the curriculum. Further, teachers learnt to collaborate in supporting one another for mutual development in data use. Instructional leadership gained prominence among school principals and they grew to be more resourceful in school collaborations as well as in communicating with parents about the performance of their children.

Operating from the premise that building capacity for data use is intrinsically a social enterprise (Halverson, Grigg, Prichett & Thomas, 2007), Farley-Ripple and Buttram (2014) explored specific ways in which capacity for data use is developed in schools. They also espoused a ‘theory of action’ which holds that data use does improve instruction on the proviso that educators have the necessary capacity (Farley-Ripple & Buttram, 2014:1). Taking cognisance of the inconclusive research findings on exactly which of the multiplicity of strategies for data use (Marsh, 2012) influences educators’ skills, Farley-Ripple and Buttram’s (2014) findings pointed to the centrality of teacher social networks towards building effective capacity for data use. Not only did teachers form professional networks that influenced their capacity for data use, but the analysis conducted by Farley-Ripple and Buttram (2014: 36) also showed that “… educators are more likely to seek advice from those with formal instructional leadership positions, which included the … principal” and subject coaches.

The second feature that draws strength from conducive conditions for data use in Ronka et al’s (2010) theory of action are data-driven actions that is founded on and informed by data of a high quality. The actions include policy decisions that are informed by data, district and school programs that are directed to addressing gaps that are identified through data use, day-to-day data-driven practices and appropriate placement of learners in schools (Ronka et al,
The quality of data determines the effectiveness of these activities and how they impact on learner achievement.

The third feature of the data use theory of action by Ronka et al (2010) is the climax of effective use of data of a high quality to drive activities. It is the resulting improvement in learner achievement (Ronka et al, 2010:2, 6). Arguably, the growing interest in large-scale assessments worldwide in general and in South Africa in particular, could tacitly be undergirded by some theory on how data should be used in making specific decisions. The contention in this study is that, unless such theoretical frames are clearly understood, and effectively applied, any information that hinges on them is not likely to lead to appropriate action.

A large part of this study included setting performance standards. The involvement of teams of experts and key stakeholders in the standard setting process ensured that the standards are responsive to different contexts in South Africa and will, therefore, be more acceptable to most users of the LSAS results.

2.16 Key principles

Emerging from the literature review are key principles and criteria that must underpin the development and use of an NRF that will adequately enhance the use of LSAS data to promote effective teaching and learning. Key among the principles is data relevance. All the case studies that were reviewed indicated the importance of curriculum relevant information for teachers to take the assessment seriously (Timperly, 2009:21; Ronka et al, 2010:3). Linked with relevance is adequacy and appropriateness of capacity for the use of data Ronka et al (2010:2). The issue of capacity permeates all research. The need for appropriate capacity among teachers and other users of data to ensure data is put under proper use seems to be paramount. In addition to requisite capacity, Ronka et al (2010) identified two other principles that they claim create conducive atmosphere for data use. They identified that, for a systemic, sustainable and learner-centred approach that will help improve learner achievement, the quality of the data that is made available as well as a permeating culture of data use were necessary conditions in the successful data-driven districts that they had studied in detail (Ronka et al, 2010:2-4).
The need for a critical minimal content and pedagogical knowledge among teachers, in particular, was identified in many studies as a critical determinant of effective use of assessment to improve teaching. The role of school leadership in providing instructional guidance on the use of data was also identified as a necessary element that creates a conducive and supportive environment for data use.

2.17 Design and quality of tests for SS

In the main, research that has been published on standard setting that distinguishes between test-centred and examinee-centred methods (Jaeger, 1989; Cizek & Bunch, 2007; Eckes, 2012). In the first method judgements are made on test items while in the second judgements are about the examinees. Among the test-centred methods, original works were started with multiple-choice items (Angoff, 1957) but over the years there has been a significant increase in setting standards from tests with either a mix of multiple-choice or essay-type examinations (Engelhard & Stone, 1998) with a similar increase in the techniques for calculating cutscores. At the same time, there is recognition that the quality of standards that are developed from test-centred methods is as good as the quality of the tests (Eckes, 2012:277).

Using test-based methods of setting performance standards is premised on the assumption that the test is a valid, reliable and credible instrument for the purpose and the selected sample. Yet in the reviewed literature on test-based methods of setting performance standards information on the psychometric properties of the tests that were used was often very superficial. To refrain from making unwarranted inferences about the test and the performance standards, it is critical that the psychometric properties of the test be checked for validity, reliability and appropriateness for the target population. As Eckes (2012:277) argues, making a judgement on which SS method to use should also take into consideration the design, format and purpose of the test to be used. Some of the critical psychometric and content features of tests that literature indicates need to be taken into account when setting test-centred standards are discussed in this section.
2.17.1 Validity of tests

Cook and Beckman (2006:7) define test “construct validity” as “… the degree to which a (test) score can be interpreted as representing the intended underlying construct”. These authors assert that evidence about the validity of a test is inherent in the test itself but also stretches beyond to include the extent to which the test represents the intended content and construct. A test is also valid if there is a known clear alignment between the thought processes of the target subjects and the intended construct that the test measures. This suggests that, for example, for a mathematics test to be considered valid, the content included in the test must be aligned with what is known about how learners learn mathematics (epistemological knowledge).

Cook and Beckman (2006:7) further assert that the results from a valid test can be correlated with results from another test that measures the same construct. The latter is an important observation regarding LSAS where performance and its implications need to be tracked over time. For instance, test designs that permit linking and equating of tests that are administered on different populations at different points in time can improve the credibility and confident use of results from LSAS of the kind of ANA in South Africa.

With the increase in the use of more modern models such as the IRT in assessment, definition and identification of key psychometric indicators of test validity and reliability have become the focus of assessment analysis. In the Rasch model Wright (1996:472) defines (person) separation as “… the number of statistically different performance strata that the test can identify in the sample”. A related concept, (person) reliability, is defined as the extent to which the test results are reproducible (Linacre, 2014:620), corresponding in some particular way with the Chronbach’s measure of internal consistency in CTT. For example, a test that is reproducible, i.e. of high reliability, is expected to have the same separation index every time it is administered on an appropriate population.

These and other concepts that are based on modern theories of assessment serve to enhance the quality of information that can be derived from LSAS. For instance, a test with a person separation index of 4 and a reliability of 0.80, in terms of Wright’s (1996:472) definition, denotes that the measurement “sensed” four distinguishable strata in the sample in terms of
the defined latent trait that the test measures. From such a test it would be logical to expect that standard setting could be used to categorise the test takers into at least four valid PLs. The reliability index of 0.80 indicates that similar results will be obtained 80% of the time every time this test is administered on an appropriate population. This level of detail and accuracy in reporting results of LSAS is critical for the NRF as it is likely to enhance confidence in the results and thus create a conducive atmosphere for data utilisation.

2.17.2 Item and person fit

All IRT models require that there be a good fit between the model and the data that is being analysed (Bond & Fox, 2007). When the model-data fit is not realised in one model, common practice in most IRT circles is to look for an alternative model that will fit the data, often from among the 1PL, 2PL and the 3PL IRT models. The Rasch model, to the contrary, requires, as a necessary and sufficient condition, that the data fits the model and not the other way round (Bond & Fox, 2007:266-268). Unlike other IRT models, in the Rach model data that does not fit the Rasch model requires an investigation rather than a modification (Bond & Fox, 2007; Fischer, 1994; Michel, 1999; Narens & Luce, 1993). However, other researchers such as Goldstein (1982) and Fan (1998) hold a different view despite sustained counter-refutations from the proponents of the Rasch model (Bond & Fox, 2007:275-276).

In a good Rasch-based measurement, both items and persons must fit the Rasch model with minimal error. The fit is indicated by two kinds of statistics, viz. the infit and the outfit mean-square values. Ideally and mathematically both the infit and outfit mean-square values must be equal to 1 in any measurement situation where the data fits the Rasch mathematical model. Values in the range “$1 \leq \text{infit/oufit} \leq 1.5$” indicate a reasonable fit (Bond and Fox, 2007). Fit statistics that are significantly less than 1 indicate local dependence which occurs when a response to one or more items depends on responses to one or other items in the test (predictability) while those much larger than 1 indicate ‘noise’ or erratic patterns of responses (Bond and Fox, 2007).

For the NRF the infit/oufit indicators provide very useful information that should lead to enlightening interactions among teachers, researchers and officials when LSAS results are disseminated and mediated. For instance, test items that show consistent “noise” or model underfit should be interrogated for possible Differential Item Functioning (DIF). Wang
Q. Moloi

(2009:90) describes how DIF manifests: “… it (DIF) occurs when test-takers having identical levels on the latent trait that the test was designed to measure but belonging to different groups, have different probabilities of endorsing (or answering correctly) a particular item”. Within the NRF, indications of DIF that are related to gender, socio-economic status (SES), racial groupings of learner populations and others, will be interrogated by standard setting teams so that assessment is seen to be fair and transparent.

2.17.3 Test targeting

Test targeting refers to whether the test was at the appropriate level of difficulty relative to the abilities of the test takers (Wright, 1996:472). Wright and Stone (1979) defined the best test design as one in which there is, among others, a carefully considered match between the average ability of examinees and the average difficulty of the test items. In the Rasch model test targeting can be estimated and represented visually in item maps where test item difficulties and person abilities are measured on the same scale. On this scale items that are well-targeted to the sample of test takers appear at the same level as the person abilities. Items that are not well targeted appear either above or below the person abilities on the item map. Items that appear above the person abilities are relatively more difficult for the sampled persons and those that appear below the person abilities are relatively easier for the test takers. Tests that are not properly targeted have large measurement errors and low reliability indexes (Bond & Fox, 2007:43).

A standards-based NRF will influence not only the reporting of LSAS results but also the quality of tests at the development stage. For instance, tests that suffer either ceiling or floor effects may result in inaccurate classification of learners according to their proficiencies in particular subjects. Koedel and Betts (2008:1) define test score “ceiling effect” as the “tendency for gains in a learner’s test score to be smaller if the learner’s initial score is toward the top end of the distribution, simply because the learner has little room for improvement given the difficulty level of the test”. Test “floor effect” is the opposite of “ceiling effect”.

On the one hand, ceiling effects occur when a test fails to measure learner ability at the top end because it does not include items of corresponding cognitive demand. On the other hand, floor effects occur when a test fails to measure learner ability at the bottom end because it
does not include enough items of low cognitive demand. In either case, mis-targeted tests provide inadequate information for diagnostic purposes in a criterion-referenced paradigm.

### 2.18 Summary

A summary of views on assessment, its usage and implications was aptly captured by Ronka et al (2010:3) as:

> “Those who advocate for more use of data in the classroom say it can give teachers concrete evidence of what instructional strategies work. But critics worry that an increasing focus on metrics could lead schools to play down intangible factors that enhance learning and inspire learners. With many measurements based on some kind of test, some critics say the drive to collect more data could exacerbate the testing culture in schools or simply create more busywork”.

Whilst Ronka et al (2010) were referring specifically to classroom assessment, it can be argued that their synopsis applies to LSAS as well. This is particularly so in the case of ANA in South Africa.

Based on the review of literature regarding the arguments made for and against models of large-scale assessments (Braun & Kanjee, 2007; Klinger, DeLuca & Miller, 2008), the value in setting and using performance standards to report learner performance (Cizek & Bunch, 2007) and the influence that LSAS may have on school and learner performance (Underwood, Zapata-Rivera & Van Winkle, 2010; Hambleton & Slater, 1994; Nzomo & Makuwa, 2006; Ravela, 2005; Bernard & Michaelowa, 2006), it is clear that “the weakest link” in the South African “value chain” of LSAS is in the absence of a national framework that specifies effective reporting, interpretation and use of the LSAS results.

There is need for greater attention to developing a national framework for reporting and using scores from LSAS. Specifically, a National Reporting Framework (NRF) should be developed that will:

- define and specify the inputs, activities and processes at each key stage of the LSAS;
- specify the type of data that must be collected;
Q. Moloi

c. specify the kinds of analyses that must be conducted; and,

d. Specify how the results should be disseminated and presented to the key users to enable them to implement and apply the results for use in improving learner performance in schools.

This study argues that reporting and using results from LSAS against performance standards criteria that are non-arbitrary and explicit, will not only improve the relevance and validity of ensuing policy decisions, but will also increase the acceptance and utilisation of these results. The methodology of the study is presented in Chapter 3 and the findings in Chapter 4.
3. RESEARCH DESIGN AND METHODOLOGY

3.1 Introduction

As was indicated in Chapter 1, the research problem that prompted this study was the observation that there was limited evidence to suggest that results from large-scale assessment surveys (LSAS) are communicated to provinces, districts and schools in meaningful ways to inform decision-making for identifying, developing and implementing relevant interventions, in order to improve learner performance in the basic education sector of South Africa. A compounding factor was the observation regarding the lack of clear guidelines or policy that stipulates how data from national assessments should be communicated and how different levels of the education system are expected to utilise the data to improve teaching and learning and hence improve performance.

This study was undertaken to develop a national framework for reporting results from LSAS and provide teachers and education officials with information that enables them to support learning and improve learners’ repertoire of skills and knowledge as specified in the national curriculum. The framework is based on performance standards that categorise learners according to what they can and cannot do as demonstrated in the assessments. Reviewed literature, both in South Africa and internationally, showed that the use of performance standards in reporting assessment results is the preferred approach compared to reporting in raw scores that do not provide information on what learners know and can do (Kanjee & Moloi, 2012; Bunch, 2007; Bejar, 2008). However, what is not known is which of the many standard setting methods that have been reported in published literature would be the most appropriate in the context of South Africa and what guidelines and reporting formats should be provided to the users of a standards-based reporting framework.

3.2 Research questions

This study set out to answer three research questions:

1. Which of the three standard setting methods: the Angoff, Objective Standard Setting and Objective Borderline Method, is most appropriate for developing performance
standards that are relevant for a National Reporting Framework in the South African context?

2. What guidelines should be provided to national, provinces, districts and schools to enhance the utilisation of information from LSAS?

3. What are the perceptions of national and provincial officials regarding comparisons between a standards-based report and a report based on the Curriculum and Assessment Policy Statement (CAPS) rating codes?

3.3 Research design

There was notable paucity of local and international research studies on developing and using national frameworks to guide recording, reporting, dissemination and use of LSAS data to improve teaching and learning. Similarly, there was very limited research available on comparing the efficacy and impact of using different methods of standard setting, particularly in the context of South Africa. For these reasons, the design of this study was largely exploratory. The University of California open source library (2014:1) defines an exploratory design as one that “… is conducted about a research problem where there are few or no earlier studies to refer to or rely upon to predict an outcome. Exploratory designs are often used to establish an understanding of how best to proceed in studying an issue or what methodology would effectively apply to gathering information about the issue”.

The nature of and the required knowledge for answering the different research questions in this study transcended the traditional boundaries that are assumed to exist between quantitative and qualitative approaches to research (Cohen, Manion & Morrison, 2011:7-8,537-539). Instead, the study adopted a mixed approach design and followed iterative processes where, for instance, qualitative descriptions of hierarchies of knowledge and skills required in the curriculum were obtained from expert educators and, subsequently, quantitative designs and techniques were employed to analyse and either refute or confirm judgements made by standard setting expert panellists, particularly in answering the first research question.

Motivating in favour of a mixed design theoretical framework, Greene (2008:13) avers that research methodologies (should) follow from the purposes and questions that the research
seeks to answer and not the other way round. Consequently, according to Cohen et al (2011:25), different methods, different sets of data collection instruments and different combinations of analysis methods may be required for the same study. The second research question in this study was answered through a literature review, unlike the first and third research questions which were answered through analysis of empirical data. The limitations of this research design as well as the arising opportunities for further research are addressed in Chapter 7.

Adopting an exploratory design and proceeding from the literature review on the use of data from LSAS to improve the quality of teaching and learning, the study is intended to develop a standards-based national framework for effective use of data but, in the process of doing so, also identify practices, technical requirements and capacity issues that must be part of the NRF to enhance use of assessment data. Following the review of relevant literature on developing performance standards and setting cutscores, the study proceeded in five main chronological steps. Firstly, various procedures were followed to determine cutscores for mathematics and language using each of the Angoff, the OSS method and the OBM with a view to comparing the resulting cutscores and identifying the most appropriate method for the South African context. It was noted in this study that some research has suggested that there is no ‘best’ method of setting performance standards and cutscores (Jaeger, 1989; Cizek & Bunch, 2007; Eckes, 2012). But there were also dissenting views arguing that some SS methods were flawed in ways that cannot be overlooked (Stone, 1995, MacCann & Stanley, 2006). This apparent dissonance in schools of thought necessitated ways of validating the cutscores from the different SS methods.

Secondly, to determine the validity of the cutscores from the different SS methods, learner data was simulated and corresponding cutscores from real and simulated datasets were compared for possible similarities or differences. This was done to test the external validity of the different cutscores. Neuman (2000:172) defines ‘external validity’ as “the ability to generalise findings (results) from a specific setting ... to a broad range of settings ...”. Theoretically, cutscores with external validity should not vary from one setting to another.

Thirdly, cutscores based on real data were compared directly across the three SS methods to determine whether the different methods yielded similar or different cutscores and, therefore,
decide on which method would be the most appropriate for the South African context. The contention of this study was that, if the notion that different SS methods can produce comparable cutscores (Cizek & Bunch, 2007; Tiratira, 2009), then two or more SS methods that produce cutscores with ‘external validity’ could be used interchangeably. If not, then other criteria may have to be used to choose the ‘best method’.

Fourthly, the different SS methods were compared in terms of the processes involved in the execution of each method and also in terms of possible impact of the resulting cutscores in categorising learners according to performance. The observations that were made during the ‘exploration’ with the different SS methods were synthesised into criteria that can be used in selecting the most appropriate or efficacious SS method. Such criteria could be included in the NRF so that future SS processes will benefit from lessons learnt previously.

In the fifth and final step, perceptions of education officials were obtained as a preliminary test on whether they would prefer standards-based reports or reports based on raw scores in assessment. By virtue of being mediators of education delivery, education officials are key to whether any innovation will be accepted or not accepted in schools. It was considered crucial, therefore, that perceptions of relevant education officials on the use of standards-based reports be tested so that relevant recommendations can be made for the NRF.

3.4 Types and sources of data

This study obtained, collected, analysed and used data from various primary and secondary sources to answer or substantiate answers to the key research questions that guided the enquiry. The types of data that were obtained from secondary sources and collected from primary sources for purposes of answering each of the research questions are discussed in the relevant sections. For each research question the source from which data was either obtained as a secondary source or collected as a primary source is also described.

Research Question 1

Which of the three standard setting methods: the Angoff, Objective Standard Setting and Objective Borderline Method, is most appropriate for developing performance standards that are relevant for a National Reporting Framework in the South African context?
3.4.1 Learner data

An important dataset that was used in this study were learner test scores (DBE, 2013a:26-28). The SS methods that were the focus of this study were all based on the test-centred approach. Learner quantitative data, in the form of test scores, was obtained from the Department of Basic Education (DBE) from a stratified random sample of grade 6 learners that had participated in the mathematics (n=8 131) and language (n=6 104) Annual National Assessments. These assessments were conducted by the DBE in the second week of September 2013. Details about sampling, the sample and how the Department administered and marked the tests can be found in the ANA 2013 Report (DBE, 2013a:16-28).

3.4.2 Test item rating scores

Test item rating scores were generated by panellists that participated in the different SS processes, i.e. the Angoff, the OSS and the OBM. The panellists in each SS team were, therefore, the primary source of this quantitative data. Each SS dataset is described and presented in the relevant section.

3.4.3 SS-process observation scores

The SS panellists were requested to evaluate each of the three SS methods on a specially-designed observation sheet. This information was important for comparing the different SS methods on a number of criteria that were identified and have been described in detail in the section on instrumentation.

3.4.4 Guidelines on use of assessment data

Guidelines were gleaned from literature that was reviewed on factors and conditions that have been shown to either impede or enhance use of assessment data to improve the quality of teaching and learning in schools. This was purely qualitative information rather than numerical data. The information was used to answer the second research question in this study.

Research Question 2
What guidelines should be provided to national, provinces, districts and schools to enhance the utilisation of information from LSAS?

To address this question, a systematic review of literature on the development and use of assessment data was conducted. The literature was specifically reviewed to identify, firstly, conditions that research and practice identified as necessary for schools and teachers to use assessment data to improve their practice and, secondly and conversely, factors that were identified as hindering or had the potential to hinder use of data by schools and teachers. One criterion that was used for selecting relevant research studies to be considered for this purpose was that the study had to have been conducted in a sample that was at least the size of a district, where district size was estimated from the South African context comprising between 75 and 300 schools (DBE, 2013c:16). Research findings from smaller samples were considered to be a sustainability risk at larger scale levels and were subsequently excluded from the selection.

**Research Question 3**

What are the perceptions of national and provincial officials regarding comparisons between a standards-based report and a report based on the CAPS rating codes?

Data to answer this question was collected from a survey of perceptions held by national and provincial officials on the relative benefits of standards-based and ‘raw score’-based reporting formats for the ANA results. The survey model is a qualitative research approach that aims at determining the characteristics of a group or describing a past or present state as it was/is (Neuman, 2000:250). A self-completed questionnaire was given to national and provincial officials who were charged with management, monitoring and support responsibilities in examinations and assessment, curriculum delivery and teacher development. The questionnaire required respondents to indicate their preference between two reporting templates, one that was based on the rating codes that are described in the CAPS document (DBE, 2011c) and the other that was based on use of performance standards.

**3.5 Instrumentation**

The instruments that were used to collect data for this study were a mix of some obtained or adapted from previous studies and others that were developed specially for data collection in
this study. The composition of the instruments, their validity as defined by Cook and Beckman (2006:7) and how they were used for specific purposes are described in this section.

3.5.1 Mathematics and language tests

As indicated in Section 3.4.1, the mathematics and language tests that formed the basis of learner data in this study were developed and administered by the DBE (DBE, 2013a). Copies of the test frameworks, tests, memoranda as well as item-level data were obtained from the DBE for the purposes of this study. Each of the mathematics and language tests had a mix of multiple-choice (MC) and open-ended (OE) items. The mathematics test had 50 items allocated a total of 75 marks. The FAL test had 46 items allocated a total of 46 marks.

The dataset from the tests was analysed using the Winsteps program Version 3.81.0 (Linacre, 2000) which operates on the principles of the Rasch Model of measurement, particularly selecting the Partial Credit Model (PCM) because each test had a mix of MC and OE items (Bond and Fox, 2007). The Winsteps program was used to calibrate item difficulties (in logits), diagnose the nature and level of fit of the data to the assumptions of the Rasch Model, i.e. the requirement for unidimensionality, and whether the items were lined up in a logical increasing order along the measured latent trait which was defined as the mathematical and language ability of the measured learners, respectively.

The tests were evaluated in terms of their content and construct validity (Cook & Beckman, 2006:7), reliability, and whether they were properly targeted for the intended population.

3.5.1.1 Validity and reliability of the tests

For each subject the test that was used for SS was based on content that was prescribed in the national curriculum documents and schools were aware that this was work expected to be covered during the first three quarters of the year before they participated in the Annual National Assessments. The content in the tests, therefore, was expected to be familiar to all the learners. The tests also displayed acceptable psychometric indexes. The person separation index of the mathematics test was 3.02 which, in terms of Wright’s (1996:472) definition, denotes that the test “sensed” three distinguishable strata in the sample in terms of the latent trait which was defined as mathematical ability or proficiency. The corresponding person reliability statistic for mathematics was 0.90. For the language test the person separation
Q. Moloi

index was 2.97 and the reliability was also 0.90. Both tests met minimum criteria for construct and measurement validity.

3.5.1.2 Test targeting

Test targeting as defined by Wright (1996:472) refers to whether there is a balanced match between the difficulty of the test and the abilities of the test takers. Using the Rasch item mapping technique as described in the literature review of this study, the match between sampled learners’ proficiencies and respective test item difficulties was examined for appropriateness of test targeting in each of the mathematics and language tests. The item maps showing test targeting in the grade 6 mathematics and language tests that were used in this study are shown in Figure 10 and Figure 11, respectively.

In Figures 10 and 11 item maps that show matches between average learner proficiencies and average item difficulties for each of the tests are presented.
From Figure 10, the grade 6 mathematics test was reasonably well targeted to the sampled learners. Item 20 was the most difficult item and Item 1.1 was the easiest. However, the gap between Item 20 and Item 22.2 was too wide not to have items. This means that the estimated measures (abilities) of learners at the top end of the performance continuum in mathematics were likely to have large measurement errors.
From Figure 11, the grade 6 FAL test was also reasonably well targeted. The most difficult item was Item 4 while the easiest was Item 16. There was a fairly wide gap between Item 4 and the next two items, viz. Items 26 and 8. Again this means that the top end learner scores in FAL were likely to have large measurement errors.

### 3.5.2 Instruments used in SS

Other data collection instruments that were used in this study included rating sheets that were used by panellists for the Angoff and OSS SS methods, respectively. This section describes the SS instruments and how their validity was ensured.

#### 3.5.2.1 Instruments used in Angoff SS
The Angoff SS method involves panellists in rating test items according to a prescribed rule and then recording the ratings on a specially prepared rater sheet. The rating sheets that were used in the Angoff SS method in this study and for each subject were adapted from previous studies (Hambleton & Pitoniak, 2010). The rater sheet listed all the items on the left and had columns for each of three rounds of ratings. It also gave some indication of the number of the item, what the item was assessing and whether the item was a MC or OE question. An extract of a rater sheet is in Appendix A.

For ease of calculation of cutscores, summary sheets aligned with the rater sheets were developed in Microsoft Excel for capturing the ratings. Formulas were included to indicate discrepancies between the ratings of various raters in order to assist raters to focus their discussions on items where ratings were widely discrepant. The summary sheets also showed the cutscores implied by the ratings in terms of both percentages and raw scores. An extract from a completed rater form summarised item ratings Appendix B.

Both the rater sheets as well as the summary sheet were adapted from instruments which had been used in a number of published standard setting studies involving the Angoff SS method (Angoff, 1971; Hambleton, 2000; Stone, 2001; Näström & Nyström, 2008; Hambleton & Pitoniak, 2010). The two instruments were submitted to two senior researchers for comment on their appropriateness and relevance for the purpose. Independently the researchers concurred that the instruments were relevant, appropriate and, therefore, valid.

3.5.2.2 Instruments used in the OSS

For the OSS method, the instrument for rating items was developed specifically for this study. The key features of the instrument were the identification of the item in terms of its number in the test, the subject content area that the item was assessing and provision of space where the rater would indicate if the item was “essential” at the Not Achieved, Partly Achieved, Achieved or Advanced level. The instrument was submitted to one senior researcher for comment on its suitability for the purpose and there was consensus that the instrument was valid. A specimen of the rating form for the OSS method has been shown in Appendix C.

3.5.2.3 SS evaluation instrument
Part of the design was to explore how the development and use of performance standards could be integrated into the NRF and whether there were standard setting methods that could work best in a framework designed for the South African context. Consequently, observations that were made in the course of implementing different standard setting methods were recorded carefully and emerging patterns were analysed, firstly, to inform the development of the NRF but, secondly, to form a basis for more structured study designs in standard setting in future. Building on key principles that were identified as critical to the selection of a SS method in the reviewed literature, a standard setting evaluation instrument (SSEI) was developed and used to evaluate the three methods of SS in this study (Appendix D).

3.5.2.4 Survey questionnaire

A special questionnaire was designed for collecting data from officials in the basic education sector on their views regarding the use of a CAPS-based vs a standards-based reporting template for reporting results of the ANA. Before administration and for purposes of validation, the questionnaire was given to two officials who were responsible for technical aspects of assessments in the Department of Basic Education for their views on its appropriateness to elicit the intended information. Both officials agreed that the questionnaire had the necessary face validity, at least. The two officials were subsequently excluded from completion of the questionnaire. The questionnaire and the reporting templates that were used in the survey are in Appendix E.

3.6 SS processes and techniques

This section traces the processes and methods that were employed in setting and evaluating the impact of performance standards developed from the different SS methods that were being compared in the study. For each process and method of SS the participants, the techniques and the data that was generated are described in detail.

3.6.1 Selection of SS panellists

The first process involved the development of performance levels (PLs) and performance level descriptors (PLDs). Traditionally SS with methods such as the Angoff begins with definition and development of PLs and PLDs, often by the same panel that sets cutscores (Hambleton, 2001). However, Cizek et al (2004) hold a different view. They propose that PLs
be developed before the SS process and be developed by a separate team from those who set standards. Their contention is that participants in the standard setting process become influenced by the set PLs and, consequently, lose objectivity in setting cutscores. In this study the PLs and corresponding PLDs were developed by a separate panel of educators and subject advisors a year prior to the session for developing cutscores.

The process of developing PLs and PLDs took place in June 2013 at a workshop that was hosted at the DBE offices in Pretoria. The participants in the session were teachers and subject advisors or curriculum specialists who were recruited from all the nine provinces of South Africa to set ANA tests in mathematics and language for grades 1-6 and grade 9. The PLs and PLDs were developed in English which is the language used for teaching and learning in the majority of schools, especially but not exclusively for grade 4 and higher. The development of PLs and PLDs was limited to grades 3, 6 and 9.

3.6.2 Development of PLs and PLDs

Prior to the workshop each participant was provided with copies of previous ANA mathematics and language tests and memoranda of the grade in which they specialised, specimens of selected PLs and PLDs from published research, a copy of a journal article on standard setting, and a copy of the CAPS document. They were encouraged to read the documents and prepare for meaningful participation in the workshop.

The first session started off in a plenary engagement where a senior researcher introduced the participants to the concepts of performance levels and and performance level descriptors as used in standard setting exercises. Participants were made aware of, and guided on how to develop, policy definitions of statements as a starting point to developing PLs. After they were shown various options for definition of policy statements and deciding on the number and labels for performance levels, they were given opportunity to discuss what they considered to be the best options for South Africa.

Discussions revolved around, among others, the advantages and disadvantages of defining either too few or too many performance levels or whether there was a reasonable compromise number of PLs. The participants were unanimous that adopting more than four levels could
compromise the valuable principles of simplicity and user-friendliness of the standards which they all agreed were important.

The plenary session resulted in consensus on the proposed number and labels of proposed PLs which the participants recommended to be applied across grades and subjects. The participants shared the view that four (4) PLs were appropriate and agreed to label them as *Not Achieved, Partly Achieved, Achieved* and *Advanced* levels.

After the plenary session, the participants were grouped according to the grade and the subject that they taught or, in the case of subject advisors, were responsible for supporting, and were asked to develop PLDs that would characterise a learner who functions at each PL. The numbers of participants per grade and subject are presented in Table 5.

**Table 5: Profile of participants in setting PLs and PLDs**

<table>
<thead>
<tr>
<th>Grade</th>
<th>Subjects</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mathematics</td>
<td>Home Language</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

From Table 5, the total number of participants in setting PLs and PLDs was 32. There were four (4) participants in each subject and grade except in grade 3 First Additional Language which was not offered in any school and, therefore, was not represented in the workshop. Across the grades in each subject there were 12 participants except in grade 3 where there were eight (8) participants.

The various groups for each grade and subject developed PLDs, with clear descriptions of skills and knowledge required as evidence of performing at a particular level, drawing heavily from their classroom experiences but also aided by copies of the CAPS document which were provided as a resource. Participants were continually conscientised to the intrinsic and embedded hierarchy in the definition of the PLs. In particular, the rigour of the PLDs in each subject and grade had to reflect the progressive increase in cognitive and skill demand from the *Not Achieved* across to the *Advanced* PL. The need to recognise and realise the hierarchy of PLs and corresponding PLDs was to be the “golden thread” throughout the SS process, regardless of the method to be used.
The last engagement in the development of PLs and PLDs were three (3) parallel plenary sessions, one for all the mathematics groups, the other two (2) for each of FAL and HL. As part of the “golden thread”, the focus of the plenary sessions was to ensure that the PLDs reflected the necessary progression across the grades in each subject. Again the process involved expression of views, counter-views, discussion and eventual consensus. The respective teams then tasked few individuals to finalise editorial corrections on the PLs and PLDs and the day’s activities were closed. Samples of grade 6 PLs and PLDs that were developed are shown in Table 6 for mathematics and in Table 6 for FAL.

### Table 6: Grade 6 mathematics PLs and PLDs

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

### Table 7: Grade 6 FAL PLs and PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Achieved</td>
<td>Partially Achieved</td>
<td>Achieved</td>
<td>Advanced level</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------</td>
<td>----------</td>
<td>----------------</td>
</tr>
</tbody>
</table>
| A learner at this level will, with intensive support, read single words, short sentences & pictures. He/she can:  
  - recognise words that were learnt previously  
  - read forward and answer short questions based on the text  
  - identify only the key character in a text | A learner at this level can, in addition to skills & knowledge in the lower PL, read texts with simple sentences. He/she can:  
  - skim & scan some poetical elements  
  - skim, scan and summarise texts  
  - use a vocabulary of 1500 words.  
  - write friendly letters  
  - extract information directly from a short text  
  - read short stanzas and answer literal questions | A learner at this level can, in addition to skills & knowledge in the lower PLs, read & interpret meaning of stories relating to real life experiences. She/he can:  
  - identify characters & plot setting  
  - identify ethical issues such as cultural/social diversity  
  - highlight the moral lesson behind a story  
  - infer information from a complex text  
  - identify key elements of poetry  
  - use dictionaries in building vocabulary of at least 1500 words  
  - understand factual information from non-fictional documents | A learner at this level can, in addition to skills & knowledge in the lower PLs, read & understand complex texts in formal and informal stories. She/he can:  
  - read poetry effectively  
  - evaluate texts  
  - identify formal and informal texts  
  - easily utilise vocabulary of at least 5000 words  
  - critique texts through book reviews and reports.  
  - analyse both formal & informal documents  
  - synthesise information from different parts of a text  
  - demonstrate comprehension of inferred information from a text |

**WRITING**

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced level</th>
</tr>
</thead>
</table>
| A learner at this level will, with intensive support, show understanding through drawings and pictures. He/she can:  
  - write single words and short sentences  
  - represent ideas with drawings  
  - spell few commonly used words correctly  
  - write brief diaries | A learner at this level can, in addition to skills & knowledge in the lower PL, write for social purposes. He/she can:  
  - write personal letters, diaries, news reports  
  - create a book cover  
  - identify similar & different texts  
  - express cause and effect relationships  
  - write simple sentences. | A learner at this level can, in addition to skills & knowledge in the lower PLs, write information in graphic and descriptive forms. She/he can:  
  - write for social purposes, e.g. frames, personal letters, diaries, dialogues & simple news reports  
  - design a book cover  
  - develop & edit key language structures. | A learner at this level can, in addition to skills & knowledge in the lower PLs, write & develop information from graphical and descriptive forms to text. He/she can:  
  - write extensively writing for social purposes  
  - develop news reports  
  - design questionnaires and adverts.  
  - integrate ideas by classifying information  
  - solve problems  
  - use relevant questioning styles to obtain information. |
PLs and PLDs were developed for grades 3 and 9 for each subject and are presented with a view to showing how the increasing complexity of skills and knowledge across the grades in each subject was reflected. The PLs and PLDs that were developed for each grade for mathematics and language shown in Appendix F for grade 3, Appendix G for grade 6 and Appendix H for grade 9.

3.7 Data collection

The main data collection sessions, which were standard setting working sessions, took place in January (Angoff Method) and March (OSS Method) in 2014, respectively, and were followed by the administration of the survey questionnaire for education officials. The OBM was applied last. Where applicable, the key processes in each SS method included training of participants on the appropriate method(s) of setting standards, implementing the appropriate SS method, capturing resulting data and calculating cutscores. There was also a survey of the perceptions of education officials on the use of performance standards against raw scores in reporting learner assessment data. This section provides an account of how data was collected and the techniques that were used to do the necessary analysis in each process.

3.7.1 Angoff SS data collection

A proposal was submitted to the DBE to consider developing and reporting assessment data in performance standards. The benefits of performance standards over raw scores in communicating and using the results of assessment were motivated and the DBE accepted the proposal. The DBE subsequently issued a letter to all nine (9) South African provinces requesting them to identify teachers and subject advisors who demonstrated outstanding competency in their profession to participate in the standard setting process. As part of the selection criteria the officials had to be either teaching mathematics or language in grades 3, 6 and 9 or were involved in supporting teachers in these grades and subjects. The officials were, therefore, recommended by their senior supervisors in the various districts on the basis of their competencies and specialisation in either mathematics or language teaching.

A key principle in the recruitment of participants was that for the standard setting process to be acceptable at national level it needed to have a reasonable representation of participants across the provinces and other historical, racial and social contexts that mark the South
African society in general and the education sector in particular. The session was hosted at the DBE offices in Pretoria on 24 January 2014. The DBE paid for the travel of the officials which included flights in some instances, accommodation and meals for the participants. The costs were, therefore, one inhibiting factor in terms of the numbers that could be included in the process.

Before the workshop the DBE sent some general information on standard setting to each nominated participant. The documents included an article on using the modified Angoff technique for standard setting, a brief summary of what standard setting entails and some example of performance level descriptors (PLDs) arrived at during a previous standard setting conducted by DBE.

The original plan was that there were going to be two main groups per grade, viz. one for mathematics and the other for language. Participants for each subject were further split into two sub-groups, Group A and Group B, to avoid a situation where eloquent individuals could dominate the discussions and thus limit the desirable diversity of views, especially across racial groupings given the observations made by Pitoniak and Yeld (2013). However, it turned out that, at grade 6 and 9 levels, teachers specialised in either Home Language (HL) or First Additional Language (FAL). Consequently, language participants had to be further split into these two subjects that were treated as distinctly. The specialisation of the participants straddled different grades and subjects as shown in Table 8.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Subject</th>
<th>Group A</th>
<th>Group B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Mathematics</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Home Language</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>Mathematics</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Home Language</td>
<td>4</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>First Additional Language</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>Mathematics</td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Home Language</td>
<td>3</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>First Additional Language</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>45</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As can be observed from Table 7, the total number of officials who were recommended and actually attended the session was 45. The numbers of participants per subject ranged between three (3) for grade 9 Home Language and eight (8) for grade 9 mathematics. The split
between Home Language and First Additional Language at grades 6 and 9 levels, which was not the case at the grade 3 level, resulted in relatively smaller groups in the languages for the two grades.

There were seven (7) panellists who participated in setting performance standards for grade 6 mathematics. All of them were educators. Four (4) of them were African Black males, and of the three (3) females one was Indian, one White and the other African Black. There were five (5) panellists in the grade 6 FAL group - two (2) Black African males, two (2) Black African females and one White female. In the further split into Group A and Group B, there were three (3) participants in Group A for each of mathematics and FAL at the grade 6 level - four (4) participants in mathematics Group B and two (2) for FAL.

The Angoff SS session started with a review of the PLs and PLDs that had been developed in 2013 as was described in Section 3.6.1. The participants endorsed the PLs and PLDs and confirmed that they were still relevant, mainly because there had been no changes in the national curriculum between the two events. They also adopted the number of PLs, the policy definitions and the labels of the PLs. The session to review PLs lasted for less than an hour mainly because the participants had been given all the materials to read before coming to the workshop.

### 3.7.1.1 Training of panellists

The training was aimed at familiarising the panellists to the theory of standard setting, the test that they were going to work with and the actual rating of test items. The training of panellists was conducted by three researchers, one of whom was a SS expert. Key activities in the training of panellists included a PowerPoint presentation by the SS expert, questions from participants and explanatory answers by the SS expert, test-taking by panellists, leading panellists on how to use the rating forms, a practical exercise to give panellists experience on how to use the rating forms and an outline of the procedure to be followed in the rating process.

The presentation was made to be as interactive as possible. For instance, the open-ended question that was given to the participants to introduce the presentation was “What is the main purpose of assessing learners?” to which different responses were initially given but the
final consensus was that the main purpose of assessment is conducted to obtain evidence of learning, with one of primary uses being improvement of learning. The presenter then introduced key concepts such as cut-scores and how they are used to differentiate learners according to their competencies, performance level descriptors and how they help to identify learners who can from those who cannot demonstrate expected knowledge and skills and possible impact that categorising learners may have for the education system.

Particular attention and proportionally more time was spent on defining a “minimally competent” or “borderline” learner. A question-and-answer method, bolstered with different examples, was used to help participants internalise this definition which was to guide the entire rating process. To simplify the conceptualisation of “minimally competent”, participants were asked to make required estimations by answering the question “How many of ten ‘minimally competent’ learners in your class will get the right answer?” for each of the items that carried one mark. For items that carried more than one mark, the variation of the same question was: “What will be the average mark of ten “minimally competent” learners on this item?”

To enable participants to familiarise themselves with the test, they were asked to individually respond to each item in the test that they were going to work with. The time taken to respond to all the items ranged between 20 and 25 minutes. Upon completion of test-taking, panellists were each given a rating sheet and asked to rate items from a different test, with the same item structure as the target test, which was deliberately used for practice purposes. To limit variations in responses that required estimating average scores of “minimally competent” learners during the practice session, panellists were restrained to select from given values, which were limited to multiples of five (5).

In the process of the practice exercise, the three researchers moved from one panellist to another observing how ratings were done while taking questions that individuals raised. Importantly, all questions and identified misinterpretations of the procedure were communicated to the SS expert/presented so that they were in turn communicated to all the panellists. This was an important step to ensure common understanding among both panellists and researchers. When all questions were exhaustively answered, and erroneous
conceptions were corrected, the presenter outlined the process that was to be followed in rating items in the target tests.

### 3.7.1.2 Rating of test items

The Angoff method centres on estimating the likely achievement of learners who function just across the borderline into a specified PL. Such learners may be referred to as “Just achieving” learners. Three such ratings had to be done for each item, one for each of the Partly Achieved, Achieved and Advanced PLs. No such rating was required for the bottom category (Not Achieved) as this lowest level was automatically defined as the zero-score level or any score below the “Partly Achieved” level.

For each item counting one mark each rater had to answer the following question. “How many of 10 just proficient learners will get this item right?” For each item counting more than one mark he/she had to estimate the mean score of 10 just proficient learners at each level. For each group, ratings for each level of each item were averaged. Then average ratings of items were combined to arrive at the cutscores for the test. The process of rating items followed three (3) rounds. This section presents an account of the purpose, the activities and the outcomes of each round.

**Round 1**

Panellists were given the specially designed form for rating items and were guided on the information that they were expected to record on the sheet. They then worked individually to rate items at each performance level from Not Achieved, Partly Achieved, Achieved and through Advance and recorded their ratings on the specially designed form. Individual ratings were captured, necessary calculations were done by the research team and feedback was presented in printed excel spread sheets which showed the mean rating of each item by the group and the range in individual ratings per item. In groups, panellists then discussed the feedback, particularly focusing on items where score ranges were in excess of 2 points.

**Round 2**

Panellists worked individually to rate items again, this time taking into consideration the arguments and counter-arguments from the group discussions. Individuals were given the
options of either changing their scores, if they were convinced by motivating arguments or leave their scores unchanged. Individual ratings were captured and feedback was again presented in printed excel spread sheets as described in Round 2. Learner scores, with calculated difficulty levels approximated by \( p \)-values, were given as additional input for consideration. The definition of \( p \)-value of an item as used in this context refers to the proportion of a well-defined population or sub-population of examinees that get the item correct in a test (Stage, 2003:2).

**Round 3**

Panellists worked individually to rate items again, this time taking into consideration the learner responses and \( p \)-values of individual items as additional input to inform their ratings. For example, if the ratings so far indicated that a particular item was estimated to be more difficult than the learner scores or \( p \)-values showed, panellists were free to either re-consider their ratings or leave the ratings as are if they were not convinced otherwise.

**3.7.1.3 Capturing of data**

On completion of each round of ratings the rating sheets were collected and captured on an Excel spread sheet by a team of appointed data capturers. Each dataset was double-captured by two persons. A senior researcher monitored the data capturing process and, when errors were identified, the affected data capturer was stopped and asked to check the entire sheet before proceeding. Cutscores were then calculated from the summarised ratings.

**3.7.1.4 Calculation of final cutscores**

The cutscores from the Angoff method were calculated by averaging ratings over items and panellists after the third round of iterations. The cutscores were presented to the participants for comments and, when everybody was satisfied that the cutscores were acceptable, the cutscores were adopted and recommended to form part of the performance standards.

**3.7.2 OSS data collection**

The OSS session focused on setting cutscores for the grade 6 tests. Therefore, only educators and subject advisors who were directly or indirectly involved in grade 6 teaching were invited on 8 March 2014. The session started with training to familiarise the panellists to the
theory of standard setting, the test that they were going to work with and the actual rating of items. There were eight (8) panellists who participated in setting performance standards for grade 6 mathematics. Four (4) were African Black female educators, two (2) were African Black male educators, one was an Indian male and was a subject advisor and one was a White male educator. Similarly, there were eight (8) panellists who participated in setting performance standards for grade 6 FAL. All of them were Black African educators, five (5) females and three (3) males.

3.7.2.1 Training of panellists

Training of panellists for the OSS method proceeded in two key phases that were followed after a practice exercise. In the first phase the researcher made a PowerPoint presentation that outlined the purpose and key concepts that underlie the OSS method which are embedded in the seminal work by Wright & Grosse (1993) with detailed techniques that were subsequently articulated very aptly by Stone (Stone, 2001:192-193). The key concepts that were defined and discussed with the participants were:

**Selectable pool**: refers to all the items in the test as a representation of the curriculum requirements

**Essential items**: refers to the sub-set of the selectable pool that comprises items which a learner must be able to answer correctly at a particular performance level

The participants were introduced to the four stages through which the OSS method normally proceeds (Stone, 2001:192-193) and were afforded opportunity to ask questions of clarity in the process.

The second phase of training was a practice exercise focused on consolidating the comprehension of “essential” and non-essential test items. Immediately after the PowerPoint presentation the participants were given a grade 3 test in each of mathematics and language as a practice exercise on how to identify “essential” items and tick them off the selectable pool. The participants were instructed that no item should be ticked more than once, i.e. an item may not be classified as simultaneously “essential” at two different performance levels.
The researcher monitored the practice exercise closely and, where misunderstanding was identified, e.g. classifying of an item as simultaneously “essential” at more than one performance level, immediate feedback was given and the affected participant was requested to make a decision as to the most appropriate performance level at which the item was “essential”. There were also instances where a participant identified as “essential” only items from one domain of a test. For example, in mathematics one participant identified only items related to Number and Operations as “essential” at one performance level and the other mathematics domains had no representative items in the same performance level. This had to be corrected with an explanation that the subject was comprised of all the domains and, therefore, “essential” items at each performance level needed to reflect this composition. The practice session was made interactive to ensure that participants understood the concepts and procedures of the OSS method.

3.7.2.2 Rating of test items

When there were no more questions of clarity from the practice exercise, copies of grade 6 mathematics and language tests and relevant rating forms were given to the respective teams. They were guided to repeat the process of identifying essential items at each of the three levels of performance, viz. Not Achieved, Achieved and Advanced levels as follows:

Stage One: Criterion Development

This stage, defined as the “content criterion” stage by Stone (2001:191), constitutes an individual activity in which experts in the affected content area identify ‘essential items’ (Stone, 2001:192-193). The essential items must of necessity represent all the domains of the test. Each judge marks (ticks off on the provided sheet) ‘essential items’ from the selectable pool for each performance level. The average difficulty (in logits) of the “essential items” constitutes the Criterion Point.

The rating process took approximately 50 minutes. When everybody indicated that they were ready to submit their completed rating forms, as a final quality check, the researcher reminded the panellists to check that firstly, their names appeared on each page of the completed rating form, secondly, they had not classified any item as simultaneously “essential” in more than one performance level, and thirdly that, items which they could not
classify as “essential” at any of the three performance levels were indicated with an asterisk (*) against the item number in the test form. The rating forms were then collected and checked for accuracy. Particular attention was paid to checking that no item was simultaneously placed as “essential” under two different performance levels.

**Stage Two: Refining Mastery Level**

This is a Group Activity to determine the level of **Mastery Level** or comprehensive competence required to achieve each performance level. Stone (2001:191) provides a guide on how panellists who use the OSS method should distinguish between a cut score and decide on “mastery” by asking the question: “Considering the nature of the professional being tested, how comfortable are you with a 50% comprehensive mastery? 60% comprehensive mastery? or 70% comprehensive mastery?” (Stone, 2001:191). In this study the question was adapted to ask: “Considering the purpose of ANA, how comfortable are you with a 50% comprehensive mastery? 60% comprehensive mastery? or 70% comprehensive mastery?” and the panellists agreed unanimously that they would be comfortable with a 60% Mastery Level in a test.

The last two stages, viz. “Stage Three: Ensuring Precision” of the cutscores and “Stage Four: Constructing the (final) Standard” (Stone, 2001:191-192) involved technical desktop work using specialised software (Winsteps Software) and were executed by the researcher in the course of calculating cutscores.

**3.7.2.3 Capturing of data**

An Excel spread sheet, with the names of the panellists in the columns and the test items in the rows, was prepared for capturing the ratings (identification of essential items) on the rating forms. In a parallel process, the item difficulties (in logits) for each test were calculated using the Winsteps Software Version 3.81.0 (Linacre, 2000), particularly selecting the Partial Credit Model (PCM) because each test had a mix of multiple-choice and open-ended items (Bond and Fox, 2007). The test item difficulties were then captured onto the prepared Excel spread sheet such that for each panellist and test item the tick (✓) on the relevant rating form was replaced with the calculated item difficulty. An extract of the captured scores (item
difficulties in logits) has been shown in Appendix I. In this format the data was now ready for analysis to determine cutscores.

3.7.2.4 Calculation of cutscores

Calculation of cutscores from the OSS method was done from the Excel sheet that was populated with item difficulty values per item per person (Appendix I). The first part was the calculation of mean item difficulties per panellist per level, viz. Not Achieved, Achieved and Advanced. This led to the calculation of the “Criterion Point”, i.e. the mean of all the panellists at a given level. The second part of the calculation involved the last two stages of this method, viz.

Stage 3: Ensuring Precision

This stage is the determination of the acceptable error band in the standard (i.e. Confidence Level). Stone (2011:192) recognises that, for any test score given to a person, the true ability of the person lies in the error band that surrounds the score. The error band is always there and needs to be accounted for in any sound measurement. If the test is trusted to be perfect, then a narrower Confidence Level, high level of strictness, is recommended to “protect the public” but if the test is known to be prone to error, then acceptance of a wider Confidence Level, low level of strictness, is recommended to “protect the individual” who takes the test.

This stage was considered to be already determined because the results of Annual National Assessments in South Africa are not used for high-stakes “pass/fail” purposes but provide information that is used to inform appropriate interventions to improve learning. A wider Confidence Level was, therefore, considered acceptable.

Stage 4: Constructing the final cut score

The final cut score for each level of performance was determined by adding the Criterion Point to the Mastery Level (both expressed in logits). To ensure low-stakes precision, the “Confidence Level” was subtracted from the sum of the Criterion Point and the Mastery Level. The procedure for calculating the Final Standard at each performance level has been summarised in Equation 2 in chapter 2.

Cut score = Criterion Point + Mastery Level ± Confidence Level x SEM  

   \(2\)
3.7.3 OBM SS method

Unlike the Angoff and the OSS methods that involve teams or panels of experts towards identifying cutscores, the OBM is a purely statistical method (Shulruf et al, 2014). Learner scores for each of grade 6 language and mathematics tests were subjected to the OBM method and cutscores were obtained immediately.

Like the OSS, the OBM method recognises and utilises the theory that the true ability of the test taker lies in the error band that surrounds the test score (Shulruf et al, 2014:3). For learners whose scores fall within the “error range” or borderline zone between clear “passes” and clear “failures”, the ubiquitous error creates the probability that borderline test-takers may be true failures or false passes. Shulruf’s et al (2014:6) statistical technique is designed to minimise such probable ambiguous decisions within the “range of error”.

The question of exactly how much “range of error” should be allowed in the OBM procedure to get precise cutscores was explored for purposes of this study. In response personal communication regarding how the size of the “error range” impacted on the precision of cutscores, Prof Shulruf (2015) indicated that he had checked different ‘ranges of error’ and the impact was minimal unless the range was huge. He then advised the author to work around the smallest acceptable range of error, extend the range and run the process again. If the differences were not statistically significant then one could assume that the range that was chosen was acceptable.

In this study the relationship between the “range of error” and the precision of the resulting cutscores was explored in the context of different datasets obtained from previous assessments conducted by the DBE and the patterns showed that, for values ranging between 1 and 15 score points, the changes in cutscores were not statistically significant and small changes in the “range of error” resulted in negligibly small changes in cutscores as shown in Appendix J. The “range of error” for this study was subsequently selected to be 10 score points around the true (ability) score and, using borderline zones at each of the Not Achieved, Achieved and Advanced levels, used Equation 3 to calculate corresponding cutscores.
3.8 Data simulations

To investigate the overall pattern and stability of the cutscores calculated from each of the OSS and OBM SS methods used in this study and, therefore, test the suitability and accuracy of these methods in setting standards in more complex situations than the real data from one event can show, Winsteps Software (Linacre, 2014) was used to generate datasets that simulated the real data. While for both the OSS and the OBM simulations could easily be based on the observed data, simulation of data for the Angoff method proved complex because of the person-intensive design of the method and the unpredictable variances that characterize human beings. Ricker (2006:57) observes that “The need for high interrater and intrarater reliability in order to validate the Angoff method presents a conundrum …” which complicates the possibility of simulating individual Angoff ratings and renders the exercise a project on its own outside the scope of this study.

For the OSS and the OBM the simulation proceeded by running Winsteps on the real dataset to produce estimates of all measures such as item difficulties and person abilities followed by the generation of estimate measures which resemble the observed real data. The following simulation conditions were specified in the program (Linacre, 2014:473):

- Use random numbers to generate simulated data;
- Generate simulated data by re-sampling from the observed data, i.e. simulated datasets to be equivalent to the real datasets;
- Resample persons, specified at 1 000 persons per simulated datafile;
- Generate a complete dataset, with no missing data; and,
- Avoid extreme scores (i.e. avoid ‘floor’ and ‘ceiling’ effects in the data by excluding zero and maximum scores, respectively).

Running Winsteps (Linacre, 2014) with data generated under the specified assumptions for each subject, three thousand (3 000) datasets in text files, with 1 000 re-sampled learners in each, were generated. All the 3 000 Winsteps output text files with simulated data were placed in one folder. To create one file that contained all the data in the 3 000 files and thus
be able to do analysis, the following MS-DOS batch routine (.bat) was followed under the
guidance of a specialist programmer:

a) Using NotePad, a text file called “extract.bat” was created and saved in the same
folder as the 3 000 data files.

b) MS-DOS commands (Appendix K) were written in the file “extract.bat” and used to
extract the same text line from all the 3 000 files and place the output into one new
text file which was named “result.csv”.

c) The command was executed by double-clicking on “extract.bat” which activated
“extract.bat” to pick out a text line from each of the 3 000 files and write the text lines
into a new text file “result.csv”.

d) The new text file “result.csv” was then imported into Excel.

e) Excel “Data”, followed by “Text to columns” were then used to split the text lines of
the new file “result.csv” into separate Excel columns.

f) Once in Excel, the consolidated datafile “result.csv” was now ready for analysis, by
using Equations 2 and 4 to calculate OSS and OBM cutscores, respectively.

3.9 Evaluation of the various SS methods

The Standard Setting Evaluation Instrument (SSEI) was used to evaluate each method of
standard setting. Each method was evaluated against critical criteria which were identified in
discussions with the SS participants on whether the SS method was undesirable (UD),
desirable (D) or highly desirable (HD), using the likely impact as a guide. Numerical codes
were subsequently assigned such that UD=0, D=1 and HD=2. An ordered comparison was
now possible where the order of the numbers, 0 to 2, represented an increasing order of
preference or desirability. A frequency count was conducted on the numeric codes to
determine total scores. The criterion with the highest frequency score is the highly desired
and one with the lowest score is the most undesirable.
The three SS methods, Angoff, OSS and OBM, were compared using the specially designed SSEI. After observations and discussions with other participants in the standard setting processes, comparison of the three methods proceeded as follows:

a. Each SS method was evaluated as to whether it was undesirable (UD), desirable (D) or highly desirable (HD) on each of the criteria on SSEI where desirability was defined in terms of the impact (stated in SSEI) the method would have on the NRF and education system at large.

b. Coding the agreements from professional discussions using Codes 0, 1 and 2 for UN, D and HD, respectively, such that UN=0, D=1 and HD=2.

c. Calculating the frequencies of each of the UN, D and HD and using the frequencies to make final decisions on the relative desirability of the three SS methods.

The resulting instrument, SSEI, is presented in Table 9.

Table 9: SSEI, the instrument used to evaluate SS methods

<table>
<thead>
<tr>
<th>No.</th>
<th>Criteria</th>
<th>Description</th>
<th>Research reference</th>
<th>Possible impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Resource intensity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Technical demands</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Precision of cutscores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Consistency of cutscores</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Content validity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Need for specialised software</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Turnaround time</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From Table 9, the key components of SSEI are the criteria used to evaluate the SS method (Column 2), a description of how the criterion is used (Column 3), the main references or sources of information on the criterion and the possible impact that the criterion in question may have if the corresponding SS method is adopted.

3.10 Necessary conditions for data use

Data to help answer the question on appropriate guidelines for an NRF was collected from research that met one or both of the following pre-set conditions:

a. Involves samples of at least 100 schools; and,
b. Presents theory(ies) of learning (and teaching) that involve data use.

A summary of studies that met the criteria, their designs and their findings was used to identify guidelines that could be considered for an NRF in the South African context.

3.11 Survey on reporting results using performance standards

The survey involved administration of a self-completed questionnaire to national and provincial education officials to determine their perceptions/views on the use of performance standards compared to raw scores in reporting results from large-scale assessment surveys. The questionnaire was distributed by e-mail with specific guidelines on its purpose and how the respondents were expected to complete the questionnaire. The respondents were requested to complete the questionnaire and e-mail their responses back to the researcher and were given deadlines within which to respond.

A common responsibility among provincial officials was that they all interacted directly with schools and teachers in curriculum- and assessment-related professional activities. They served on, and represented their respective provinces in, curriculum and assessment structures that were charged with advising the DBE on relevant policy and policy implementation issues. Each of the nine (9) South African provinces is represented by at least one official in each of the curriculum and assessment structures. Together with their national counterparts, these officials hold bi-monthly meetings where they reflect on policy implementation issues and, when necessary, make recommendations to the DBE for the review of existing policies or make proposals for development of new policies. A minimum of four (4) officials drawn from curriculum management, teacher development, public examinations and national assessments, represent each province on this forum.

Due to their liaison and facilitating role in the provinces, these provincial officials, together with their national counterparts, constituted a convenient purposive, but not necessarily representative, sample for the validation of the NRF. The respondents, who comprised both national and provincial officials, were requested to complete the questionnaire by indicating a reporting template that they would prefer for different uses.

The number and responsibilities of officials who responded to the questionnaire on the standards-based and raw score-based reporting formats have been summarised in Table 10.
Table 10: Distribution and responsibilities of respondents to questionnaire

<table>
<thead>
<tr>
<th>Province</th>
<th>Areas of responsibility</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assessment</td>
<td>Curriculum</td>
</tr>
<tr>
<td>Eastern Cape</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Free State</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Gauteng</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>KwaZulu-Natal</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Limpopo</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mpumalanga</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Northern Cape</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>North West</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Western Cape</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>DBE</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>Percentage</td>
<td>46.0</td>
<td>23.1</td>
</tr>
</tbody>
</table>

From Table 10, a total of 39 responses were received from seven (7) of the nine (9) provinces of South Africa. Officials from the Eastern Cape and Limpopo were not available to respond to the questionnaire at the time of data collection. Also not available during the period were officials from Curriculum and Teachers Development in each of the provinces of Free State, KwaZulu-Natal, Mpumalanga and North West while from Northern Cape non-participation was from the Assessment and Teacher Development Units. In total 46% of the participants were from Units that were responsible for assessment activities, 23.1% from Curriculum Units and 30.8% were in Teacher Development sections.

3.12 Statistical tests of significance

At a number of stages in the design of this study there were specific quantitative comparisons to be made. For instance, cutscores were ranked and compared across various SS methods and between real and simulated datasets. Comparisons were made to test whether there were differences in impact of cutscores from different SS on the distribution of learners across PLs in each subject. To avoid making Type I errors in the comparisons, i.e. accepting a difference when there is no difference between two statistics, statistical tests of significance were conducted, where appropriate (Nachar, 2008:14-15, 20). Non-parametric tests were the most appropriate for the data in this study because the datasets that were compared did not meet the requirements of continuous and normally-distributed data and the samples were very small (Mann & Whitney, 1947; Nachar, 2008:13). The Mann Whitney U Test, which is a
non-parametric version of the parametric independent samples t-test, was used to test for statistical significance of differences between cutscores generated from two different populations of SS methods at a given time. For comparing the impact of cutscores generated from different SS methods the Chi-square test was used since the data was categorical (Neuman, 2000, Monahan & Ankenmann, 2005).

Basically each of the two tests is used to test the null hypothesis that there is no difference between two given statistics against the alternative hypothesis that there is a difference (Nachar, 2008:14-15). Because there was no interest in the direction of differences, where they existed, two-tailed tests were conducted in all the statistics that were compared.

3.13 Summary

In this chapter the design of the study was described in detail and the reasons behind the choice of methodology substantiated, which mainly related to the exploratory nature of this study in order to obtain relevant explanations to the research questions of interest. It was indicated in the chapter that the absence of readily available researched techniques and data collection instruments to answer specific aspects of the study provided a unique opportunity for consideration of the most appropriate approaches to addressing the concerns of this study. Consequently, the study has been set in a rigidly chronological sequence allowing for use of different instruments and analysis techniques at each stage.

A detailed description of the data collection instruments, how they were either adapted or developed specifically for the study as well as accounts of the validity of respective results were presented. Because the majority of the data was collected through different standard-setting exercises, a substantial part of the chapter was dedicated to describing the participants and their profiles, the technical details of how each SS method was executed and how the cutscores were calculated.

Included in the chapter are measures that were taken to evaluate each standard setting method and also to examine perceptions of education officials on formats of reporting assessment performance in terms of performance standards. There is an elaborate account of how a specific instrument was developed and utilised to evaluate each SS process according to agreed criteria among the participants. Similarly, the process of developing and using a
specialised instrument to measure the views or perceptions of education specialists on the appropriateness of the performance standards in reporting learner performance was described in fair detail.

Finally, a description of the statistical methods that were used to test the significance of differences among cutscores and impact data that were generated from different SS methods was given with well-motivated reasons for the choice of the particular tests. The study findings are presented in Chapter 4.
4. RESEARCH FINDINGS

4.1 Introduction

The purpose of this study was to develop a standards-based National Reporting Framework (NRF) that would guide the recording, analysis, reporting and utilisation of data from large-scale assessment surveys (LSAS) to enhance teaching and learning towards improvement of learner performance. Chapter 3 described the methodology that was used in the research design, the data collection and the data analyses that were adopted in this study. This chapter presents the findings of the study. Findings related to each research question are presented in numerical and graphical formats followed by pertinent discussions.

4.2 Findings on Research Question 1

This research question made a comparison between the Angoff and Objective Standard Setting (OSS) method, as well as the Objective Method (OBM) where appropriate, to inform a choice for the most appropriate method applicable to the South African context. The standard setting (SS) methods, which were applied using mathematics and English First Additional Language (FAL) data, were compared in terms of the following criteria:

a) The similarities or differences of their generated cutscores at various performance-levels (PLs);

b) The consistency of the cutscores when applied to simulated data, thus indicating whether or not they could be generalised to different scenarios;

c) The impact of the various cutscores on the learners’ categorisations in each subject;

d) The extent to which members of the SS panels found the method desirable and preferred.

Research Question 1

Which of the three standard setting methods: the Angoff, Objective Standard Setting and Objective Borderline Method, is most appropriate for developing performance standards that are relevant for a National Reporting Framework in the South African context?
To answer this question, standard setting sessions were convened in Pretoria between 2013 and 2014 and teams of educators participated in determining PLs and cutscores and developing performance standards using each of the Angoff and the OSS methods. Standard setting on the OBM was executed by the researcher as a desktop exercise. This technique, unlike most SS methods, does not require panels. Using the relevant SS methods as described in Chapter 3, cutscores from real data were calculated for each dataset, namely the mathematics and language datasets. The results were presented as percentages in bar graphs for each subject and SS method.

4.2.1 Mathematics cutscores

Mathematics cutscores were calculated from the item ratings that were completed by eight (8) panellists using the Angoff method of SS and eight (8) others using the OSS method. One of the panellists in the OSS method produced ratings and cutscores that violated the “golden thread” of hierarchical increase from the Not Achieved through the Advanced level. The ratings that were produced by this panellist are shown in Table 11.

Table 11: Mathematics cutscores that violate the PL hierarchy

<table>
<thead>
<tr>
<th>Moss (not a real name)</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cutscores in logits</td>
<td>-0.82</td>
<td>0.34</td>
<td>0.31</td>
</tr>
<tr>
<td>Cutscores in percentages</td>
<td>30.6</td>
<td>58.5</td>
<td>57.8*</td>
</tr>
</tbody>
</table>

From Table 11, Moss (not a real name) misjudged “essential” items at the Achieved and Advanced levels, 58.5% and 57.8%, respectively, resulting in cutscores that violate the fundamental principle of the hierarchical structure of PLs according to which the cutscore at the Advanced level should be higher than the cutscore at the lower Achieved level and not the other way round. Consequently, the OSS ratings from this particular panellist were excluded from the final calculation of cutscores, leaving seven (7) compliant panellists in the group.

Mathematics cutscores that were calculated from the ANA data using the Angoff \((n_1 = 8)\), OSS \((n_2 = 7)\) SS methods and verified with the OBM are shown in Figure 12.
From Figure 12, at the *Partly Achieved level* mathematics cutscores from the Angoff and OSS methods were fairly similar but substantially higher for the OBM. However, at both the *Achieved* and *Advanced levels*, cutscores from the Angoff and OBM were similar but substantially lower for the OSS method. A Mann-Whitney U test (two-tailed test) that was used to examine the statistical significance of the differences among the SS methods indicated that:

a) At the *Partly Achieved level*, the median mathematics cutscore from the OSS method (Mdn = 27.7) was not different from the median cutscore from the Angoff method (Mdn = 24.5), $U = 22.0, p = .05$. Also, the median cutscore from the Angoff method (24.5) was not different from the cutscore generated from the OBM (Mdn = 35.0), $U = 9, p = .05$ but the median cutscore from the OSS method (Mdn = 27.7) was different from the cutscore generated from the OBM method (Mdn = 35.0), $U = 0.0, p = .05$.

b) At the *Achieved level*, the median mathematics cutscore from the Angoff method (Mdn = 56.7) was different from the median cutscore from the OSS method (Mdn = 47.5), $U = 10.0, p = .05$). Also, the median cutscore from the OSS method (Mdn = 47.5) was different from the cutscore generated from the OBM (Mdn = 57.0), $U = 1.0,$
At the Advanced level, the median cutscore from the OSS method (Mdn = 61.1) was different from the median cutscore from the Angoff method (Mdn = 72.4), $U = 0.0$, $p = .05$. Also, the cutscore from the OSS method (Mdn = 61.1) was different from the cutscore generated from the OBM method (Mdn = 76.0), $U = 0.0$, $p = .05$ but the median cutscore from the OBM (Mdn = 76.0) was not different from the cutscore generated from the Angoff method (Mdn = 72.4), $U = 11.0$, $p = .05$.

The findings on whether the three SS methods generated mathematics cutscores that were different or not different at each performance level are summarised in Table 12.

<table>
<thead>
<tr>
<th>Difference between cutscores from</th>
<th>Is the cutscore difference significant at each PL?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Partly Achieved</td>
</tr>
<tr>
<td>OSS &amp; OBM</td>
<td>Yes</td>
</tr>
<tr>
<td>OSS &amp; Angoff</td>
<td>No</td>
</tr>
<tr>
<td>Angoff &amp; OBM</td>
<td>No</td>
</tr>
</tbody>
</table>

From Table 12, the OSS method & OBM consistently generate mathematics cutscores that are different at the Partly Achieved, Achieved and Advanced levels. Also, the OSS & Angoff methods generally generate mathematics cutscores that are different except at the Partly Achieved level where the two SS methods generate mathematics cutscores that are not different. However, the Angoff & OBM consistently generate mathematics cutscores that are not different at each of the three PLs. It can be concluded, therefore, that, while the Angoff method & OBM may be used interchangeably to set mathematics cutscores because they tend to generate cutscores that are not different, the OSS & OBM as well as the OSS & Angoff may not be used interchangeably because they tend to generate mathematics cutscores that are different.

### 4.2.2 Language cutscores

English FAL cutscores were calculated from the item ratings that were completed by five (5) panellists using the Angoff method of SS and eight (8) others using the OSS method. Three
(3) of the panellists in the OSS method produced ratings and cutscores that violated the “golden thread” of hierarchical increase from the Not Achieved through the Advanced level. By definition of PLs the cutscore at the Achieved level should be higher than the cutscore at the Partly Achieved level and the cutscore at the Advanced level should be the highest. The ratings that were produced by these panellists are shown in Table 13.

**Table 13: FAL cutscores that violate the PL hierarchy**

<table>
<thead>
<tr>
<th></th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lina (not a real name)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cutscores in logits</td>
<td>-0.30</td>
<td>0.55</td>
<td>0.41*</td>
</tr>
<tr>
<td>Cutscores in percentages</td>
<td>42.6</td>
<td>63.3</td>
<td>60.2*</td>
</tr>
<tr>
<td>Minie (not a real name)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cutscores in logits</td>
<td>0.66*</td>
<td>-0.40</td>
<td>0.16</td>
</tr>
<tr>
<td>Cutscores in percentages</td>
<td>65.8*</td>
<td>40</td>
<td>54.1</td>
</tr>
<tr>
<td>Rose (not a real name)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cutscores in logits</td>
<td>-0.12</td>
<td>-0.16*</td>
<td>0.16</td>
</tr>
<tr>
<td>Cutscores in percentages</td>
<td>46.9</td>
<td>46.1*</td>
<td>54.1</td>
</tr>
</tbody>
</table>

From Table 13, Lina (no a real name) misjudged “essential” items at the Achieved and Advanced levels, 63.3% and 60.2%, respectively, Minie (no a real name) misjudged “essential” items at the Partly Achieved and Achieved levels, respectively, and Rose (not a real name) also misjudged “essential” items at the Partly Achieved and Achieved levels, respectively. Consequently, the OSS ratings from these three (3) panellists were excluded from the final calculation of cutscores, leaving five (5) compliant panellists in the group.

Language cutscores were also calculated from the ANA data using the Angoff ($n_1 = 5$), OSS ($n_2 = 5$) SS methods and were verified using the OBM. The cutscores are shown in Figure 13.
Figure 13: Language scores calculated from three SS Methods

As noted in Figure 13, the language or FAL cutscores from the three SS methods were dissimilar at all three PLs. The minimum difference in cutscores was observed at the *Partly Achieved level* between the Angoff and OBM cutscores. To examine the statistical significance of the differences in cutscores from the different SS methods, a Mann-Whitney U test of significance was used. The Mann-Whitney U test indicated that:

a) At the *Partially Achieved* level, the median FAL cutscore from the OSS method (Mdn = 39.7), was different from the cutscore from the Angoff method (Mdn = 33.2), U = 0.0, p = .05. However, the median cutscore from the Angoff method (Mdn = 33.2), was not different from the cutscore generated from the OBM (Mdn = 33.0), U = 9.0, p = .05 but the cutscore from the OBM (Mdn = 33.0) was different from the cutscore generated from the OSS method (Mdn = 39.7), U = 0.0, p = .05.

b) At the *Achieved level*, the median cutscore from the Angoff method (Mdn = 64.6) was different from the median cutscore obtained from the OSS method (Mdn = 50.5), U = 0.0, p = .05 and also different from the cutscore from the OBM (Mdn = 55.0), U = 0.0, p = .05. But the median cutscore from the OSS method (Mdn = 50.5) was not different from the median score from the OBM (55.0), U = 0.0, p = .05.

c) At the *Advanced level*, the median cutscore from the Angoff method (Mdn = 85.2) was different from the median cutscore from the OSS method (Mdn = 62.2), U = 0.0,
\[ p = .05 \text{ and also different from the median cutscore from the OBM (Mdn = 76.0), } U = 0.0, p = .05. \] Similarly, the median cutscore from the OBM (Mdn = 76.0) was different from the median cutscore from the OSS method (Mdn = 62.2), \( U = 1.0, p = .05. \)

The findings on whether the three SS methods generated FAL cutscores that were different or not at each performance level are summarised in Table 14.

**Table 14: Summary of significant differences among FAL cutscores by SS method**

<table>
<thead>
<tr>
<th>Difference between cutscores from</th>
<th>Is the cutscore difference significant at each PL?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Partly Achieved</td>
</tr>
<tr>
<td>OSS &amp; Angoff</td>
<td>Yes</td>
</tr>
<tr>
<td>OBM &amp; Angoff</td>
<td>No</td>
</tr>
<tr>
<td>OBM &amp; OSS</td>
<td>Yes</td>
</tr>
</tbody>
</table>

From Table 14, the three SS methods generally generate FAL cutscores that are different at the different PLs. Exceptions were the OBM & Angoff and the OBM & OSS which generated cutscores that were not different at the *Partly Achieved* and the *Achieved* levels, respectively. It can be concluded, therefore, that the Angoff, OSS and OBM SS methods may not be used interchangeably to set FAL cutscores. Other criteria need to be explored that can be used to identify the best method for a specific context.

### 4.2.3 Cutscores from real and simulated data

To validate each of the SS methods, the calculation of cutscores for each of mathematics and language were repeated using simulated data. The purpose of the validation process was to determine whether for each SS method, and at each performance level, the cutscores from real and simulated data were different or similar. The thesis of this study is that if the same SS method generates similar cutscores from real and simulated datasets, a logical conclusion could be that the SS method is consistent across different contexts and, therefore, can be regarded to be dependable (Neuwman, 2000:170). However, if a SS method generates different cutscores from real and simulated datasets, this could imply that the SS method is context-sensitive and, therefore, may not be dependable and suitable for an NRF.

As indicated in Section 3.8, learner data were simulated using the Winsteps program. For the Angoff method, simulation of persons to generate simulated item ratings proved to be
complex and was not undertaken. Thus, in this study only the OSS method and the OBM were validated by comparing cutscores from real and simulated data.

4.2.3.1 Mathematics cutscores from real and simulated data

Mathematics cutscores from both real and simulated data from the OSS method, \( n_1 = 7 \) and \( n_2 = 7 \), respectively, and the OBM, \( n_1 = 5 \) and \( n_2 = 5 \), respectively, are shown in Figure 14.

![Mathematics cutscores calculated from real and simulated data](image)

**Figure 14: Mathematics cutscores calculated from real and simulated data**

From Figure 14 the mathematics cutscores calculated using the OSS method on real data tended to be lower than corresponding cutscores calculated from simulated data at all the PLs. For the OBM, there were very striking similarities among corresponding cutscores from real and simulated data at all PLs. For each SS method, corresponding cutscores from real and simulated data were examined for statistical significance of differences.

From the OSS method, a Mann-Whitney U test indicated that:
a) at the Partially Achieved level, the mathematics median cutscore from the real data (Mdn = 27.7) was not different from the median cutscore obtained from simulated data (Mdn = 31.9), \( U = 11.0, p = .05 \).

b) at the Achieved level, the mathematics median cutscore from the real data (Mdn = 47.5) was not different from the median cutscore obtained from simulated data (Mdn = 53.7), \( U = 12.0, p = .05 \).

c) at the Advance level, the mathematics median cutscore from the real data (Mdn = 61.1) was not different from the median cutscore obtained from simulated data (Mdn = 65.2), \( U = 8.5, p = .05 \).

Similarly, using the OBM, a Mann-Whitney U test indicated that:

a) at the Partially Achieved level, the mathematics median cutscore from the real data (Mdn = 35.0) was not different from the median cutscore obtained from simulated data (Mdn = 35.0), \( U = 11.5, p = .05 \).

b) at the Achieved level, the mathematics median cutscore from the real data (Mdn = 57.0) was not different from the median cutscore obtained from simulated data (Mdn = 56.0), \( U = 7.5, p = .05 \).

c) at the Advanced level, the mathematics median cutscore from the real data (Mdn = 76.0) was not different from the median cutscore obtained from simulated data (Mdn = 75.0), \( U = 12.0, p = .05 \).

Overall, mathematics cutscores generated from both OSS method and OBM, using real data and data simulated to represent different scenarios, were not different. This suggests that the two SS methods can be used dependably to generate valid mathematics cutscores across different contexts. One could conclude that, on the basis of dependability, each of the OSS and the OBM could be a candidate for inclusion in an NRF that will regulate reporting different assessments at different times.
4.2.3.2 Language cutscores from real and simulated data

Language (FAL) cutscores from both real and simulated data from the OSS method, n₁ = 5 and n₂ = 5, respectively, and the OBM, n₁ = 5 and n₂ = 5, respectively, are shown in Figure 15.

![Figure 15: Language cutscores calculated from real and simulated data](image)

From Figure 15 it can be observed that, although for real data corresponding mathematics cutscores from the OSS method and the OBM tended to be different, corresponding cutscores from simulated for the two SS methods were remarkably similar. For instance, at the Partly Achieved level each of the two SS methods generated 37% as the simulated mathematics cutscore while corresponding cutscores at the Advanced level were also very close at 75% and 77%, respectively. The Mann-Whitney U test was used to test the differences among corresponding real and simulated language cutscores for statistical significance.

For the OSS method, a Mann-Whitney U test indicated that:
Q. Moloi

a) at the *Partially Achieved level*, the FAL median cutscore from the real data (Mdn = 39.7) was not different from the median cutscore obtained from simulated data (Mdn = 39.2), $U = 10.0, p = .05$.

b) at the *Achieved level*, the FAL median cutscore from the real data (Mdn = 50.5) was not different from the median cutscore obtained from simulated data (Mdn = 51.0), $U = 11.0, p = .05$.

c) at the *Advanced level*, the FAL median cutscore from the real data (Mdn = 62.2) was not different from the median cutscore obtained from simulated data (Mdn = 62.7), $U = 12.0, p = .05$.

Using the OBM, a Mann-Whitney U test indicated that:

a) at the *Partially Achieved level*, the FAL median cutscore from the real data (Mdn = 33.0) was not different from the median cutscore obtained from simulated data (Mdn = 34.7), $U = 11.5, p = .05$;

b) at the *Achieved level*, the FAL median cutscore from the real data (Mdn = 55.0) was not different from the median cutscore obtained from simulated data (Mdn = 54.7), $U = 12.0, p = .05$; and,

c) at the *Advanced level*, the FAL median cutscore from the real data (Mdn = 76.0) was not different from the median cutscore obtained from simulated data (Mdn = 76.0), $U = 12.0, p = .05$.

As was the case with mathematics, language cutscores generated from both OSS method and OBM, using real data and data simulated to represent different scenarios, were not different. Again this suggests that the two SS methods can be used dependably to generate valid language cutscores across different contexts, which is a necessary condition for a SS method if it will be included in an NRF that will regulate reporting different assessments at different times.
4.2.4 Impact of cutscores

Standard practice in SS is to evaluate possible impact of cutscores on the affected sampled learners and, by extension, on the target population of learners (Hambleton, 2001). The evaluation involves determining the percentage of learners who will fall into the various categories or PLs marked by the cutscores. Possible differential impact of the Angoff, OSS method and the OBM cutscores was investigated and estimated in terms of the numbers of learners categorised into the various PLs as shown in Figure 16 for mathematics (n=8 131).

![Figure 16: Distribution of mathematics learners across the performance levels](image)

From Figure 16, mathematics cutscores from each of the Angoff, OSS and OBM categorise the majority of learners at the **Not Achieved** and **Partly Achieved levels** while the smallest number of learners are categorised at the **Advanced level**. Cutscores from the Angoff and the OSS methods categorise equal numbers of learners, 2 229 or 27.4%, at the **Not Achieved level**. The highest number of learners, 4 567 or 56.2%, are categorised at the **Partly Achieved level** by the Angoff method cutscore while the cutscore from the OBM categorises the lowest number of learners, 417 or 5.1%, at the **Advanced level**.

A chi-square test of independence was calculated to examine whether there was a relation (two-tailed test) between the categorisation of learners into the four PLs and the SS method that generated the cutscores. There was a significant relation between the frequency of learners in each category and the SS method that was used to generate the categorising
Q. Moloi

mathematics cutscores, $\chi^2(6, N = 8\,131) = 1\,460.3, \ p \leq 0.05$. The categorisation of learners into the PLs depended on the SS method used to generate the mathematics cutscores.

Fitting regression lines from Microsoft Excel for each data related to a particular SS method and displaying R-squared values in Figure 16 provided some measure, albeit inconclusive, of variation in the data from the modelled relationship in each SS method. The modelled relationship is that of the cutscore associated with a SS method and the resulting number of learners in each PL. The R-squared is defined as the percentage of the dependent variable variation that is explained by a linear model (Neuwman, 2000:337).

The R-squared values related to the three SS methods used to set mathematics cutscores were $R_A^2=0.388$, $R_{OSS}^2=0.376$ and $R_{OB}^2=0.948$ for the Angoff, OSS and OBM, respectively. To avoid clutter, only the regression line and the relevant R-squared value related to the OBM, which was the highest of the three, are shown in Figure 16. According to the definition of R-squared, the regression line explains 95% of the OBM learner data variation. In the other two SS methods the R-squared values show that the relevant regression line explains 39% of the Angoff data and 38% of the OSS data. All other things being equal, the OBM would, therefore, be the preferred SS method for mathematics since its related model explains the highest amount of variation or has the best fit to the model.

The possible impact of cutscores that were calculated for language (n=6\,106) using the Angoff, OSS and OBM, is presented in Figure 17.

![Figure 17: Distribution of language learners across the performance levels](image-url)
From Figure 17, language cutscores from each of the Angoff, OSS and OBM categorise the majority of learners at the *Not Achieved* and *Partly Achieved* levels while the smallest number of learners are categorised at the *Advanced* level. Cutscores from the Angoff method categorise the highest number of learners, 3,510 or 57.5%, at the *Partly Achieved* level and the lowest, 52 or 0.9%, at the *Achieved* level. At the *Achieved* level the OBM categorises the highest number of learners, 1,488 or 24.4%.

A chi-square test of independence was calculated to examine whether there was a relation (two-tailed test) between the categorisation of learners into the four (4) PLs and the SS method that generated the language cutscores. There was a significant relation between the frequency of learners in each category and the SS method that was used to generate the categorising cutscores, $\chi^2(6, N = 6106) = 2,527.4$, $p \leq .05$. The categorisation of learners into the PLs depended on the SS method used to generate the language cutscores.

Fitting regression lines for data related to each of the three SS methods and displaying $R$-squared values as in Figure 17 showed $R$-squared values of $R_A^2 = 0.429$, $R_{OB}^2 = 0.594$ and $R_{OSS}^2 = 0.786$ for the Angoff, OBM and OSS, respectively. To avoid clutter, only the regression line and the relevant $R$-squared value related to the OSS, which was the highest of the three, are shown in Figure 17. According to the definition of $R$-squared, the regression line explains 79% of the OSS learner data variation. In the other two SS methods, the $R$-squared values show that the relevant regression line explains 43% of the Angoff data and 59% of the OBM data. All other things being equal, the OSS would, therefore, be the preferred SS method for language since its related model explains the highest amount of variation.

An important observation that can be made from the use and interpretation of $R$-squared values is that while the OBM seems to be the preferred method for explaining the impact of cutscores in mathematics, for language the OSS appears to be the best SS method. This raises an important question as to whether the best SS method should apply across subjects or different methods may work better in different subjects.

A summary of whether the differences among the cutscores from the Angoff and OSS SS methods and their differential impact, on the distribution of learners according to performance, were statistically significant is presented in Section 4.3.
4.2.5 Summary of key findings on cutscores

From the comparison of cutscores generated by the three SS methods that were the focus of this study; namely, the Angoff, OSS and OBM, the following summary highlights the key findings.

4.2.5.1 Mathematics cutscores

On the one hand, for mathematics, the Angoff method & the OBM consistently generated cutscores that were not different across the three PLs, viz. the Not Achieved, Partly Achieved and Advanced levels. This suggests that, based on cutscores alone, the Angoff method and the OBM may be used interchangeably to set mathematics cutscores and, therefore, either of the two or both could be included in the NRF since neither leads to any unique cutscores. Other criteria are necessary to make an informed definite choice between the two SS methods.

On the other hand, the OSS method compared to the OBM and also to the Angoff method generated different cutscores among them and across the PLs. This suggests that, based on cutscores alone, the OSS method may not be used interchangeably with any of the other two SS methods because each pair tends to generate mathematics cutscores that are different. Other criteria must be used to identify the most suitable SS method for the South African context.

4.2.5.2 Language cutscores

Overall, the Angoff, OSS and OBM SS methods generated different corresponding language cutscores among them, and based on cutscores only, the three SS methods may not be used interchangeably to set FAL cutscores. Other criteria need to be explored that can be used to identify the best method for inclusion in the NRF that is designed for a specific context.

4.2.6 Reliability test for the SS methods

A SS method was deemed to be reliable and, therefore, dependable, if it generated similar corresponding cutscores on real and simulated datasets.
4.2.6.1 Reliability of SS method for mathematics

Overall, mathematics cutscores generated from both OSS method and OBM, using real data and data simulated to represent different scenarios, were not different. This suggests that the two SS methods can be used dependably to generate valid mathematics cutscores across different contexts. One could conclude that on the basis of dependability, each of the OSS and the OBM could be a candidate for inclusion in an NRF that will regulate reporting different assessments at different times. The Angoff SS method was not subjected to this test.

4.2.6.2 Reliability of SS method for FAL

As was the case with mathematics, language cutscores generated from both OSS method and the OBM, using real data and data simulated to represent different scenarios, were not different. This suggests that one or both of the two SS methods can be used dependably to generate valid language cutscores across different contexts which is a necessary condition for a SS method if it will be included in an NRF that will regulate reporting different assessments at different times. The Angoff SS method was not subjected to this test.

4.2.7 Impact of cutscores

The impact of cutscores was defined in terms of the number or proportion of learners that get categorised into particular PLs if the cutscore were to be mandatory. There was a significant relationship between the cutscore generated by a SS method and the percentage or proportion of learners that the cutscore categorises into different PLs. For each of mathematics and FAL the categorisation of learners into the PLs depended on the SS method used to generate the relevant cutscore. An important observation that was made, and needs a more systematic approach to verify, is whether it would be best to use different methods for different subjects in setting performance standards.

4.2.8 Evaluation of the SS methods by panellists

The panellists in each SS session were asked to evaluate each SS method and were given criteria against which the evaluation was to be done as was described in Section 3.9. Using the same criteria and standard setting evaluation instrument (SSEI), the author also evaluated the OBM in which the other panellists were not involved. The coded results of the SS
evaluation on SSEI are summarised in Table 15. The codes ranged from 0 to 2 where 0 indicates an undesirable and, therefore, non-preferred attribute of the SS method and 2 indicates the most desirable and most preferred attribute.

Table 15: The SSEI results on the Angoff, OSS and OBM SS methods

<table>
<thead>
<tr>
<th>No.</th>
<th>Criteria</th>
<th>SS Method (coded)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Angoff OSS OBM</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Resource intensity</td>
<td>0 0 1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Technical/conceptual demands</td>
<td>0 0 1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Precision of cutscores</td>
<td>0 1 2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>Consistency of cutscores</td>
<td>0 1 2</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>Content validity</td>
<td>1 2 0</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>Need for specialised software</td>
<td>2 0 2</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>Turnaround time</td>
<td>0 1 2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 5 10</td>
<td></td>
</tr>
</tbody>
</table>

From Table 15 the SSEI scores on the Angoff, OSS and OBM were 3, 5 and 10 indicating that the OBM had the highest number of desirable features while the Angoff method had the least. Details of distinguishing features of each method are presented below.

- **Resource intensity**

According to the panellists, the Angoff method was the least desirable in terms of resource intensity. It was found to be very costly. The standard setting process lasted for 2 days (16 hours) requiring overnight accommodation, travel expenses and daily allowances for panellists. Four capturers were employed to ensure capturing was accurate and completed promptly. Two senior researchers offered their technical services at no cost, but this remains a potential cost that must be included in the budget if this method is to be used.

The OSS session lasted for a half day (four (4) hours). Capturing was done later and it took one capturer another half day to complete (four (4) hours). The OBM is a purely desktop exercise and was completed in an average of four (4) hours. However, both the OSS and OBM required specialised expertise and experience in the use of IRT (Rasch) analysis.
Technical/conceptual demands

The Angoff method required high levels of technical and conceptual skills. Programming the capturing spreadsheets and ensuring the consistency and accuracy of the calculations necessitated a level of skill that is not readily available. The panellists had to remain conceptually focused to consistently operationalize the definition of a “minimally competent borderline candidate”. The group discussions that were interspersed between the rating tasks required a different dimension of conceptual and social engagement which brought about necessary change but had its own demands.

Data capturers had to be fast and accurate because they had to complete the capturing at certain fixed intervals, i.e. during tea and meal times so as not to disrupt the flow of work.

The OSS method requires highly specialised technical skills, e.g. only a person with knowledge of the Rasch analysis and the software that is specially designed for this methodology could do the analysis. To a much lesser degree, the OBM also requires some level of computer skills (using Microsoft Excel) but a fairly high statistical competency level. However, once the concept of “borderline candidates” has been grasped the method becomes routine.

Precision of cutscores

Precision was used to refer to the extent of variations in score values when ratings were iterated. Precision of Angoff cutscores was observed to be influenced largely by intra- and inter-rater consistencies. Angoff cutscores stabilise mainly after the third round of ratings intermitted with group discussions, most probably because of power of persuasion from among panel members. Iterations did not change OSS cutscores in any significant way, most probably because specialists tended to be consistent in what they considered to be “essential content” which necessarily meant that calculated cutscores remained robust. As it would be expected, OBM cutscores were the most precise because, by design, this technique is meant to eliminate uncertainties at the borderline zone of cutscores and does not involve any human judgement.
• **Consistency of cutscores**

Consistency of cutscores was used to refer to changes in cutscores from one subject to another or from one context to another, e.g. between different contexts as represented by real and simulated data. All three SS methods were found to be reasonably consistent within and across subjects. For instance, all three methods showed that the majority of learners were functioning at the Not Achieved level in mathematics but in language the majority of the learners were at the Partly Achieved level. This means that, from the three SS methods, a consistent observation is that learner competencies in language are generally higher than in mathematics.

• **Content validity**

Both the Angoff and the OSS methods require inputs from curriculum specialists. Standard setting in the Angoff method was, more than the OSS, preceded by a curriculum-intensive process of developing policy definitions, setting both performance levels and performance descriptors, as well as taking the testing to familiarise with its specific content. Not only do panellists have to discuss curriculum content at the beginning of the respective sessions, but in the case of the Angoff method they also need to keep focused on the skills and content knowledge that a minimally competent learner requires at a given level. The resulting performance standards from the Angoff method are steeped in the curriculum and, therefore, have high levels of content validity. In constantly having to think about “essential” items at each level, panellists in the OSS method also bring their curriculum knowledge to bear on the standard setting process and the resulting performance standards have high levels of validity.

The OBM is a purely statistical procedure that is focused on eliminating measurement errors and uncertainties around the cutscore border zone. No consideration is given to curriculum content in this SS method. As such, the OBM can be best utilised in verifying cutscores from another SS method rather than being used on its own.

• **Turnaround times**

The turnaround times for each SS method were determined from the beginning of the orientation session to the final cutscores, without including time it takes to refine and present
the cutscores in graphs or other formats. The Angoff process lasted for 48 hours, the OSS for 8 hours and the OBM took an average of 4 hours.

4.2.9 Summary of findings from Research Question 1

In both mathematics and FAL the OSS and Angoff SS methods generally generate cutscores that are different, with one exception observed at the Partly Achieved Level in mathematics. The two methods also have different impact on distribution of learners according to performance, entirely so in language but only partly in mathematics. The two methods may not, therefore, be used interchangeably. The OSS generates cutscores that are similar across different contexts and scenarios as represented by the simulations that were devised for this study. The consistency of the OSS method across different contexts was verified through the OBM. Of course, a limitation of this study was that appropriate simulations for individual Angoff scores were not feasible. Until this limitation could be addressed, the generalisability of the OSS method renders this method more preferable than the Angoff method.

An evaluation of the three different SS methods in terms of selected criteria for efficacy showed that the OSS was, in most respects, more beneficial than either the Angoff method or the OBM. Specifically, the findings show that:

a) Compared to the Angoff method, the OSS method makes relatively low demands on finances and personnel time;

b) The OSS technical and conceptual processes are easier to be understood and implemented by panellists resulting in reduced variability of cutscores compared to the Angoff method; and,

c) Of the three methods compared in this study, the OSS method generates cutscores that are the richest and the most appropriate in terms of curriculum and subject content, followed by the Angoff method.

Taking all issues into consideration, the overall finding in this study is that the OSS was the most appropriate SS method for developing performance standards that are relevant to the South African context.
4.3 Findings on Research Question 2

Flowing from the first research question, the second question was intended to elicit information on guidelines that could be provided to provinces, districts and schools on how to use data within the NRF and within a standard-setting paradigm.

**Research Question 2**

What guidelines should be provided to national, provinces, districts and schools to enhance the utilisation of information from LSAS?

4.3.1 Conditions for data use

To answer this research question, literature was reviewed from relevant studies that met two main criteria, viz. they involved samples of a minimum of 100 schools and focussed on improving teaching and learning that involved data use. The rationale was that success in smaller numbers of schools would fail to provide lessons related to possible problems of scale. In the South African context there are districts with more than 300 schools.

A summary of literature reviewed to identify conditions that research shows to be conducive or detrimental to data use by schools and teachers has been presented in Table 6. In Table 6 the relevant study is identified, the design and purpose of each study is also presented, and the conditions that either enabled or constrained use of assessment data are summarised. Where available, the key messages from the researchers were quoted verbatim. Part of the analysis comprised summarising the key messages of the researchers under “Key lessons”. The findings from the analysis are summarised in Table 16.
Table 16: Summary of necessary conditions for data use

<table>
<thead>
<tr>
<th>Necessary condition</th>
<th>Key lessons</th>
<th>Study</th>
<th>Implications for guidelines within NRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum relevance</td>
<td>Teachers are more likely to use curriculum-relevant assessment data, otherwise they ignore it</td>
<td>Timperly, 2009; Griffin, 2009.</td>
<td>Reports to be based on SS methods that reflect adequate curriculum content</td>
</tr>
<tr>
<td>Format of presentation</td>
<td>Use of statistical jargon discourages teachers from interacting with reports</td>
<td>Hambleton &amp; Slater, 1997; Raivoce, s.a.; Timperly, 2009</td>
<td>Report in formats with a good balance of text, visuals &amp; limited use of statistical jargon. Where necessary, use simple statistics</td>
</tr>
<tr>
<td>Interactive dissemination strategies</td>
<td>Engaging users enhances prospects of data use</td>
<td>Kanjee &amp; Makgamatha, 2002</td>
<td>Mediate the results during dissemination</td>
</tr>
<tr>
<td>Capacity for data use</td>
<td>Teachers need capacity building on how to read, interpret and utilise assessment data</td>
<td>Raivoce, s.a.; Timperly, 2009</td>
<td>Provide training, tools &amp; support to districts and schools</td>
</tr>
<tr>
<td>Learning communities on data use</td>
<td>Data use was an area of collective interest</td>
<td>Griffin, 2009; Timperly, 2009</td>
<td>Identify and utilise district &amp; school champions on data use</td>
</tr>
<tr>
<td>Influence of school leadership</td>
<td>Where school leadership took interest in and facilitated use of data, teachers were quick to use data in their teaching</td>
<td>Timperly, 2009</td>
<td>Include data use in all leadership training for district officials &amp; school principals. Disseminate results targeted to teachers and school management, e.g. school reports</td>
</tr>
</tbody>
</table>

From Table 16, most of the conditions that were found to promote data use had both direct and indirect relevance to the use of performance standards. For instance, curriculum relevance was one of the criteria that were identified as critical for a SS method that will be relevant in South Africa.

In a successful New Zealand initiative that involved more than 300 schools in improving teaching and learning through effective data use, Timperly (2009:21) observed: “We have now identified a number of conditions required for the use of assessment data to have the impact we hoped for… Teachers need sufficient knowledge of the meaning of the assessment
data to make appropriate adjustments to practice”. Kanjee and Makgamatha (2002:viii) in South Africa made similar observations following a successful initiative to support districts which, in turn, had to support schools. In the guidelines that the NRF must provide, curriculum relevance must be key to keeping teachers and other users interested in the reports.

The observation that reports from assessment need to have a good balance of text, visuals and simple statistics has implications for selection and training of data analysts and report writers. In standards-based reports that are envisaged in the NRF, texts must necessarily expound on knowledge and skills that are indicated in the PLDs. Arguing for development and use of clear performance descriptors in assessment reports in a cross-national study that involved six countries in the Pacific Islands Region, Raivoce (s.a.:1) warned: “Teachers will not take up attractive sounding ideas, albeit based on extensive research, if these are presented as general principles which leave entirely to them the task of translating them into everyday practice”. Guidelines in the NRF will need to be specific on what teachers must do with data.

In a treatise on teacher use of assessment data, particularly commenting on the use of raw scores in reporting learner performance in the Australian context, Griffin (2009:3) averred: “It is only when the numbers have a meaningful interpretation that measurement and assessment begin to merge and they build a link to curriculum.” This argument underpins the proposal to have a standards-based NRF and guide provinces and districts on meaningful ways of reporting and utilising assessment data.

Capacity building for data use was identified as one of the key features in the theory of action propagated by Ronka et al (2010). The finding from successful initiatives on data use to improve teaching and learning is that teachers cannot be assumed to be able to interrogate assessment reports on their own. Although the DBE has started distributing guidelines on how to use ANA data (DBE, 2011b), findings from the literature indicate that “School leaders need to be able to have the conversations with teachers to unpack this meaning (in the reports)” (Timperly, 2009:21). Both school leaders and peers have a role to play in developing a culture of data use as Griffin (2009:1) argues: “that peer accountability is more important than system accountability in helping teachers to use test data to improve learning. When peer accountability is in place its effect on decision making for intervention is direct.”
4.3.2 Summary of findings from Research Question 2

The reviewed research indicates that guidelines that should be provided to national, provinces, districts and schools as part of the standards-based NRF to enhance the utilisation of information from LSAS must be characterised by the following principles whose efficacy was attested to in the various studies:

a) Ensuring relevance of the assessment and the assessment reports to the curriculum that teachers implement;

b) Reporting results in formats with a good balance of text, visuals & limited use of simple statistics, where necessary;

c) Use of interactive strategies during the dissemination of the LSAS results to teachers and school leaders;

d) Building the capacity of users, especially teachers, school leaders and district officials, on data use as well as providing appropriate tools for the purpose;

e) Establishing learning communities for data use rather than limiting capacity to individuals; and,

f) Including data use in all leadership training for district officials & school principals.

These principles, and similar principles contained in legal and policy documents that regulate education delivery in South Africa, were used as a basis to compile specific guidelines that are recommended as part of the NRF, are presented in Chapter 5 and are given practical expression in the examples of standards-based reports in Chapter 6.

4.4 Findings on Research Question 3

The third research question in this study sought to get a preliminary validation of a standards-based reporting format. The question sought to find out if officials would endorse or reject a standards-based reporting template for ANA. Responses to this question were considered to be critical to the acceptance of the NRF that is based on performance standards.

Research Question 3
What are the perceptions of national and provincial officials regarding comparisons between a standards-based report and a report based on the CAPS (DBE, 2011c) rating codes?

4.4.1 Survey findings

To answer this research question a specially-designed questionnaire was e-mailed to national and provincial officials requesting their views on the use of CAPS-based (T1) compared to the standards-based (T2) templates for reporting ANA results. The respondents were provided with the two reporting templates and asked to indicate their preference for each in terms of the attributes that were given (Appendix G). In Table 17 the attributes that the respondents were to consider in choosing between T1 and T2 are listed in the second column. The percentages of the respondents who selected each of the templates are indicated in the third and fourth columns, respectively. The mean percentage on each template is shown in the last row.

Table 17: Summary of national and provincial views on T1 and T2 (n=39)

<table>
<thead>
<tr>
<th>No.</th>
<th>Attribute of report</th>
<th>% Preferring T1</th>
<th>% Preferring T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Useful</td>
<td>49</td>
<td>51</td>
</tr>
<tr>
<td>2</td>
<td>Information-rich</td>
<td>3</td>
<td>97</td>
</tr>
<tr>
<td>3</td>
<td>Clear</td>
<td>8</td>
<td>92</td>
</tr>
<tr>
<td>4</td>
<td>Meaningful</td>
<td>8</td>
<td>92</td>
</tr>
<tr>
<td>5</td>
<td>Curriculum-specific</td>
<td>3</td>
<td>97</td>
</tr>
<tr>
<td>6</td>
<td>Knowledge-deficiency specific</td>
<td>27</td>
<td>73</td>
</tr>
<tr>
<td>7</td>
<td>Skills-deficiency specific</td>
<td>11</td>
<td>89</td>
</tr>
<tr>
<td>8</td>
<td>Teacher's choice</td>
<td>19</td>
<td>81</td>
</tr>
<tr>
<td>9</td>
<td>Subject Advisor's choice</td>
<td>8</td>
<td>92</td>
</tr>
<tr>
<td>10</td>
<td>Parent's choice</td>
<td>46</td>
<td>54</td>
</tr>
<tr>
<td>11</td>
<td>(Can) Use with confidence</td>
<td>14</td>
<td>86</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>18</td>
<td>82</td>
</tr>
</tbody>
</table>

From Table 17, the percentage of national and provincial officials who endorsed the attributes of the standards-based reporting template, T2, was generally higher than those who endorsed the same attributes in relation to the raw score-based reporting template, T1. Overall, the mean percentage of respondents who endorsed the attributes in relation to the standards-based
reporting template (T2) was 82% compared to 18% who endorsed the raw score-based reporting template (T1).

The attributes of T2 that were endorsed by the largest percentage of the respondents, 97% each, were “information-rich” and “curriculum-specific”. In all 73% and 89% of the respondents endorsed T2 because it was specific in identifying knowledge and skills deficiencies, respectively.

The other attributes of T2 that was endorsed by a significantly high percentage of the respondents, each receiving 92% of support, were “clear”, “meaningful” and “Subject Advisor’s choice”. In other words, the majority of the respondents preferred a reporting template that is information-rich, curriculum-specific, clear and meaningful. Of the 39 respondents 92% indicated that T2 was preferred for use by subject advisors, 81% indicated preferred use by teachers and 54% indicated they would prefer T2 for use by parents.

On the overall usefulness of the two reporting templates, there was an almost even split, 49% and 51%, between respondents who found T1 and T2 “useful”, respectively. This was a curiously tight margin given that on the specific aspects of the two templates the margins were observably wide and in favour of the standards-based T2. A limitation of self-completed instruments, which effectively becomes a limitation of this study, is that one can never be certain on whether the respondents interpreted the stimuli as intended. Regardless, the narrow margin was still in favour of the standards-based T2.

4.4.2 Verification of survey findings

The respondents were asked to provide other comments that they wanted to make about the two reporting templates. Enough space was provided for free writing in the questionnaire. Typical open-ended comments that were given by respondents are presented in this section. For anonymity, the respondents have been identified according to whether they were in a province (Province and a fictitious number) or they were DBE officials.
Comments on information content of reports

In their comments, the respondents indicated that a standards-based report provides substantive feedback to all target audiences and as such, the report enables users to make meaningful interventions. A typical response was:

“Template 2 is the obvious choice as it gives all the information that everybody in the system needs to know including the parents. This type of reporting will make interventions easier without having to deal with two different reports (i.e. technical as well as diagnostic report). Reporting in this way in itself is diagnostic in nature and allows everybody to intervene immediately.” Province 1, Assessment.

There was also recognition of and preference for detail and specificity in the content of reports which need the respondents perceived was best addressed by a standards-based report as can be observed in the following comments:

“The more granular the results are provided, the more helpful the information becomes.” Province 2.

“I agree that Template 2 contains more useful and relevant information which would structure processes to a greater extent.” Province 3, Curriculum.

The respondents seemed to agree that even with standard-based reporting there should be targeting specific audiences with specific content. For instance, a perception was shared that parents would need to be given feedback that would be different from other audiences such as teachers and district officials. A typical comment was:

“Reporting template two (T2) is much more meaningful but maybe not for our parents who are not acquainted with the curriculum content.” Province 2

But the fact that a report based on raw scores or percentages (T1) was not adequate was emphasised in many ways as can be observed from this comment:
Q. Moloi

“Template 1 (based on raw scores as percentages) does not in any way assist the process and cannot help anyone come up with intervention strategies to improve the system.” DBE official.

“I would have loved to also see the percentages column in Template 2 (T2) as the percentages are important. This template 2 (T2) will be a step in the correct direction, though it should not be a standardised template in terms of the remarks. It needs to be province-specific.” Province 4, Assessment.

**Comments on use of 4 or 7 PLs**

Regarding the proposal to report on four performance levels as against the seven that are currently prescribed in the curriculum documents, the respondents seemed not amenable to changing the status quo even though they acknowledged the meaningfulness of the standards-based reporting format (T2). The conflict between the need to comply with the policy that is in force and the realisation of the benefits that could be derived from an innovative way of reporting was evident:

“I think you have to maintain the 7 level descriptors (Template 1) (T1) but the info on template 2 (T2) is more meaningful.” Province 3, Curriculum.

“Template 2 (T2): include 7 point scale rating for it to be aligned to the National policy and promotion requirements for Grades R to 12.” DBE official.

“Template 2 (T2) gives more information compared to template 1 (T1), performance levels should remain the same as in template 1 (T1), which is in line with CAPS. The score range defining performance levels should be included as well in template 2 (T2).” Province 5, Assessment.

The use of four performance levels was actually perceived to be a “shortfall” or weakness by some of the respondents:

“Template 2 (T2) has a shortfall in that the level it has only 4 levels. These broad levels tend to bunch performance of learners together instead of clearly distinguishing between performance as it happens when the levels are many like in template 1.” DBE official.
“Template 2 (T2) is more appropriate to report to all the stakeholders but the level definitions /descriptors are not in line with CAPS policy. It would be more relevant if all level descriptors (1-7) are used.” Province 3, Assessment.

However, there were also perceptions of a need for change as can be inferred from this comment:

“I like template 2 (T2) more for the following reasons:

1. It consists of 4 levels of achievement.

2. It describes the achievement level in terms of the content that must be achieved in a particular level.” Province 2, Curriculum.

Comments on the usefulness of the reporting templates

The respondents were asked to indicate the extent to which they perceived the relative usefulness of each of the reporting templates. Their perceptions were influenced largely by their official responsibilities but also by their knowledge of the South African basic education system. Some of their responses are worth noting:

“The 2 templates are suitable, each serving a different purpose; I would not replace one with the other. Template 1 (T1) gives a quick and easier glance at the report, which should ideally be followed by Template 2 (T2), which does explain the performance reflected in the Template 1 (T1).” Province 6, Assessment.

“For the end user, teachers and advisers will particularly enjoy the qualitative data from template 2 (T2) whereas parents generally want a straightforward report, possibly with practical recommendations for improvement.” Province 2, Assessment.

Overall, there was evident support for the standards-based reporting template. There were, however, indications from the free responses that the respondents, as practitioners who are responsible for implementing the existing raw-score assessment framework, could have been influenced largely by the status quo rather than independent convictions. Comments such as “The 2 templates are suitable, each serving a different purpose; I would not replace one with the other.”, “I think you have to maintain the 7 level descriptors (Template 1) (T1) but the
information on template 2 (T2) is more meaningful” and “I would have loved to also see the percentages column in Template 2 (T2) as the percentages are important. This template 2 (T2) will be a step in the correct direction ...” are indications of a possible dilemma the respondents had to resolve between the perceived need to comply with the status quo and what their professional judgement suggested could be a preferred option.

4.4.3 Summary of findings on Research Question 3

What are the perceptions of national and provincial officials regarding comparisons between a standards-based report and a report based on the CAPS rating codes?

Because of the critical role that they play in implementing and monitoring the delivery of education, national and provincial officials in curriculum, teacher development and assessment were surveyed to understand their perceptions about the relative importance of standards-based reporting templates and reporting templates that are based on raw scores called “CAPS codes”. From their ratings of the two reporting templates, which were corroborated in the comments that the officials wrote, there was overwhelming preference for a reporting template that provides feedback in rich curriculum-specific information that is clear and meaningful. More than 80% of the respondents preferred a standards-based reporting template because their perception was that this reporting format was more likely than the other one to provide the kind of feedback that they perceived was desirable.

Whilst indicating unequivocal support for a standards-based reporting template, the respondents also seemed not to be predisposed to possible changes to some of the current features of reporting as prescribed in the curriculum policy documents. They suggested that a meaningful reporting format was desirable but it should still retain the seven levels of reporting that are prescribed. The tendency to resist a paradigm change is one obstacle that will need to be addressed in the introduction of a standards-based NRF. A strong motivation on why fewer PLs will be critical to explaining the value and use of the NRF.

On the whole the respondents were supportive of a standards-based reporting framework which, according to them, would make reporting meaningful and result in well-targeted interventions for improvement. This kind of attitude creates an ideal atmosphere for
introducing an NRF that promotes reporting large-scale assessment results in performance standards and thus enhancing the chances of effective learning and teaching in schools.

4.5 Summary

This chapter presented findings to the three research questions of this study. The first research question used a comparative approach to find the most appropriate SS method that could be adopted into and as part of the proposed NRF and will serve as the basis for analysis and reporting of performance standards. The second research question reviewed appropriate published literature to explore recommended guidelines for effective use of assessment data with a view to enhance teaching and learning towards improving learner performance. The exploration was intended to identify guidelines that could be incorporated into the NRF to promote LSAS data utilisation in provinces, districts and schools. The third research question solicited views of education officials on whether they would support a raw score-based or a standards-based reporting framework for assessments. The main findings of the study are summarised in this section.

Appropriate SS method for South Africa

Three SS methods were compared, viz. the Angoff, OSS method and the OBM. The two latest methods, the OSS and OBM were first checked for validity and were found to yield similar cutscores in different contexts represented by real and simulated data sets. The validity and generalisability of the two methods was, therefore, confirmed. A comparison of cutscores generated from each of the three methods showed consistent patterns within and across subjects, both in terms of the cutscores as well as in the possible impact that the cutscores would have, if adopted. However, a statistical test of significance (Mann-Whitney U test) indicated that the Angoff and OSS methods did not only generate cutscores that are different at all performance levels, but the impact of the cutscores was also different, except at the Not Achieved Level in mathematics, where there was no difference in the impact made by cutscores from either method. This exceptional observation will need further investigation.

The three SS methods also differed in other ways. The Angoff and OSS methods, unlike the OBM, provide cutscores that embed important curriculum content so that groups of learners on opposite sides of a cut score can also be characterised by specific knowledge and skills
that distinguish them. However, unlike the Angoff, the OSS method generated cutscores that were robust, i.e. did not change significantly with iterations, and thus provides a stable amount of curriculum content within the categories marked by cutscores.

In terms of efficacy, i.e. turnaround times, the Angoff method was found to be the most labour- and resource-intensive method with the longest turnaround time, viz. two days (16 hours). The method was also found to be conceptually demanding as it required focused attention on the operationalization of a “minimally competent” hypothetical candidate. The view of the panellists was that this focused conceptual demand rendered the method prone to errors arising from concentration lapses among panellists. The OSS method was more efficacious in that it required far less time, viz. a half day (4 hours) and only required teachers to reflect on the content and skills they teach learners on a day to day basis. However, the method requires technical/analytical skills and software that may not be in common supply.

Whilst the OBM was the most precise in terms of eliminating ambiguities around cutscores, the method is completely devoid of any curriculum content in its conceptualisation. For educational purposes, it could be argued that the OBM would best be used in complementing another SS method that reflects curriculum content.

Although the findings on which SS method was the most appropriate were inconclusive because each method had its unique strengths and weaknesses for the context of South Africa, the OSS method displayed superior attributes related to curriculum content validity, robustness and generalisability of cutscores and efficacy of the SS process. An interesting finding was that cutscores from different SS methods impacted differently on different subjects in terms of categorising learners into various PLs. This finding may need to be researched further because it could imply that the NRF should specify the SS method to be used in setting performance standards for each subject.

Guidelines for effective implementation of the NRF

In this study perspectives and information from appropriate and successful empirical studies on data use to improve teaching and learning were interrogated. They seemed to converge on common findings that have direct implications on how provinces and districts should be guided on how to report and utilise assessment data. Such guidelines need to ensure that
assessment data is relevant to the curriculum and to practices with which teachers are already familiar. There is concurrence from across regions and continents that effective data use requires appropriate capacity building for both School Management Teams (SMTs) and teachers if it is to lead to improvements in teaching and learning. All reports highlighted the need for a collective approach and culture to data effective data use for improving both practice and outcomes.

**Views of education officials on standards-based reporting**

National and provincial officials who are responsible for managing delivery of curriculum, conducting assessments and supporting teacher development were requested to indicate their preference between a raw score-based and a standards-based framework for reporting assessment data. Overall, the education officials expressed views and comments that endorsed a standards-based reporting framework for assessments. While reasons for their choices varied, the main reason cited was the level of reported detail and usefulness for feedback that the standards-based framework provides. One recommendation was that the two formats be used together rather than selecting one. This view could also be sensed in suggestions that the percentage of learners who perform at each PL must be reflected in the report template. It would appear that the officials are not averse to simple statistics that accompany the detailed textual description.

A curious observation was on recommendations to maintain the seven levels of reporting that are currently recommended in policy (DBE, 2011c). It was clear that, while officials preferred an information-rich reporting template, they were also concerned about possible policy implications regarding any changes that would be required if a four-level reporting template was adopted. In the wake of many changes that have happened in curriculum in the last twenty years, it comes as no surprise that education officials seem to fear change, regardless of the benefits that change would bring.
5. A NATIONAL REPORTING FRAMEWORK FOR SOUTH AFRICA

The National Reporting Framework (NRF) proposed in this chapter is designed to address specific gaps that have been identified in the policy and legislative framework that regulates assessment in general and LSAS in particular, focusing on the reporting and use of assessment results in South Africa. Specific instances of such gaps are described under the rationale for the NRF. In this chapter the proposed NRF is presented beginning with the rationale and definition of the NRF, followed by the underlying philosophy, the target users of the NRF and examples of key components of the framework.

5.1 Rationale

Although the basic education sector in South Africa has made significant strides in developing and implementing policies that seek to accelerate transformation and enhance quality of educational delivery and performance, overall performance of the sector is continually blighted by perennial poor performance of learners as witnessed in low scores in the ANA (DoE, 2005), TIMSS (Reddy, 2006), PIRLS (Howie, et al, 2007) and SACMEQ (Moloi & Chetty, 2010) studies. With specific reference to LSAS, some of the explanations for the non-responsiveness of the system to various improvement interventions include limited use of information in identifying and implementing interventions, due primarily to reporting, dissemination and inadequate clarity in the feedback that goes to districts and schools after each cycle of assessment.

The rationale behind the NRF is that effective utilisation of assessment data can only be realised when assessment is part of a seamless policy network, the assessment results meet the requirements and needs of the end-users, and key role-players, teachers and school leaders in particular, are equipped with necessary skills and appropriate tools to enhance their use of assessment information. Some of the gaps that the NRF intends to fill in the policy network related to reporting and using assessment data are highlighted as part of the rationale.

Section 8 (1) of the National Education Policy Act, 1996 (Act. 27 of 1996), mandates the Minister to direct that standards of education provision, delivery and performance be monitored (DoE, 1996). Subsequently, the Assessment Policy which was promulgated in...
December 1998 (DoE, 1998) provides for conducting systemic evaluation at the key transitional stages, viz. grades 3, 6 and 9, as one of the monitoring mechanisms at the disposal of the Minister to assess the effectiveness of the education system, and the extent to which the vision and goals of the system are being achieved. Later, the Sector Plan, *Action Plan to 2014: Towards the realisation of schooling 2025*, (DBE, 2011a) introduced ANA as a large-scale adaptation of the sample-based systemic evaluation for monitoring educational quality. Thus legislation is clear that educational performance must be monitored and also that one vehicle for executing this responsibility is the application of LSAS for assessing, reporting on (and using) learner performance in specific subjects and grades.

While the legislative framework for monitoring performance in education is quite comprehensive, no matching guidance is provided on how the results of measured learner performance should be reported, disseminated and utilised for purposes of improving the quality of educational performance. Schedule 4 of the Constitution of the Republic of South Africa (1996:109) categorises pre-tertiary education as a concurrent responsibility of both the national and provincial governments, while the NEPA Act (1996) specifically authorises the Minister to monitor and, therefore, provide feedback on the outcomes of large-scale assessments. Structurally, the DBE should provide support to provinces, provinces to support districts, districts to support schools and schools must account to parents. For this system to be effective, the national department must provide a national framework that guides the how information is reported and used at all these levels. However, no such framework is currently in place; hence the proposed NRF is expected to fill this gap.

In keeping with the provisions of NEPA (1996), the Revised National Curriculum Statement (NCS) Grades R-12 (DoE, 2002:18), serves as a broad ‘standard’ for delivery and monitoring of educational quality. The Curriculum and Assessment Policy Statement (CAPS), which replaced the NCS, provides protocols for assessment of learner performance in all approved subjects and also specifies how the results of assessments should be reported (DBE, 2011c). Arguing for results-oriented mutual accountability among stakeholders in education, the National Development Plan (NDP) directs that ANA results shall serve as a reliable measure of the health of the education system and further stipulates that the results should be reported “to parents and the community in a way that makes the data easy to interpret” (NDP, 2012:311).
According to the Action Plan (2011a) the NCS generally, and CAPS specifically, is expected to define and describe what the “required” minimum levels of literacy and numeracy should be at each of the grades 3, 6 and 9 that are specified in the national goals. However, the CAPS (DBE, 2011c) merely quantifies the expected levels of competency that should be reported from assessments, including large-scale assessments. It is important to consider and address some of the limitations set by this arrangement.

Firstly, the levels prescribed in the CAPS rating codes (DBE, 2012a) are not associated with any descriptions regarding what learners should know and be able to do. In the absence of such descriptions, no information is available that teachers, parents and/or learners can use to identify and thus address weaknesses and improve on strengths. For example, a score of 56% in a test provides no information on what should be done for improving learning and teaching.

Secondly, the CAPS categories are between fixed cutscores, which are raw scores expressed in percentages, and these fixed percentage bands do not accommodate for variations in the difficulty of tests. For instance, learners who score in the range of 0%-29% correct responses are categorised as functioning at the “Not Achieved” level and those who score in the range of 80%-100% as functioning at the “Meritorious” level, regardless of the difficulty of the specific test. Yet it is known that on an easy test percentage correct raw scores tend to be higher than in a difficult test while in the same test, learners of higher ability score higher than their counterparts of lower ability (Bond & Fox, 2007). So, a meritorious achievement in an easy test may not necessarily be so in a difficult test. In standards-based categories test difficulty is taken into consideration during the rating of test items and the setting of cutscores.

Thirdly, besides the information limitation of percentage scores, a higher workload burden 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 is unrealistic to expect a teacher to keep track of and provide differentiated support across seven categories of learners in a class.

Fourthly, reporting performance against seven levels could compromise the accuracy of information and meaningfulness of associated reports. An increase in the number of reporting
levels will invariably always be accompanied by a decrease in the accuracy with which skills required for adjacent levels can be distinguished. For instance, distinguishing between skills and knowledge that characterise the “Adequate Achievement” and the “Moderate Achievement” levels, Levels 2 and 3, respectively, could be less convincing than distinguishing between “Not Achieved” and “Adequate Achievement” levels. Similarly, a distinction between Levels 4 and 5, “Adequate Achievement” and “Substantial Achievement”, respectively, could be equally less clear.

It is due to these limitations that the reporting learner performance in the NRF is based on the standard setting approach whereby each performance level is characterised by minimum skills and knowledge and not necessarily the number of items that the learner answers correctly. For relative ease and accuracy of reporting, as well as to facilitate the use of information for effecting relevant interventions, four reporting levels are proposed in the NRF: Not Achieved, Partially Achieved, Achieved, and Advanced achievement.

In practice, this means that three pairs of adjacent levels prescribed in the National Protocol for Assessment (DBE, 2012a) have been collapsed into one level each thus reducing the number of reporting levels to four (4). Specifically, Levels 2 and 3, “Elementary Achievement” and “Moderate Achievement”, respectively, have been collapsed into one level defined as “Partly Achieved”. Similarly, Levels 4 and 5 and Levels 6 and 7 have been collapsed into “Achieved” and “Advanced”, respectively. A summary of the key distinguishing features of the CAPS design (DBE, and the proposed standard-based reporting framework is shown in Table 18.
Table 18: Summary of key features of CAPS and standard setting designs

<table>
<thead>
<tr>
<th>Rating Code</th>
<th>Percentage</th>
<th>Description of Competence</th>
<th>Performance Level</th>
<th>Percentage Cutscores</th>
<th>Performance Level Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0-29</td>
<td>Not Achieved</td>
<td>Not Achieved</td>
<td>Determined in a SS process for the test</td>
<td>Concise description of skills &amp; knowledge demonstrated by a learner at this level</td>
</tr>
<tr>
<td>2</td>
<td>30-39</td>
<td>Elementary Achievement</td>
<td>Partly Achieved</td>
<td>Determined in a SS process for the test</td>
<td>Concise description of skills &amp; knowledge demonstrated by a learner at this level</td>
</tr>
<tr>
<td>3</td>
<td>40-49</td>
<td>Moderate Achievement</td>
<td>Achieved</td>
<td>Determined in a SS process for the test</td>
<td>Concise description of skills &amp; knowledge demonstrated by a learner at this level</td>
</tr>
<tr>
<td>4</td>
<td>50-59</td>
<td>Adequate Achievement</td>
<td>Achieved</td>
<td>Determined in a SS process for the test</td>
<td>Concise description of skills &amp; knowledge demonstrated by a learner at this level</td>
</tr>
<tr>
<td>5</td>
<td>60-69</td>
<td>Substantial Achievement</td>
<td>Advanced</td>
<td>Determined in a SS process for the test</td>
<td>Concise description of skills &amp; knowledge demonstrated by a learner at this level</td>
</tr>
<tr>
<td>6</td>
<td>70-79</td>
<td>Meritorious Achievement</td>
<td>Advanced</td>
<td>Determined in a SS process for the test</td>
<td>Concise description of skills &amp; knowledge demonstrated by a learner at this level</td>
</tr>
<tr>
<td>7</td>
<td>80-100</td>
<td>Outstanding Achievement</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As can be observed from Table 18, the key distinguishing features between the CAPS and the standards-based reporting designs include:

a) Use of functionally well-defined performance levels (PLs) instead of numerical “rating codes” to categorise learners in terms of their performance. In the standards-based design learner performance at each level is described and not just rated or coded as is the case in the CAPS design.

b) Determining cutscores through a systematic standard-setting process on each test instead of using fixed cutscores which are not responsive to variations in test difficulty.

c) Use of detailed performance level descriptors (PLDs) to describe the repertoire of knowledge and skills that learners at a particular PL demonstrate instead of using a generic summary “description of competence”. The dearth of skills and knowledge description in the CAPS assessment framework limits the utility value of the assessment results and compromises the contribution that the effective use of assessment results can make towards improving the quality of teaching and learning.
The proposal of a standards-based rather than score-based NRF is intended to address this limitation.

Govender (2014) conducted a comprehensive study on how district officials perceive and utilise ANA results to support schools. A significant finding of the study was that districts relied heavily on ANA guidelines and exemplars as “standardised” instruments for quality assessment and depended almost exclusively on ANA results as an indicator of the quality of teaching and learning in primary schools. These findings indicate the potential power that ANA results have as leverage for improving the quality of education. By enhancing the clarity, precision and information-richness of ANA results through the use of a standards-based NRF, the impact of the use of the assessment results could be both deepened and accelerated.

Moreover, as reported in this study, an overwhelming majority of provincial and district officials responsible for either curriculum delivery or national assessments indicated a very strong preference for a standard-based rather than a score-based framework for reporting assessment results.

5.2 Definition of the NRF

The NRF is defined as a guideline document on the format and content to be included in a report from large-scale assessment of learners to facilitate effective recording, reporting and use of data to improve teaching and learning.

5.3 Underlying philosophy for reporting LSAS results

The philosophy underlying the reporting of assessment data in general, and LSAS results in particular, holds that, providing feedback from assessment in formats that are user-friendly, meaningful and information-rich will motivate target users to use the assessment results. If the feedback is coupled with relevant capacity building in a sustained culture of data use, the provision of meaningful data will lead to effective data use, enhance evidence-based decision-making regarding the implementation of relevant interventions for improving learning and result in improved performance in schools.
The proposal for an NRF is premised on this philosophy and further maintains that a standards-based framework will not only enhance the meaningfulness and use of feedback from LSAS but will actually help address the perennial problem of underperformance in South Africa. Building on the philosophy, the study set out to develop a standards-based NRF that will incorporate guidelines on how assessment data should be analysed, recorded, reported and utilised effectively to impact on achievement. The key tenets underpinning the NRF are:

a) a specific focus on the criterion-referenced paradigm in assessment

b) using performance standards to report results in meaningful ways to provide relevant evidence that will enhance capacity to improve practice and

c) building capacity for data use at all levels of the system and particularly in the teaching and learning space.

5.4 Purpose of the NRF

The purpose of the NRF is to enhance the utilisation of assessment data to inform decision-making at the different levels of the education system to improve teaching and learning in schools. The main thrust of the NRF is to encourage a shift towards reporting learner performance in terms of the knowledge and skills that characterise learners who function at various levels of the competency continuum instead of using numerical data that has been shown to have serious limitations in communicating assessment results in meaningful ways.

5.5 Content of the NRF

The NRF provides guidelines on the formats in which data from large-scale assessments, including ANA, must be recorded and presented as well as the priority content that must be included in reporting the results in ways that will help schools and teachers identify what learners know or do not know and what learner can or cannot do, and thus provide substantive evidence for classroom teachers and school leaders on how to address specific learning needs of learners. It also provides districts with substantive evidence in determining the nature of the support that must to be given to schools and teachers for improving learning and teaching. At the national and provincial levels, the NRF provides information to enhance
both the PED and the DBE decision-making processes regarding provision of appropriate resources, capacity development initiatives and implementation of policy frameworks that are responsive to identified needs of schools, teachers and learners.

The NRF is intended to be utilised by the DBE, PEDs, districts and schools in the process of recording, reporting and utilising assessment data (information) from LSAS to improve learning and teaching. Parents and learners will be direct beneficiaries of the NRF in terms of the quality specificity and clarity of information on what has been achieved and what still needs to be achieved. The framework must also serve as a tool that will help the education system in addressing issues of equity. It must highlight and foreground the plight of poor and marginalised learners by making it possible to focus interventions on these learners.

a) National level uses

The DBE will use the NRF to analyse and report data from LSAS to monitor changes over time in the proportions of learners who achieve acceptable levels of performance across provinces and districts. The NRF will provide specific information on repertoires of knowledge and skills that learners demonstrate or fail to demonstrate within and across provinces and districts, across schools in different poverty quintile levels, across grades and across genders. Validated with other sources of information, e.g. teacher profiles, information in the NRF will indicate whether there are equity issues to be addressed in terms, for instance, of distribution of teachers with specific subject skills, necessary support and possible bottlenecks in policy implementation.

The key users of the NRF at the DBE level will be the National Assessment, Teacher Development and Curriculum Development Units. The Assessment Unit needs to ensure that tests that shall be used for the LSAS have sufficient number of items at each performance level to make sure that reported scores are stable and reliable. Used with proper test equating designs, the NRF will indicate whether trends in performance in specific subject domains show increases, declines or remain unchanged over time.

For the Teacher Development and Curriculum Development Units, the NRF shall provide specific information on the skills and knowledge deficiencies of learners. This information can be used to identify areas of the curriculum in which teachers and learner’s require
additional support, and should provide guidelines for developing relevant interventions for teachers and learners, for example Continuous Professional Development courses or revisions of teachers guides and/or learners workbooks

b) PED level uses

Provinces will use the NRF to create reports for monitoring the level and quality of learner performance within and across districts and across all the other categories that have been identified for DBE users, focusing especially on poor and marginalised learners. The main users of the NRF at the PED level shall be Curriculum implementation, Assessment and monitoring and support Units. From the distribution and trends in what learners can or cannot do in the curriculum, provinces shall be able to identify which districts, schools, school categories, (e.g. quintiles) require specific support to improve learning and teaching and the specific nature and content of the interventions relevant to the province. Moreover, this information should also serve for guiding how interventions are implemented, for example, determining appointment and deployment of subject advisors with relevant competencies.

c) District level uses

As the closest delivery points to schools, districts can use the NRF to generate reports that identify schools, school management teams and teachers who require either general or specific support to be able to improve learner performance. From the performance levels and standards that learners attain, districts will be able to deduce corresponding deficiencies in teacher content knowledge and pedagogic skills and decide on whether interventions are required at group or individual teacher levels. Moreover, districts will be able to track changes in the performance of poor and marginalised learners at the lowest or Not Achieved levels over time.

Standards-based reports on learner achievement will have far reaching implications on whole-school development and improvement levels. For example, from the levels and quality of knowledge and skills that learners are either able or not able to demonstrate in the assessment, districts will be able to identify instructional management skills which distinguish among schools in similar socio-economic conditions. They will then be able to initiate appropriate interventions for capacity building on school leadership with a focus on
instructional leadership whereby school principals promote effective teaching and learning (Hallinger and Heck, 1997) and ensure that the needs of poor and marginalised learners are appropriately and adequately met.

d) School level uses

At school level the NRF is designed to support teachers, School Management Teams (SMTs) and parents to generate and report relevant data for identifying learning successes and needs of individual learners. The inherent use of performance descriptors in standard setting and the resulting hierarchies of skills and knowledge will enable teachers to identify learners at risk in terms of cumulative skills repertoires and to develop appropriate teaching strategies to scaffold learning and enhance chances of successful learning by individual learners.

Employment of targeted interventions linked to identified knowledge and skill deficiencies, as opposed to one-size-fits-all approaches, can reduce repetition rates because each learner will be taught at the level of their learning need and readiness to learn.

One sustainable use of the NRF at school level will be building the capacity of teachers to teach, assess and use assessment data to improve teaching and learning. Within the framework of standard setting, teachers can develop clearer understanding of typical skills that characterise effective learning at various levels of performance. Knowing which simpler skills are antecedents of more complex skills will assist teachers learn how to sequence learning programmes and enhance chances of success of learners because every child shall be assisted at a relevant level of learning need and readiness to learn.

By avoiding use of intimidating statistical jargon in standards-based reports and focusing on clear and easy to understand descriptions of skills and knowledge, parental understanding of the outcomes of assessment in general, and LSAS results in particular, can also be deepened. Consequently, parents will have access to meaningful information regarding performance levels of their children and can interact on a more informed basis with their children around assessment reports. Many studies have shown that parental involvement increases the chances of effective learning among affected children (Flouri & Buchanan, 2004).
5.6 Guidelines for effective data use

The key principles that underpin the need for, and the use of, the NRF are closely linked to identified legislative and policy stipulations (NEPA (1996), NCS (DoE, 2002:18), CAPS (DBE, 2011c), National Protocol for Assessment (DBE, 2012a) as well as the NDP (2012:311) regarding the conduct, reporting and use of assessment in the South African basic education sector. Each of these legislative and policy instruments provides a basis for guiding the purpose for which performance data should be collected, how the data should be analysed and the results communicated as well as actions that need to be taken as a consequence of emerging evidence. To enhance the use of data from LSAS for improving the quality of teaching and learning, it is important that the NRF be framed around the principles that are provided in the existing legal and policy framework for assessment and its use.

NEPA (1996:8) stipulates that the monitoring and evaluation of the standards of education performance in South Africa shall be undertaken through analysis of data gathered by specified means, including LSAS, reporting the results and requiring relevant authorities to remedy situations of underperformance within specified timeframes. The NCS (DoE, 2002:18) requires that the minimum knowledge and skills that are expected of learners, as well as the standards that must be achieved, be made clear and specific. A simplified version of the NCS, CAPS (2011c), specifies the knowledge and skills that each learner must acquire in each grade and subject. In the National Protocol for Assessment Grades R – 12 (DBE, 2012), schools are guided to report learner achievement by indicating combinations of quantitative national codes and percentages as well as qualitative comments, presumably on what learners know and can do (DBE, 2012).

There were also principles that research indicated underpinned conditions that are conducive to effective data use for improving the quality of teaching and learning in schools. These should be viewed as complementary rather than additional to the documented legal and policy principles and include:

a) ensuring relevance of the assessment and the assessment reports to the curriculum that teachers implement;
b) reporting results in formats with a good balance of text, visuals & limited use of simple statistics, where necessary;

c) using interactive strategies during the dissemination of the LSAS results to teachers and school leaders;

d) building the capacity of users, especially teachers, school leaders and district officials, on data use as well as providing appropriate tools for the purpose;

e) establishing learning communities for data use rather than limiting capacity to individuals; and,

f) including data use in all leadership training for district officials & school principals.

Based on the cited principles, the NRF highlights key features of results of LSAS and provides specific guidelines on what must be done to ensure that data is reported, disseminated and utilised effectively to improve teaching and learning. The specific guidelines include:

- **Relevant and curriculum-specific feedback**: providing carefully nuanced information on “required” levels of competency and what learners at each level can and cannot do in literacy and numeracy at the key grades as alluded to in the Action Plan (DBE, 2011a) and the Presidential injunction (2010) both of which make reference to acceptable levels of competency.

- **Standardisation**: - setting, clarifying, maintaining and communicating standards to which schools and learners should aspire to, as specified in the CAPS (DBE, 2011c).

- **Efficiency**: - enhancing the utility value of assessments, particularly large-scale assessments, and thus sharpening and deepening the impact of assessment on the quality of teaching and learning (DBE, 2011a).

- **High quality**: ensuring that data from assessments is accurate, error-free and of a high quality for users to interpret it with ease as indicated in the NDP (2012:311).
Q. Moloi

- **Capacity building**: providing on-going training on the development of assessment instruments, on how to report assessment results and on different ways in which assessment information can be used to inform effective teaching and learning (DBE, 2011a).

- **User-friendliness**: presenting assessment results in formats that are non-intimidating, attractive, easy to read and meaningful as alluded to in the NDP (2012:311).
5.7 **Technical features of the NRF**

The NRF consists in two key components, viz. the processes that must be followed and the outputs which are mainly reports that must be compiled, distributed and utilised at various levels.

### 5.7.1 Technical processes

The NRF will require that, after the assessment has been administered and tests have been marked, the following process activities must be executed, not necessarily in the listed sequence:

a) Coding and scoring of learner responses

b) Data capturing and cleaning

c) Appointment of standard setting panellists, a minimum of 15 professionals for each test

d) Training of panellists on standard setting

e) Setting of performance standards for each subject

f) Data analysis and categorisation of learners according to performance levels

g) Compilation and approval of reports by the DBE

h) Dissemination of reports to various audiences

i) Sustained support to build and bolster the confidence of teachers in the use of data to improve learning.

### 5.7.2 Key NRF outputs

Various standards-based reports, targeting various audiences, shall be the key outputs produced in terms of the NRF. A standards-based report for a grade shall be:
a) Short and precise, not exceeding four (4) A4 pages in length – this could vary depending on the level, if possible, but conveying critical information to enable users to respond to specific issues. Target users of the report have many other documents that they must read and may be turned off by a voluminous report.

b) Produced in both hard and soft form to meet the needs of users in different contexts.

c) Printed using different colour codes for parents, schools, districts, provinces and the national level.

5.8 Reporting and contents of reports

The NRF provides guidelines on how reporting of assessment data should be conducted at the different levels of the system and specifies, as a minimum requirement, types and formats of reports that must be produced, disseminated and utilised to inform decisions that will lead to continuous improvement of the quality of teaching and learning in schools. Reporting of LSAS results must be planned and prioritised to reach provinces, districts and schools before the end of the school year, or at least before the on the first day of a school year or earlier. Reporting of the assessment results must be based on performance standards that comprise well-defined performance levels (PLs) and are qualified by concise and accurate performance level descriptors (PLDs) which describe the knowledge and skills that a learner at a given Performance Level is expected to demonstrate. Reporting must recognise and articulate “content standards” that are specified in the curriculum and “performance standards” which indicate “how much” knowledge and skills a learner has acquired or developed.

5.8.1 Performance levels

For relative ease and accuracy of reporting, as well as to facilitate the use of information for effecting relevant interventions, four reporting levels are proposed in the NRF. Each level has been provided with a relevant label that describes what performance at the specific level means. Thus Level 1 has been labelled – “Not Achieved”, Level 2 – “Partially Achieved”, Level 3 – “Achieved”, and Level 4 “Advance”. Moreover, for each Performance Level, a policy definition is provided that lists the practical significance of each level. In addition, the NRF also provided information on progression and intervention implications for each level, noted in Table 19.
Table 19: Policy definitions and their practical significance per Performance Level

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

5.8.2 Performance level descriptors

To add meaning to performance levels across different subject and grade areas, detailed descriptions of what learners at specific levels can or cannot do are provided. Two important features of the PLs and PLDs must be borne in mind when interpreting standard-based reports. Firstly, the NRF presents the knowledge and skills displayed by learners in a hierarchical structure.

The hierarchy of PLs means that the knowledge and skills at a given PL are cognitively more demanding than those at a lower PL and less demanding than those at a higher PL, i.e. ‘Not Achieved’ < ‘Partly Achieved’ < ‘Achieved’ < ‘Advanced’ in terms of cognitive demand. Secondly, a learner who displays knowledge and skills that are characteristic of a given PL is expected to also have a high probability of displaying knowledge and skills at lower PLs, but is unlikely to display knowledge and skills at higher PLs. For instance, a learner who functions at the Achieved level is expected to also demonstrate the knowledge and skills at Partly Achieved and Not Achieved levels but is unlikely to display Advanced level knowledge and skills. Illustrations of PLs, PL definition and PLDs, based on the grade 6
CAPS for mathematics and English First Additional Language (FAL), are reflected in Table 20 and Table 21, respectively.

**Table 20: Grade 6 mathematics PLs and PLDs**

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>A learner at this level will, with intensive support, recognise basic number systems. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs:</td>
</tr>
<tr>
<td>• count forward only with whole numbers</td>
<td>• count forward and backwards in decimals</td>
<td>• count, recognise and do calculations using fractions &amp; decimals</td>
<td>• critically read, interpret &amp; analyse with awareness of sources of error &amp; manipulation to draw conclusions &amp; make predictions</td>
</tr>
<tr>
<td>• count objects not exceeding 10</td>
<td>• recognise place value up to 9 digits</td>
<td>• represent multiples, factors &amp; prime numbers</td>
<td>• list possible outcomes for simple experiments including tossing a coin, rolling a die &amp; spinning a spinner</td>
</tr>
<tr>
<td>• add whole numbers up to 10</td>
<td>• round off number up to 1000</td>
<td>• find percentages of whole numbers</td>
<td>• distinguish between volume, surface area &amp; dimensions of rectangular prisms</td>
</tr>
<tr>
<td>• draw simple pictures of objects</td>
<td>• add and subtract up to 9 digits</td>
<td>• solve problems involving finances and measurement</td>
<td>• solve problems involving different time zones</td>
</tr>
<tr>
<td>• Measure length of lines</td>
<td>• do simple calculations using ordinary fractions and decimals</td>
<td>• compare rate and ratio</td>
<td>• estimate, record, compare &amp; convert between SI units (including mass, temperature, distance and capacity)</td>
</tr>
<tr>
<td>• Name few SI units</td>
<td>• read digital and analogue time</td>
<td>• identify &amp; describe numeric and geometric patterns</td>
<td>• organise &amp; record data</td>
</tr>
<tr>
<td>• measure area &amp; perimeter of objects</td>
<td>• measure using basic SI units</td>
<td>• use and describe transformations</td>
<td>• calculate the median &amp; mode of data</td>
</tr>
<tr>
<td>• draw simple bar graphs</td>
<td>• draw pictographs</td>
<td>• locate and describe movement on a grid</td>
<td>• list possible outcomes &amp; predict “likelihood” of events.</td>
</tr>
</tbody>
</table>
Table 21: Grade 6 FAL PLs and PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, read single words, short sentences &amp; pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read texts with simple sentences. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; interpret meaning of stories relating to real life experiences. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; understand complex texts in formal &amp; informal stories. She/he can:</td>
</tr>
<tr>
<td>- recognise words that were learnt previously</td>
<td>- skim &amp; scan some poetical elements</td>
<td>- identify characters &amp; plot setting</td>
<td>- read poetry effectively</td>
</tr>
<tr>
<td>- read forward &amp; answer short questions based on the text</td>
<td>- skim, scan &amp; summarise texts</td>
<td>- identify ethical issues such as cultural/social diversity</td>
<td>- evaluate texts</td>
</tr>
<tr>
<td>- identify only the key character in a text</td>
<td>- use vocabulary of 1500 words.</td>
<td>- highlight the moral lesson behind a story</td>
<td>- identify formal and informal texts</td>
</tr>
<tr>
<td></td>
<td>- write friendly letters</td>
<td>- infer information from a complex text</td>
<td>- easily utilise vocabulary of at least 5000 words</td>
</tr>
<tr>
<td></td>
<td>- extract information directly from a short text</td>
<td>- identify key elements of poetry</td>
<td>- critique texts through book reviews &amp; reports.</td>
</tr>
<tr>
<td></td>
<td>- read short stanzas &amp; answer literal questions</td>
<td>- use dictionaries in building vocabulary of at least 3000 words</td>
<td>- analyse both formal &amp; informal documents</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- synthesise information from different parts of a text</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- demonstrate comprehension of inferred information from a text</td>
</tr>
<tr>
<td><strong>WRITING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, show understanding through drawings and pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL, write for social purposes. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write information in graphic &amp; descriptive forms. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write &amp; develop information from graphical &amp; descriptive forms to text. He/she can:</td>
</tr>
<tr>
<td>- write single words &amp; short sentences</td>
<td>- write personal letters, diaries, news reports</td>
<td>- write for social purposes, e.g. personal letters, diaries, dialogues &amp; simple news reports</td>
<td>- write extensively for social purposes</td>
</tr>
<tr>
<td>- represent ideas with drawings</td>
<td>- create a book cover</td>
<td>- identify similar &amp; different texts</td>
<td>- develop news reports</td>
</tr>
<tr>
<td>- spell few commonly used words correctly</td>
<td>- identify cause &amp; effect relationships</td>
<td>- express cause &amp; effect relationships</td>
<td>- design questionnaires &amp; adverts.</td>
</tr>
<tr>
<td>- write brief diaries</td>
<td>- write simple sentences &amp; draw pictures.</td>
<td></td>
<td>- integrate ideas by classifying information</td>
</tr>
</tbody>
</table>

Table 20 and Table 21 form the quintessential feature of reporting according to performance standards which, in turn, form the pivotal element of the NRF. All standard-based reports are derived and compiled from the PLs and PLDs as exemplified in these two tables.
After marking learner scripts and calculating individual total scores, and guided by the cutscores from the standard setting process, one simply determines the percentage of learners who perform at each of the levels in Table 20 or Table 21, whichever is applicable. This information can then be presented either in tabular or visual formats, e.g. in bar graphs or pie charts to guide the reader. Regardless of the format in which the results are presented, a standards-based report must be accompanied by the information with PLs and PLDs.

Standard-based reports for different audiences are designed and geared to inform possible interventions. For teachers the PLs and PLDs that characterise a specific category of learners have implications for the content that must either be revised or taught, the most appropriate pedagogical approach to follow, the types of learning support materials to use in learner interaction sessions and the types and levels of assessment complexity to follow up with. Standards-based reports make possible or lead to differentiated rather than one-size-fits-all approaches.

### 5.8.3 Interpretation of standards-based results

After setting cutscores and performance standards, test takers are categorised into specific PLs, based on their performance in the test. A typical standards-based report will show the percentage of learners who function at each PL. The report will be accompanied by a table of policy definitions and a table of PLs and PLDs for the relevant subject. Interpretation of results then proceeds logically in three easy steps that include:

a) identifying the PL in which a learner(s) are categorised in the report,

b) using the table of policy definitions to read off how the PL is defined, what the implications for progression and intervention are at this PL and, importantly,

c) using the table of PLs and PLDs (exemplified in Table 20 and Table 21), read off the specific knowledge and skills that the learner(s) can be expected to or not to demonstrate.

For example, if a standards-based report shows that x% of test takers function at the Partly Achieved level, reading off directly from the table of policy definitions (e.g. Table 19) shows that these learners:
Q. Moloi

i. demonstrate partial understanding of the knowledge and skills required to function at this grade level,

ii. are unlikely to succeed in the next grade without support, and

iii. require specific intervention to address knowledge gaps, while also requiring additional and continued support to progress to the required Achieved level.

The final step is to consult the subject-specific PLs and PLDs to read off directly what specific subject knowledge and skills these learners do or do not demonstrate. In the illustrative example of the $x\%$ learners who function at the Partly Achieved level, say in mathematics, one consults the relevant table of PLs and PLDs (e.g. Table 20) to read off directly that these learner(s) can do, among others and in addition to skills and knowledge in the lower Not Achieved level, :

i. count forward and backwards in decimals

ii. recognise place value up to 9 digits

iii. round off numbers up to 1000

...

In addition, one will know that these learners will not be able to demonstrate knowledge and skills that characterise the higher Achieved level.

To sum up, interpreting standards-based reports is basically a matter of reading off relevant information directly from the policy definitions and from the relevant PLs and PLDs. The user of the report is deliberately made to benefit immensely, in terms of the simplicity and user-friendliness of the documents, from the highly technical analysis and detailed compilation of cutscores, PLs and PLDs that is carried out by experts in the standard-setting phase. Otherwise the simplicity and user-friendliness do not in any way diminish the information-richness of standards-based reports.
5.9 **Key features of performance standards**

It is important to highlight key features of performance standards that give a unique character to the NRF. The performance standard indicates the level at which the learner is functioning by listing the specific knowledge and skills that the learner has mastered, is able to learn optimally and, therefore, the level at which effective teaching should be pitched. The PL also provides information on the percentage of learners in a class, school or district who function at a specific level.

The hierarchical nature of performance standards makes it possible for teachers to set challenging activities that can be given to learners at a particular PL in order to enhance their cognitive development without risking frustration or boredom. For instance, for a learner who functions at the Partly Achieved level, activities that are pitched at the Achieved level would be challenging enough with possible successful completion whereas activities pitched at the Advanced level would be far above the proficiency level of this learner and, therefore, likely to be frustrating. Conversely, activities that are pitched at levels that are below where the learner is ready to learn can be boring and detrimental to learning.

For each subject, the performance standard provides an indication of the content that must be taught and learnt at each PL. This is made possible by the involvement of subject specialists in the setting of standards. The implications for teaching which characterise each PL already suggest possible teaching methodologies to be considered for remediation or enrichment. Performance standards make it possible for teachers, school leaders and district officials to develop short-term, flexible and targeted curricula for continuing professional development of teachers. The implications for progression that are associated with each PL make it possible for schools and teachers to be proactive in providing additional support that addresses specific learning needs of learners.

5.10 **Setting of performance standards**

Immediately after the national assessment has been completed, the National DBE, which is entrusted with setting and monitoring standards, must appoint competent educators and assessment experts to conduct a standard setting process to determine cut-scores that mark the transition from one PL to the next on the knowledge and skills continuum.
Although there are various methods and combinations of standard setting methods, the study that led to the development of this NRF showed that the OSS method is a preferred choice for the South African context. Compared to two other methods that were used in the study, the OSS method found to be the most appropriate because it:

a) involves the lowest costs
b) was found to be easily repeatable and, therefore, reliable
c) produces the most consistent cutscores
d) its cutscores have the highest levels of precision or lowest margins of error
e) has the highest content validity, i.e. it is the most information-rich SS method
f) can be applied with readily available software
g) has the lowest turnaround time.

Using suitably qualified expertise, the DBE will, after the administration of each LSAS, facilitate the analysis of data and the writing of the essential elements in each of the national, provincial, district and school reports, templates for which are proposed in the next section. This approach is necessary for standardisation, which is one important principle underpinning the NRF.

5.11 Report templates

Reports form an integral part of the NRF. Typical examples of standards-based reports for national, provincial, district, school and learner levels are presented in the following sections. The examples are largely based on the mathematics and language results from the grade 6 ANA verification study of 2013, as well as simulated data to illustrate important design features of reporting performance standards in the NRF. In some graphs and tables percentages may not add up to 100% due to rounding off.

The NRF provides illustrative templates to be used to report assessment results at national, provincial, district and school levels. Use of easy-to-understand, short and precise reports (no more than 5 pages for a report) will encourage users to interact meaningfully with the
information. Assessment data in the reports shall be presented in user-friendly statistics, graphs and other visual formats and texts that are easy to understand, error-free, statistical jargon-free, factual and accurate in the information that they convey. Where applicable, a brief definition and explanation of special statistical data must be provided to assist readers to understand the totality of what the report communicates without being distracted by unfamiliar words, characters and mathematical or statistical formats. In addition, reports must be disseminated in both electronic and hard-copy formats.

The reports for the different levels of the education sector will, at a minimum, provide information that will help users at each level to answer the following questions:

a) What is the current level of learner performance in the provinces, districts, schools and classrooms?

b) How is performance distributed within provinces, districts, schools and classrooms?

c) Are there any gender-related differentials in learner performance?

d) Is performance uniform or differentiated across content sub-domains in each subject?

e) Are there any poverty-related differentials in learner performance?

From each cycle of assessment one national report will be produced and, where data is available, a report for each of the provinces, districts, schools and learners. For instance, if assessment was confined to representative samples of schools across provinces, the reporting will end at provincial level whereas if a universal assessment that involved all schools and learners was administered, reporting will be possible at all the mentioned levels.

For benchmarking purposes, a summary of performance at national and across all provinces must be shared with each province. Similarly, each district must receive a summary of national, provincial as well as other districts’ performance.

The reports must present information in different but appropriate formats for easy reading and understanding. For instance, brief texts will be appropriate for describing the knowledge and skills that learners at each PL are able to demonstrate in terms of performance standards. Statistical information such as percentages of learners who function at various PLs can be
tabulated. Bar graphs will be appropriate for representing trend information such as changes over time in percentages of learners who function at various PLs while pie charts will be most appropriate for representing the spread of performance at a given point in time. In a good balance, a mix of textual, visual and statistical formats, applied relevantly, will not only enhance the understanding of the reports but will also aid informed decision-making and appropriate action.

A copy of the performance standards as presented in Table 20 for mathematics and Table 21 for language shall be attached to and shall form an integral part of each of the national, provincial, district, school and individual learner report for easy reference. Each report shall be organised in sections that respond to the guiding questions in Section 5.101 above where appropriate. The following organisation and format of reporting shall apply at national, provincial, district, school and individual learner levels:

- **General information**: The particulars of the entity or level affected by the report shall appear at the top of the first page of the report. The year in which the assessment was conducted, the subject and the affected grade must also appear on the first page. An introduction guides the reader on how to interpret the information in the report.

- **Overall performance**: This section of the report must provide information that responds to the question: What is the current level of performance? Information for responding to this question shall be obtained from the percentages of learners who function at the various PLs as well as the PLDs. It shall be summarised in a clearly labelled pie chart to show, visually and numerically, the spread of learners across the PLs.

- **Performance distribution**: How is performance distributed across provinces, districts or schools? Information to answer this question will be tabulated, also presented in bar graphs, showing the percentage of learners functioning at various PLs per province in the national report, per district in the provincial report, and per school in the district report. This information is important for identifying provinces, districts and schools that require additional support with either resources or capacity building.

- **Performance by sub-groups**: This section of the report must identify sub-groups of
learners who may require additional support for improvement of learning. It must answer questions such as: Are there gender- or language-related differentials in learner performance? Ensuring equitable provision of education to all learners, regardless of gender or language, is a national priority. Information to help respond to this question shall be presented in a table that shows percentages of boys and girls and language of learning and teaching (LOLT) sub-groups functioning at the various PLs. Where learners are taught and tested in different languages, the percentages of language groups functioning at various PLs must also be represented in a tabular form, with tables clearly distinguished by appropriate headings and labels.

e. **Performance by content sub-domains**: Is performance uniform or differentiated across content sub-domains in each subject? Information for responding to this question must be summarised in tabular form showing the percentage of learners who function at various PLs in each subject sub-domain. For meaningful and targeted support it is important to identify specific sub-domains where there is deficiency and/or strength in specific skills and knowledge.

f. **Performance by quintile**: Social redress is one of the key priorities of government in South Africa. Information on differential performance of learners across quintiles is important for monitoring whether there are any poverty-related differentials in learner performance. Information to respond to this question will be presented in a tabular form showing percentages of learners from each quintile group who function at different PLs. This analysis must be conducted and reported at national and provincial levels but the numbers of schools and learners may be too limited to do the same at district and school levels. Specimens of national, provincial, district, school and learner reports are presented in Chapter 6.
6. EXEMPLARS OF STANDARDS-BASED REPORTS

6.1 Introduction

This chapter is dedicated to presenting the types and examplars of reports that are an integral part of the National Reporting Framework (NRF). The reports are based on performance standards and, where appropriate, basic statistical information is presented to augment information that is given in the policy definitions of performance levels (PL) and performance level descriptors (PLDs).

Each report is a complete self-contained entity and can be read independent of others. Therefore, the presentation of results in each report begins with two or three important pieces of information. The first piece of information that is common among all the reports is a table that presents the definitions of performance levels and their implications for learner progression and necessary interventions. This information is intended to assist the reader of the different (and separate) reports in interpreting the typical characteristics of learners classified by performance levels. The second and the third tables present detailed information on the specific knowledge and skills that a learner in each PL can be expected to demonstrate in mathematics and language, respectively.

6.2 Types of reports in the NRF

Five examplars are presented: a national, provincial, district, school and learner reports. Each report begins with tabulated information on policy definitions of PLs and their implications for learner progression and interventions. Depending on the target audience of the report, either both the mathematics and language PLDs are presented or at least one of the two is included as explained in the next section.

6.2.1 National report

The target audience of the national report are national policy- and decision-makers who are responsible for the monitoring and evaluation of education delivery across the provinces. As such, the national report should provide a synopsis of performance in all the subjects that are included in the large-scale assessment surveys (LSAS). The presentation of results in the national report is preceded by the table of policy definitions of PLs followed by the PLDs for
each subject. For instance, in assessments where only mathematics and a language are tested, the policy definitions are followed by mathematics and then language PLDs. The national report provides information that will help policy-makers in:

a) monitoring the overall performance of learners in the education system;

b) identifying provinces that require extraordinary support to improve learner performance; and,

c) tracking and redressing indications of inequitable delivery of education to all learners.

### 6.2.2 Provincial reports

The target audience of provincial reports are provincial decision-makers who are responsible for ensuring equitable delivery of high quality education in all schools in the province. The provincial report provides a synopsis of performance in all the subjects that are included in the LSAS. Like the national report, the presentation of results in the provincial report is preceded by the table of policy definitions of PLs followed by tabulated PLDs for each subject. The provincial report provides information that will help decision-makers in:

a) monitoring the overall performance of learners in the province;

b) benchmarking learner performance against other provinces with similar contextual conditions;

c) identifying districts that require extraordinary support to improve learner performance; and

d) tracking and redressing indications of inequitable delivery of education to all learners.

### 6.2.3 District reports

The target audience of district reports are district managers, circuit managers (where they exist), subject advisors and curriculum delivery support staff that interact with and support schools on a regular basis. Unlike the national and the provincial reports that provide a synoptic overview of performance across all the subjects that are included in the LSAS, the district report separates reporting per subject so that specialists in a subject can be able to
detach a relevant report and use it independent of their counterparts in other subjects. Like the national and provincial reports, the presentation of results in the district report is preceded by the table of policy definitions of PLs followed by tabulated PLDs for each subject. The district report provides information that will help education delivery and support staff in districts in:

a) supporting all schools in a district according to identified professional needs of teachers and learning needs of learners;

b) benchmarking learner performance across schools and districts in similar contexts;

c) identifying schools that require extraordinary support to improve learner performance; and

d) tracking and redressing indications of inequitable delivery of education to all learners.

### 6.2.4 School reports

The target audience of school reports are subject teachers, heads of schools, School-Management Teams (SMTs), parents and learners. Like the district report, the school report separates reporting per subject so that teachers of specific subjects can be able to detach a relevant report and use it independent of their counterparts in other subjects. Like the rest of the reports, the presentation of results in the school report is preceded by the table of policy definitions of PLs followed by tabulated PLDs for each subject.

The school report provides information that will help teachers in:

a) supporting individual learners to improve in specific subjects and subject sub-domains according to identified learning needs;

b) benchmarking learner performance across districts and provinces in similar contexts;

c) identifying learners that require extraordinary support to improve teaching and enhance learning; and

d) tracking the performance of individual learners over time.
6.2.5 Learner reports

The learner report provides important feedback to both the learner and their parents who must be encouraged to assist or motivate their children to grow interest in their school work. The report indicates the PL at which the learner is functioning in each subject. It also provides a brief summary of the progression and intervention implications of functioning at a particular PL in each subject. In addition to indicating the PL at which the learner functions, the learner report also provides information at subject sub-domain level. Where there may not be enough number of items in a test to cover a meaningful scope of each sub-domain, the results at this level can still be summarised in percentage mean scores to augment what has already been communicated in the policy definitions and in performance level descriptors.

The learner report provides information that will assist teachers, parents and the learner himself/herself to:

a) identify the unique learning needs of the learner and be able to provide individualised support;

b) monitor how the learner develops in relation to his/her peers in a class or grade; and

c) track the progress that the learner makes over time in the same grade and across grades
This report presents results of the large-scale assessment survey (LSAS) of 2013 regarding learner performance in mathematics and FAL at the grade 6 level. Information is reported using performance standards which provide information on the specific level at which a learner functions (Performance Level (PL) shown in Table 1) and the set of knowledge and skills that the learner could or could not display (Performance Level Descriptors (PLDs) shown in Table 2 and 3 for mathematics and FAL, respectively). It is important that the results be read in conjunction with Tables 1, 2 and 3.

### Table 1: Definition and implications of performance standards

<table>
<thead>
<tr>
<th>Performance Level</th>
<th>‘Performance Level’ definition</th>
<th>Progression Implications</th>
<th>Intervention implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced</td>
<td>comprehensive understanding of the knowledge and skills required to function at this grade level</td>
<td>has high likelihood of success in the next grade</td>
<td>requires little or no academic intervention but needs to be provided with more challenging tasks to maximise their full potential</td>
</tr>
<tr>
<td>Achieved</td>
<td>sufficient understanding of the knowledge and skills required to function at this grade level</td>
<td>has a reasonable likelihood of success in the next grade</td>
<td>may require some assistance with complex concepts to progress to the Advanced level</td>
</tr>
<tr>
<td>Partly Achieved</td>
<td>partial understanding of the knowledge and skills required to function at this grade level</td>
<td>is unlikely to succeed in the next grade without support</td>
<td>requires specific intervention to address knowledge gaps, while also requiring additional and continued classroom support to progress to the required Achieved level</td>
</tr>
<tr>
<td>Not Achieved</td>
<td>very limited understanding of the knowledge and skills required to function at this grade level</td>
<td>unlikely to succeed in the next grade without significant support</td>
<td>require specific intervention to address knowledge gaps, while also requiring additional teaching time and extensive and continued support within the classroom context, to progress to the required grade (achieved) level</td>
</tr>
</tbody>
</table>

### Table 2: Grade 6 mathematics PLs and PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>A learner at this level will, with intensive support, recognise basic number systems. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs:</td>
</tr>
<tr>
<td>- count forward only with whole numbers;</td>
<td>- count, recognise and do calculations using fractions &amp; decimals;</td>
<td>- count, recognise and do calculations using fractions &amp; decimals;</td>
<td>- critically read, interpret &amp; analyse with awareness of sources of error &amp; manipulation to draw conclusions &amp; make predictions;</td>
</tr>
<tr>
<td>- count objects not exceeding 10;</td>
<td>- represent multiples, factors &amp; prime numbers;</td>
<td>- represent multiples, factors &amp; prime numbers;</td>
<td>- distinguish between volume, surface area &amp; dimensions of rectangular prisms;</td>
</tr>
<tr>
<td>- add whole numbers up to 10;</td>
<td>- find percentages of whole numbers;</td>
<td>- find percentages of whole numbers;</td>
<td>- solve problems involving different time zones;</td>
</tr>
<tr>
<td>- draw simple pictures of objects;</td>
<td>- solve problems involving finances and measurement;</td>
<td>- solve problems involving finances and measurement;</td>
<td>- estimate, record, compare &amp; convert between SI units;</td>
</tr>
<tr>
<td>- Measure length of lines;</td>
<td>- compare rate and ratio;</td>
<td>- compare rate and ratio;</td>
<td>- identify &amp; describe numeric and geometric patterns;</td>
</tr>
<tr>
<td>- Name a few SI units;</td>
<td>- identify &amp; describe numeric and geometric patterns;</td>
<td>- identify &amp; describe numeric and geometric patterns;</td>
<td></td>
</tr>
<tr>
<td>- measure area &amp; perimeter of objects; and,</td>
<td>- record and compare volume, surface area &amp; dimensions of rectangular prisms;</td>
<td>- record and compare volume, surface area &amp; dimensions of rectangular prisms;</td>
<td></td>
</tr>
<tr>
<td>- draw simple bar graphs.</td>
<td>- draw simple bar graphs.</td>
<td>- draw simple bar graphs.</td>
<td></td>
</tr>
</tbody>
</table>
Q. Moloi

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>• measure using basic SI units; and,</td>
<td>• use and describe transformations;</td>
<td>(including mass, temperature, distance and capacity);</td>
<td>• organise &amp; record data;</td>
</tr>
<tr>
<td>• draw pictographs.</td>
<td>• locate and describe movement on a grid;</td>
<td>• calculate the median &amp; mode of data; and,</td>
<td>• list possible outcomes &amp; predict “likelihood” of events.</td>
</tr>
<tr>
<td>• differentiate between sample &amp; population; and,</td>
<td>• draw &amp; interpreting bar and pictographs.</td>
<td>• identify characters &amp; plot setting;</td>
<td>• use and describe transformations;</td>
</tr>
<tr>
<td>• locate and describe movement on a grid;</td>
<td>• use and describe transformations;</td>
<td>• identify ethical issues such as cultural/social diversity;</td>
<td>• evaluate texts;</td>
</tr>
<tr>
<td>• locate and describe movement on a grid;</td>
<td>• use and describe transformations;</td>
<td>• highlight the moral lesson behind a story;</td>
<td>• identify formal and informal texts;</td>
</tr>
<tr>
<td>• locate and describe movement on a grid;</td>
<td>• use and describe transformations;</td>
<td>• infer information from a complex text;</td>
<td>• easily utilise vocabulary of at least 5000 words;</td>
</tr>
<tr>
<td>• locate and describe movement on a grid;</td>
<td>• use and describe transformations;</td>
<td>• identify key elements of poetry;</td>
<td>• critique texts through book reviews &amp; reports;</td>
</tr>
<tr>
<td>• locate and describe movement on a grid;</td>
<td>• use and describe transformations;</td>
<td>• use dictionaries in building vocabulary of at least 3000 words;</td>
<td>• analyse both formal &amp; informal documents;</td>
</tr>
<tr>
<td>• locate and describe movement on a grid;</td>
<td>• use and describe transformations;</td>
<td>• understand factual information from non-fictional documents.</td>
<td>• synthesise information from different parts of a text; and,</td>
</tr>
<tr>
<td>• locate and describe movement on a grid;</td>
<td>• use and describe transformations;</td>
<td>• demonstrate comprehension of inferred information from a text.</td>
<td>• demonstrate comprehension of inferred information from a text.</td>
</tr>
</tbody>
</table>

Table 3: Grade 6 FAL PLs and PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>READING</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; interpret meaning of stories relating to real life experiences. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; interpret meaning of stories relating to real life experiences. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; interpret meaning of stories relating to real life experiences. She/he can:</td>
</tr>
<tr>
<td>• recognise words that were learnt previously;</td>
<td>• identify characters &amp; plot setting;</td>
<td>• identify characters &amp; plot setting;</td>
<td>• read poetry effectively;</td>
</tr>
<tr>
<td>• read forward &amp; answer short questions based on the text; and,</td>
<td>• identify ethical issues such as cultural/social diversity;</td>
<td>• highlight the moral lesson behind a story;</td>
<td>• evaluate texts;</td>
</tr>
<tr>
<td>• identify only the key character in a text.</td>
<td>• infer information from a complex text;</td>
<td>• use dictionaries in building vocabulary of at least 3000 words;</td>
<td>• identify formal and informal texts;</td>
</tr>
<tr>
<td>• use and describe transformations;</td>
<td>• identify key elements of poetry;</td>
<td>• understand factual information from non-fictional documents.</td>
<td>• easily utilise vocabulary of at least 5000 words;</td>
</tr>
<tr>
<td>• locate and describe movement on a grid;</td>
<td>• use and describe transformations;</td>
<td>• demonstrate comprehension of inferred information from a text.</td>
<td>• critique texts through book reviews &amp; reports;</td>
</tr>
<tr>
<td>• differentiate between sample &amp; population; and,</td>
<td>• locate and describe movement on a grid;</td>
<td>• analyse both formal &amp; informal documents;</td>
<td>• analyse both formal &amp; informal documents;</td>
</tr>
<tr>
<td>• draw &amp; interpreting bar and pictographs.</td>
<td>• locate and describe movement on a grid;</td>
<td>• synthesise information from different parts of a text; and,</td>
<td>• synthesise information from different parts of a text; and,</td>
</tr>
<tr>
<td>• locate and describe movement on a grid;</td>
<td>• locate and describe movement on a grid;</td>
<td>• demonstrate comprehension of inferred information from a text.</td>
<td>• demonstrate comprehension of inferred information from a text.</td>
</tr>
</tbody>
</table>

WRITING

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>A learner at this level will, with intensive support, show understanding through drawings and pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write information in graphic &amp; descriptive forms. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write information in graphical &amp; descriptive forms to text. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write &amp; develop information from graphical &amp; descriptive forms to text. He/she can:</td>
</tr>
<tr>
<td>• write single words &amp; short sentences;</td>
<td>• write for social purposes, e.g. personal letters, diaries, news reports;</td>
<td>• write extensively for social purposes;</td>
<td>• write extensively for social purposes;</td>
</tr>
<tr>
<td>• represent ideas with drawings;</td>
<td>• create a book cover;</td>
<td>• develop news reports;</td>
<td>• develop news reports;</td>
</tr>
<tr>
<td>• spell few commonly used words correctly; and,</td>
<td>• identify similar &amp; different texts;</td>
<td>• design questionnaires &amp; adverts;</td>
<td>• design questionnaires &amp; adverts;</td>
</tr>
<tr>
<td>• write brief diaries.</td>
<td>• express cause &amp; effect relationships; and,</td>
<td>• integrate ideas by classifying information;</td>
<td>• integrate ideas by classifying information;</td>
</tr>
<tr>
<td>• write brief diaries.</td>
<td>• write simple sentences &amp; draw pictures.</td>
<td>• solve problems;</td>
<td>• solve problems;</td>
</tr>
<tr>
<td>• write brief diaries.</td>
<td></td>
<td>• use relevant questioning styles to obtain information; and,</td>
<td>• use relevant questioning styles to obtain information; and,</td>
</tr>
<tr>
<td>• write brief diaries.</td>
<td></td>
<td>• listen with effective comprehension.</td>
<td>• listen with effective comprehension.</td>
</tr>
</tbody>
</table>
The results in Figure 1 indicate the following in relation to grade 6 level:

### Mathematics
- 45% of learners in South Africa showed very limited understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without significant support;
- 5% of learners demonstrated partial understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without support;
- Only 50% of learners demonstrated sufficient and comprehensive understanding of the knowledge and skills required to function at this grade level and have a reasonable to high likelihood of success in the next grade.
- There is need to increase both the number and the professional capacity of subject advisors so that they can support teachers in apply effective methods of teaching this critical subject.

### First Additional Language
- 35% of learners showed very limited understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without significant support;
- 25% of learners demonstrated partial understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without support;
- 50% of learners demonstrated sufficient and comprehensive understanding of the knowledge and skills required to function at this grade level and have a reasonable to high likelihood of success in the next grade.
- Teachers of English FAL, who are not speakers of the language, require focused training in effective methods of teaching a foreign language.
When aggregated by province, the mathematics results in Figure 2 indicate that:

- In Province P4, which is a predominantly urban province, 45% of learners demonstrated sufficient and comprehensive understanding of the knowledge and skills required to function at this grade level and, if everything remains the same, they have a reasonable to high likelihood of success in the next grade.

- In the rest of the provinces, between 40% and just under 50% of the learners showed very limited understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without significant support.

- Province P8, which is a predominantly rural province, is particularly at risk because almost half of the learners in this province are unlikely to succeed in the next grade without significant support as they demonstrate very limited understanding of the knowledge and skills required to function at this grade level.
Figure 3: Overall national performance in FAL

The FAL results in Figure 3 indicate that:

- In Provinces P1 and P8, around 50% of learners demonstrated sufficient and comprehensive understanding of the knowledge and skills required to function at this grade level and have a reasonable to high likelihood of success in the next grade.

- However, in Provinces P4 to P7, the majority of learners, between 60% and 70%, functioned between the Not Achieved and Partly Achieved levels. These learners demonstrated either very limited or partial understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without significant support.

- Generally, patterns of inequitable distribution of the quality of performance are evident and this phenomenon often affects the poorest and most marginalised people mainly in rural provinces like P4 and P6. The Department needs to accelerate redress measures.
Table 4: National performance by gender

<table>
<thead>
<tr>
<th>Performance Levels</th>
<th>Mathematics</th>
<th>First Additional Language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boys</td>
<td>Girls</td>
</tr>
<tr>
<td>Not Achieved</td>
<td>33</td>
<td>26</td>
</tr>
<tr>
<td>Partly Achieved</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>Achieved</td>
<td>33</td>
<td>40</td>
</tr>
<tr>
<td>Advanced</td>
<td>18</td>
<td>11</td>
</tr>
</tbody>
</table>

The results as reflected in Table 4 indicate marked gender differentials in that:

- **49%** of each of boys and girls functioned at the *Not Achieved* and *Partly Achieved* levels. These learners demonstrated either very limited or partial understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without significant support. The situation could be aggravated by the shortage of suitably qualified mathematics teachers, especially in rural areas.

- Generally, the same percentage of boys and girls (51% each) demonstrated sufficient to comprehensive understanding of the knowledge and skills required to function at this grade level and have a reasonable to high likelihood of success in the next grade.

- **The percentages** of boys and girls who functioned at the *Not Achieved* and *Partly Achieved* levels are **69%** and **59%**, respectively. These are learners who demonstrated either very limited or partial understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without significant support.

- **More girls** (41%) than boys (31%) demonstrated sufficient to comprehensive understanding of the knowledge and skills required to function at this grade level and have a reasonable to high likelihood of success in the next grade.

Table 5: National performance by LOLT

<table>
<thead>
<tr>
<th>Performance Levels</th>
<th>Mathematics (%)</th>
<th>First Additional Language (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Don’t speak LOLT</td>
<td>Speak LOLT</td>
</tr>
<tr>
<td></td>
<td>Don’t speak LOLT</td>
<td>Speak LOLT</td>
</tr>
<tr>
<td>Not Achieved</td>
<td>33</td>
<td>26</td>
</tr>
<tr>
<td>Partly Achieved</td>
<td>33</td>
<td>16</td>
</tr>
<tr>
<td>Achieved</td>
<td>23</td>
<td>40</td>
</tr>
<tr>
<td>Advanced</td>
<td>11</td>
<td>18</td>
</tr>
</tbody>
</table>

179
From Table 5, the results indicate that, in both mathematics and FAL, the fact that a learner speaks or does not speak the LOLT outside the teaching and learning environment has an influence on their performance. Specifically, the results indicate that:

**Mathematics**
- 66% of learners who do not speak the LOLT outside the formal teaching and learning environments, compared to 42% who do, function at the *Not Achieved* and *Partly Achieved* levels. They are unlikely to succeed in the next grade without significant support.
- 34% of those who do not speak the LOLT and 58% of those who speak the LOLT function at the *Achieved* and *Advanced* levels. These are the learners who have a reasonable to high likelihood of success in the next grade.

**First Additional Language**
- 60% of learners who do not speak the LOLT outside the formal teaching and learning environments, compared to 40% who do, function at the *Not Achieved* and *Partly Achieved* levels. They are unlikely to succeed in the next grade without significant support.
- 40% of those who do not speak the LOLT and 60% of those who speak the LOLT function at the Achieved and Advanced levels and they have a reasonable to high likelihood of success in the next grade.

**Figure 4: National performance in mathematics (mean score %) by sub-domains**

The results, presented in mean score percentages in Figure 4, indicate that:
Overall performance in the different mathematics sub-domains ranges from very low in the sub-domain of Probability to just satisfactory in Numbers & Operations.

The highest performance in mathematics, mean score of 56%, is in Numbers & Relationships, followed by Functions & Algebra with a mean score of 46%. Given that these two sub-domains have wide application in different areas of mathematics, the national policy-makers need to evaluate the content knowledge of teachers and encourage provinces to strengthen continuous professional development of mathematics teachers.

The lowest performance, and the most worrisome, is in Probability. This topic was introduced in grade 6 as part of the streamlined National Curriculum Statement (NCS).

A possible explanation for underperformance in this sub-domain could be that teachers of mathematics may not have been given adequate orientation into the latest version of the curriculum and thus many may have neglected the new topics. The national policy-makers need to ensure that curriculum changes are matched with commensurate support.

The results summarised in Figure 5 show that:

- learner performance in FAL is generally acceptable in Reading & Viewing, mean score of 60%, but the other sub-domains evidently require intensive intervention;
- the unacceptably low performance in Language Structure, mean score of 25%, may be ascribed to limited practice in regular written work in schools which may require that the numbers and utilisation of subject advisors be monitored more closely in the provinces;
• performance in Comprehension is very low and this is an area of concern because this skill affects learning in all subjects; and,

• there may be need to review methods of teaching FAL in school.

Figure 6: National performance in mathematics by quintile

As shown in Figure 6, performance varies widely across the quintiles and the following observations can be made:

• The highest percentage of learners who functioned at the Achieved and Advanced levels (60%) are in Quintile 5, compared to 10% for Q1 schools, and between 40 and 49% for Q2 to Q4 schools. These are learners who have a reasonable to high likelihood of success in the next grade.

• The majority of learners in Quintile 1 (90%) functioned at the Not Achieved and Partly Achieved levels. These are learners who demonstrate limited to partial understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without extended support.

• The marked disparities in performance among learners in Quintile 5 and Quintile 1 schools suggest that more decisive redress measures should be considered in ensuring that the poor and marginalised learners in relatively less-resourced schools are given extra support and that their teachers are targeted for focused and intensive continuous
professional development courses. It has been observed that Quintile 1 schools are generally fee-free schools and are thus less likely than their better-resourced counterparts in Quintile 5 to attract expert teachers. It is recommended that teachers of mathematics be incentivised financially to take up teaching posts in poor schools.

Figure 7: National performance in mathematics by quintile and sub-domains

The results in Figure 7, summarised in mean score percentages show that:

- performance in the various mathematics sub-domains generally decreases from Quintile 5 to Quintile 1 which suggests that the socio-economic status of the school that a learner attends may explain some of the overall performance that is observed;

- performance in Quintile 3 and Quintile 4 does not seem to differ much and, therefore, common rather than differentiated interventions may be considered in the two quintile sectors, an approach that could save state resources remarkably as more could be achieved with less; for instance, schools in these quintiles could share textbooks and other learning and teaching support materials (LTSM); and,

- in all quintiles the mean scores are highest in Number & Operations, invariably followed by Functions & Algebra, and the lowest in Probability and that these differentials are more pronounced and impact the heaviest on the marginalised learners in the poorest sector of the school system.
A more nuanced analysis of performance in the various sub-domains is given in Table 6.

Table 6: National Quintile performance (mean score %) by mathematics sub-domain

<table>
<thead>
<tr>
<th>Sub-domains</th>
<th>Not Achieved</th>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quintile 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number &amp; Operations</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>15</td>
<td>15</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>Measurement</td>
<td>10</td>
<td>25</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>Data Handling</td>
<td>20</td>
<td>10</td>
<td>25</td>
<td>45</td>
</tr>
<tr>
<td>Probability</td>
<td>10</td>
<td>10</td>
<td>30</td>
<td>50</td>
</tr>
<tr>
<td><strong>Quintile 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number &amp; Operations</td>
<td>30</td>
<td>35</td>
<td>45</td>
<td>55</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>Measurement</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Data Handling</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>Probability</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td><strong>Quintile 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number &amp; Operations</td>
<td>35</td>
<td>45</td>
<td>55</td>
<td>65</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Measurement</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>Data Handling</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Probability</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td><strong>Quintile 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number &amp; Operations</td>
<td>45</td>
<td>50</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>65</td>
</tr>
<tr>
<td>Measurement</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Data Handling</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Probability</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td><strong>Quintile 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number &amp; Operations</td>
<td>50</td>
<td>65</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>45</td>
<td>60</td>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td>Measurement</td>
<td>40</td>
<td>50</td>
<td>55</td>
<td>65</td>
</tr>
<tr>
<td>Data Handling</td>
<td>35</td>
<td>40</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Probability</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 6 presents information that will inform specific and appropriately targeted interventions. For example, in Quintile 1 and Quintile 2 the mean score percentages for Data Handling and Probability at the Achieved and Advanced levels range between 25% and 50%, respectively, whereas in Quintile 5 the range is between 40% and 55%. For optimal impact, a differentiated intervention strategy would be recommended for the two disparate categories.
of schools provided care is taken not to discriminate among schools. One way could be to rotate teachers among the quintiles after, say every three years.

A key observation that the Education Department should address is the significant differences among learners at the high PLs compared to those at the lower PLs. While the key challenge is to ensure that the learners at the Not Achieved and Partially Achieved levels are provided with relevant support to perform at the required grade level, it is also critical to ensure that the specific needs of learners at the Achieved and Advanced levels are addressed. At the very least, the danger of learners regressing into lower performance levels needs to be avoided.

Figure 8: National performance in FAL by quintile and sub-domains

From Figure 8, the results show that:

- performance in all the FAL sub-domains generally tends to decrease steeply from Quintile 5 to Quintile 1, which suggests that the socio-economic status of the school may explain a markedly great deal of learner performance; for instance, mean score percentages for Reading & Viewing decrease from 70% in Quintile 5 schools to 40% in Quintile 1 schools;

- the Quintile 1 mean score percentages in the critical competencies of Language Structure and Comprehension diminish so low that it would be doubtful if learners in this quintile could learn meaningfully and with understanding in any subject;
• in all quintiles the mean scores are highest in Reading & Viewing and the lowest in Comprehension; and,

• even in Q5 schools, learners performed poorly in the sub-domain of Comprehension, indicating a cause for concern at the higher grades for these learners.

From Table 7 the following observations can be made:

• Mean score percentages at the Not Achieved and Partly Achieved levels in Quintile 1 and Quintile 2 schools are particularly low, and lowest in the critical competencies of Language Structure and Comprehension. Learners in these categories of schools are taught in English as a first additional language although it is common knowledge that

<table>
<thead>
<tr>
<th>Sub-domains</th>
<th>Not Achieved</th>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quintile 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading &amp; Viewing</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>56</td>
</tr>
<tr>
<td>Language Structure</td>
<td>7</td>
<td>15</td>
<td>25</td>
<td>40</td>
</tr>
<tr>
<td>Comprehension</td>
<td>4</td>
<td>10</td>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td><strong>Quintile 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading &amp; Viewing</td>
<td>17</td>
<td>20</td>
<td>35</td>
<td>65</td>
</tr>
<tr>
<td>Language Structure</td>
<td>15</td>
<td>15</td>
<td>35</td>
<td>33</td>
</tr>
<tr>
<td>Comprehension</td>
<td>5</td>
<td>10</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td><strong>Quintile 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading &amp; Viewing</td>
<td>20</td>
<td>40</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Language Structure</td>
<td>25</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Comprehension</td>
<td>15</td>
<td>30</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td><strong>Quintile 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading &amp; Viewing</td>
<td>25</td>
<td>45</td>
<td>55</td>
<td>65</td>
</tr>
<tr>
<td>Language Structure</td>
<td>30</td>
<td>45</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Comprehension</td>
<td>20</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td><strong>Quintile 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading &amp; Viewing</td>
<td>35</td>
<td>50</td>
<td>65</td>
<td>75</td>
</tr>
<tr>
<td>Language Structure</td>
<td>35</td>
<td>45</td>
<td>55</td>
<td>65</td>
</tr>
<tr>
<td>Comprehension</td>
<td>30</td>
<td>45</td>
<td>50</td>
<td>55</td>
</tr>
</tbody>
</table>
these learners seldom if ever get opportunity to speak the language outside the teaching and learning contexts. The DBE needs to roll out the initiative to introduce English as a subject from grade 1 so that by the time learners reach grade 6 they shall have adequate exposure to the language that they will eventually use as a LOLT.

- Although learners in Quintile 5 schools tend to perform relatively higher than their counterparts in the lower quintiles, the results indicate that Quintile 5 learners also experience difficulties in Comprehension. Their mean score percentage in Comprehension was the lowest compared to the other sub-domains.

- It is remarkable to observe that there are learners in Quintiles 1 and 2 who obtained mean scores up to 65% and function at the Advanced level notwithstanding the socio-economic conditions of their schools. Teachers, with the support of subject advisors, need to ensure that these learners are assisted to maintain and improve on this performance.
PROVINCIAL REPORT

Large-scale Assessment of Learners – 2013

Grade 6 Mathematics and First Additional Language

This report presents the results from the large-scale assessment survey (LSAS) of 2013 on provincial levels and quality of performance in mathematics and FAL at the grade 6 level. Information is reported using performance standards. A “performance standard” describes performance in terms of the level at which a learner functions (Performance Level (PL)) shown in Table 1 and the set of knowledge and skills that the learner could or could not display (Performance Level Descriptors (PLDs) shown in Table 2 & 3 for mathematics and FAL, respectively. It is important that the results be read in conjunction with Tables 1, 2 and 3.

Table 1: Definition and implications of performance standards

<table>
<thead>
<tr>
<th>Performance Levels</th>
<th>PL definition</th>
<th>Progression Implications</th>
<th>Intervention implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced</td>
<td>comprehensive understanding of the knowledge and skills required to function at this grade level</td>
<td>has high likelihood of success in the next grade</td>
<td>requires little or no academic intervention but needs to be provided with more challenging tasks to maximise their full potential</td>
</tr>
<tr>
<td>Achieved</td>
<td>sufficient understanding of the knowledge and skills required to function at this grade level</td>
<td>has a reasonable likelihood of success in the next grade</td>
<td>may require some assistance with complex concepts to progress to the Advanced level</td>
</tr>
<tr>
<td>Partly Achieved</td>
<td>partial understanding of the knowledge and skills required to function at this grade level</td>
<td>is unlikely to succeed in the next grade without support</td>
<td>requires specific intervention to address knowledge gaps, while also requiring additional and continued classroom support to progress to the required Achieved level</td>
</tr>
<tr>
<td>Not Achieved</td>
<td>very limited understanding of the knowledge and skills required to function at this grade level</td>
<td>unlikely to succeed in the next grade without significant support</td>
<td>require specific intervention to address knowledge gaps, while also requiring additional teaching time and extensive and continued support within the classroom context, to progress to the required grade (achieved) level</td>
</tr>
</tbody>
</table>
Table 2: Grade 6 mathematics PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>A learner at this level will, with intensive support, recognise basic number systems. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL:</td>
</tr>
<tr>
<td>- count forward only with whole numbers;</td>
<td>- count forward and backwards in decimals;</td>
<td>- count, recognise and do calculations using fractions &amp; decimals;</td>
<td>- critically read, interpret &amp; analyse with awareness of sources of error &amp; manipulation to draw conclusions &amp; make predictions;</td>
</tr>
<tr>
<td>- count objects not exceeding 10;</td>
<td>- recognise place value up to 9 digits;</td>
<td>- represent multiples, factors &amp; prime numbers;</td>
<td>- list possible outcomes for simple experiments including tossing a coin, rolling a die &amp; spinning a spinner;</td>
</tr>
<tr>
<td>- add whole numbers up to 10;</td>
<td>- round off number up to 1000;</td>
<td>- find percentages of whole numbers;</td>
<td>- distinguish between volume, surface area &amp; dimensions of rectangular prisms;</td>
</tr>
<tr>
<td>- draw simple pictures of objects;</td>
<td>- add and subtract up to 9 digits;</td>
<td>- solve problems involving finances and measurement;</td>
<td>- solve problems involving different time zones;</td>
</tr>
<tr>
<td>- Measure length of lines;</td>
<td>- do simple calculations using ordinary fractions and decimals;</td>
<td>- compare rate and ratio;</td>
<td>- estimate, record, compare &amp; convert between SI units (including mass, temperature, distance and capacity);</td>
</tr>
<tr>
<td>- Name few SI units;</td>
<td>- read digital and analogue time;</td>
<td>- identify &amp; describe numeric and geometric patterns;</td>
<td>- organise &amp; record data;</td>
</tr>
<tr>
<td>- measure area &amp; perimeter of objects; and,</td>
<td>- measure using basic SI units;</td>
<td>- use and describe transformations;</td>
<td>- calculate the median &amp; mode of data; and,</td>
</tr>
<tr>
<td>- draw simple bar graphs.</td>
<td>- and,</td>
<td>- locate and describe movement on a grid;</td>
<td>- list possible outcomes &amp; predict “likelihood” of events.</td>
</tr>
<tr>
<td></td>
<td>draw pictographs.</td>
<td>- differentiate between sample &amp; population;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- draw &amp; interpreting bar and pictographs.</td>
<td></td>
</tr>
</tbody>
</table>
### Table 3: Grade 6 FAL PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, read single words, short sentences &amp; pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL, read texts with simple sentences. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; interpret meaning of stories relating to real life experiences. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; understand complex texts in formal &amp; informal stories. She/he can:</td>
</tr>
<tr>
<td>• recognise words that were learnt previously;</td>
<td>• skim &amp; scan some poetical elements;</td>
<td>• identify characters &amp; plot setting;</td>
<td>• read poetry effectively;</td>
</tr>
<tr>
<td>• read forward &amp; answer short questions based on the text; and,</td>
<td>• skim, scan &amp; summarise texts;</td>
<td>• identify ethical issues such as cultural/social diversity;</td>
<td>• evaluate texts;</td>
</tr>
<tr>
<td>• identify only the key character in a text.</td>
<td>• use vocabulary of 1500 words;</td>
<td>• highlight the moral lesson behind a story;</td>
<td>• identify formal and informal texts;</td>
</tr>
<tr>
<td></td>
<td>• write friendly letters;</td>
<td>• infer information from a complex text;</td>
<td>• easily utilise vocabulary of at least 5000 words;</td>
</tr>
<tr>
<td></td>
<td>• extract information directly from a short text; and,</td>
<td>• identify key elements of poetry;</td>
<td>• critique texts through book reviews &amp; reports;</td>
</tr>
<tr>
<td></td>
<td>• read short stanzas &amp; answer literal questions.</td>
<td>• use dictionaries in building vocabulary of at least 3000 words; and,</td>
<td>• analyse both formal &amp; informal documents;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• understand factual information from non-fictional documents.</td>
<td>• synthesise information from different parts of a text; and,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• demonstrate comprehension of inferred information from a text.</td>
</tr>
<tr>
<td><strong>WRITING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, show understanding through drawings and pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL, write for social purposes. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write information in graphic &amp; descriptive forms. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write &amp; develop information from graphical &amp; descriptive forms to text. He/she can:</td>
</tr>
<tr>
<td>• write single words &amp; short sentences;</td>
<td>• write personal letters, diaries, news reports;</td>
<td>• write for social purposes, e.g. personal letters, diaries, dialogues &amp; simple news reports;</td>
<td>• write extensively for social purposes;</td>
</tr>
<tr>
<td>• represent ideas with drawings;</td>
<td>• create a book cover;</td>
<td>• design a book cover;</td>
<td>• develop news reports;</td>
</tr>
<tr>
<td>• spell few commonly used words correctly; and,</td>
<td>• identify similar &amp; different texts;</td>
<td>and,</td>
<td>• design questionnaires &amp; adverts;</td>
</tr>
<tr>
<td>• write brief diaries.</td>
<td>• express cause &amp; effect relationships; and,</td>
<td>• develop &amp; edit key language structures.</td>
<td>• integrate ideas by classifying information;</td>
</tr>
<tr>
<td></td>
<td>• write simple sentences &amp; draw pictures.</td>
<td></td>
<td>• solve problems;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• use relevant questioning styles to obtain information; and,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• listen with effective comprehension.</td>
</tr>
</tbody>
</table>
Figure 1: Overall provincial performance

The results in Figure 1 indicate that at grade 6 level:

**Mathematics**
- 45% of learners in South Africa show very limited understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without significant support.
- 35% of learners demonstrate partial understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without support.
- only 20% of learners demonstrate sufficient and comprehensive understanding of the knowledge and skills required to function at this grade level and have a reasonable to high likelihood of success in the next grade.

**First Additional Language**
- 42% of learners show very limited understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without significant support.
- 27% of learners demonstrate partial understanding of the knowledge and skills required to function at this grade level and are unlikely to succeed in the next grade without support.
- 31% of learners demonstrate sufficient and comprehensive understanding of the knowledge and skills required to function at this grade level and have a reasonable to high likelihood of success in the next grade.
Figure 2: Provincial performance in mathematics by district

The provincial mathematics results in Figure 2, broken down to district level, show that:

- In District D5 45% of learners functioned at the *Achieved* and *Advanced* levels and, therefore, demonstrated sufficient and comprehensive understanding of the knowledge and skills required to function at this grade level and thus have a reasonable to high likelihood of success in the next grade.

- Next to District D5, District D3 had a reasonable percentage of learners (38%) who also functioned at the *Achieved* and *Advanced* levels and thus demonstrated sufficient and comprehensive understanding of the knowledge and skills required to function at this grade level so that they have a reasonable to high likelihood of success in the next grade.

- The lowest levels of performance were in District D4 where 80% of learners functioned at the *Not Achieved* and *Partly Achieved* levels. These are learners who demonstrated limited to partial understanding of the knowledge and skills required to function at this grade level and are, consequently, unlikely to succeed in the next grade without extensive and properly targeted support.
For FAL the provincial results, reported at district level in Figure 3, indicate that:

- 40% of learners in District D2 and 38% in D4 functioned at the *Achieved* and *Advanced* levels. These are learners who demonstrated sufficient and comprehensive understanding of the knowledge and skills required to function at this grade level and thus have a reasonable to high likelihood of success in the next grade.

- In the rest of the districts there were fewer learners who functioned at the acceptable levels, viz. *Achieved* and *Advanced* levels

- In District D5 more than 70% of learners functioned at the *Not Achieved* and *Partly Achieved* levels. This is of great concern because it means that the majority of learners in this district are unlikely to succeed in the next grade without extensive and properly targeted support.
Table 4: Provincial performance by gender and LOLT

<table>
<thead>
<tr>
<th>Performance Levels</th>
<th>Mathematics</th>
<th>FAL</th>
<th>Mathematics</th>
<th>FAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Girls</td>
<td>Boys</td>
<td>Girls</td>
<td>Boys</td>
</tr>
<tr>
<td>Not Achieved</td>
<td>25</td>
<td>32</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>Partly Achieved</td>
<td>53</td>
<td>51</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>Achieved</td>
<td>13</td>
<td>10</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Advanced</td>
<td>9</td>
<td>7</td>
<td>15</td>
<td>10</td>
</tr>
</tbody>
</table>

The results in Table 4 show that:

**Results by gender**

- 78% of girls, compared to 83% of boys, functioned at the Not Achieved and Partly Achieved levels.
- Only 22% of girls, compared to 17% of boys, functioned at the Achieved and Advanced levels.
- There is need for gender-sensitive interventions in the teaching and learning of mathematics in schools.

**Results by LOLT**

- 69% of learners who do not speak the LOLT, compared to 35% who do, functioned at the Not Achieved and Partly Achieved levels.
- 31% of learners who do not speak the LOLT, compared to 65% who do, functioned at the Achieved and Advanced levels.
- There is need to accelerate initiatives of prioritising the training of teachers for the majority of schools that teach in English as a first additional language.

**First Additional Language**

- 60% of girls, compared to 80% of boys, functioned at the Not Achieved and Partly Achieved levels.
- 40% of girls, compared to only 20% of boys, functioned at the Achieved and Advanced levels.
- Boys must be motivated to participate in regular literacy and

- 69% of learners who do not speak the LOLT, compared to 35% who do, functioned at the Not Achieved and Partly Achieved levels.
- 31% of learners who do not speak the LOLT, compared to 65% who do, functioned at the Achieved and Advanced levels.
- There is need to accelerate initiatives of prioritising the training of teachers for the majority of schools that teach in English as a first additional language.

- 50% of learners who do not speak the LOLT, compared to 30% who do, functioned at the Not Achieved and Partly Achieved levels.
- 50% of learners who do not speak the LOLT, compared to 70% who do, functioned at the Achieved and Advanced levels.
- Non-speakers of English who use it as a LOLT must be exposed to a variety of reading materials to expand their vocabulary and use of
reading programmes.

- Generally, more attention should be given to assisting boys to improve their skills in both numeracy and literacy.

- The results indicate that exposure or lack of exposure to the LOLT needs to be addressed as a priority as it affects the majority of learners.

Figure 4: Provincial performance in mathematics by sub-domains

From the presented results in Figure 4 the following observations can be made:

- The sub-domain of Numbers & Relationships seems to be most understood (mean score of 65%) followed by Functions & Algebra at just 50%.

- Performance in the rest of the sub-domains is low to very low, ranging from a mean score of 25% in Data Handling to a mere 10% in Probability.

- Probability seems to be a challenging area not only in the province but also at national level.
The results in Figure 5 indicate that:

- Except for the sub-domain of Reading & Viewing, with a mean score of 68%, performance in FAL is generally low in this province.
- Comprehension skills, which are critical for learning across all areas, are unacceptably low at mean score of 33%.
- Although performance in Language Structure is slightly higher that in Comprehension, it remains too low for a FAL.

Figure 6: Provincial performance in mathematics by quintile
The results in Figure 6 indicate that:

- There is a huge disparity in performance from Quintile 1 through Quintile 5, which could suggest inequity in other factors that influence learning and teaching.

- An overwhelming 71% of learners in Quintile 1 functioned at the Not Achieved and Partly Achieved levels while only 40% of their counterparts in Quintile 5 function at these levels. This means that almost 7 out of every 10 Quintile 1 learners in this province, compared to 4 in Quintile 5, are unlikely to succeed in the next grade without extensive and properly targeted support.

![Figure 7: Provincial performance in FAL by quintile](image)

From the presented results in Figure 7, the following observations can be made:

- 60% of learners in Quintile 5, compared to 25% in Quintile 1, functioned at the Achieved and Advanced levels in FAL. Learners in these PLs stand reasonable to high likelihood of success in the next grade.

- It is interesting to note that the percentages of learners who function at the Not Achieved and Partly Achieved PLs in Quintiles 3 and 4, 41% and 46%, respectively, are fairly similar even though socio-economic conditions in these two categories of schools are often very dissimilar.
From Figure 8 the results show that:

- Learners in Quintile 5 schools tend to perform relatively well and much higher than their counterparts in Quintile 1 and Quintile 2 in all the sub-domains.

- In all quintiles performance is distinctly highest in the sub-domain of Number & Operations followed by Functions & Algebra and lowest in Probability.

- Quantiles 1 and 2 seem to be particularly under-performing in this province. Focused redress measures may be necessary to provide impetus for improvement in this sector, as this result has serious implications for delivery of an equitable system of high quality education.

- Point about low performance in Q5 schools in Probability and implications for intervention.
From Figure 9, the results show that:

- There is a marked disparity between performance in Quintile 5 and Quintile 1.
- Performance in the sub-domains generally decreases from Quintile 5 to Quintile 1 which suggests that the socio-economic status of the school may explain some of the performance.
- Learner mean scores are the highest in Reading & Viewing and the lowest in Comprehension across all five quintiles.
- Low levels of performance across the Comprehension sub-domain indicate that the problem could be of national rather than provincial proportions. Interventions to be considered could include:
  
  i. engaging services of experts to ensure that the literacy and language national curricula are appropriate and correctly interpreted in schools;
  
  ii. adopting modern methods of teaching literacy and language, especially to non-speakers of a language that is used as a LOLT; and,
  
  iii. encouraging teaching of language and comprehension across all subjects.
DISTRICT REPORT
Large-scale Assessment of Learners – 2013
Grade 6 Mathematics and First Additional Language

This report presents the results from the large-scale assessment survey (LSAS) of 2013 on district levels of performance in mathematics and English First Additional Language (FAL) at the grade 6 level. Information is reported using performance standards. A “performance standard” describes performance in terms of the level at which a learner functions (Performance Level (PL)) shown in Table 1 and the set of knowledge and skills that the learner could or could not display (Performance Level Descriptors (PLDs) shown in Table 2 for mathematics and in Table 3 for FAL. It is important that the results be read in conjunction with Tables 1, 2 and 3.

Table 1: Definition and implications of performance standards

<table>
<thead>
<tr>
<th>Performance Levels</th>
<th>PL definition</th>
<th>Progression Implications</th>
<th>Intervention implications</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A learner at this PL demonstrates:</td>
<td>The learner:</td>
<td>The learner:</td>
</tr>
<tr>
<td>Advanced</td>
<td>comprehensive understanding of the knowledge and skills required to function at this grade level</td>
<td>has high likelihood of success in the next grade</td>
<td>requires little or no academic intervention but needs to be provided with more challenging tasks to maximise their full potential</td>
</tr>
<tr>
<td>Achieved</td>
<td>sufficient understanding of the knowledge and skills required to function at this grade level</td>
<td>has a reasonable likelihood of success in the next grade</td>
<td>may require some assistance with complex concepts to progress to the Advanced level</td>
</tr>
<tr>
<td>Partly Achieved</td>
<td>partial understanding of the knowledge and skills required to function at this grade level</td>
<td>is unlikely to succeed in the next grade without support</td>
<td>requires specific intervention to address knowledge gaps, while also requiring additional and continued classroom support to progress to the required Achieved level</td>
</tr>
<tr>
<td>Not Achieved</td>
<td>very limited understanding of the knowledge and skills required to function at this grade level</td>
<td>unlikely to succeed in the next grade without significant support</td>
<td>require specific intervention to address knowledge gaps, while also requiring additional teaching time and extensive and continued support within the classroom context, to progress to the required level</td>
</tr>
</tbody>
</table>

Figure 1: District performance in mathematics

- % Not Achieved
- % Partly Achieved
- % Achieved
- % Advanced
Table 2: Grade 6 mathematics PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
</table>
| A learner at this level will, with intensive support, recognise basic number systems. He/she 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; and,  
• draw simple bar graphs. | A learner at this level can, in addition to skills & knowledge in the lower PL:  
• 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; and,  
• draw & interpreting bar and pictographs. | A learner at this level can, in addition to skills & knowledge in the lower PLs:  
• critically read, interpret & analyse with awareness of sources of error & manipulation to draw conclusions & make predictions;  
• list possible outcomes for simple experiments including tossing a coin, rolling a die & 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; and,  
• list possible outcomes & predict “likelihood” of events. |

From Figure 1, the results indicate that:

- 55% of learners in the district functioned at the Not Achieved and Partly Achieved levels, with an overwhelming majority at the Not Achieved level. These learners demonstrated very limited to partial understanding of the knowledge and skills required to function at this grade level, and are unlikely to succeed in the next grade without significant support.

---

4 For specific skills and knowledge that learners at this PL can demonstrate see Table 2.
45% of learners functioned at the *Achieved* and *Advanced* levels. They demonstrated sufficient to comprehensive understanding of the knowledge and skills required to function at this grade level and have a reasonable likelihood of success in the next grade.

**Table 3: District mathematics performance by gender and LOLT**

<table>
<thead>
<tr>
<th>Performance Levels</th>
<th>Mathematics</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Girls</td>
<td>Boys</td>
</tr>
<tr>
<td>Not Achieved</td>
<td>40</td>
<td>48</td>
</tr>
<tr>
<td>Partly Achieved</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td>Achieved</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>Advanced</td>
<td>18</td>
<td>8</td>
</tr>
</tbody>
</table>

The results in Table 3 show that:

**Mathematics results by gender**

- 60% of girls, compared to 80% of boys, functioned at the *Not Achieved* and *Partly Achieved* levels.
- 40% of girls, compared to only 20% of boys, functioned at the *Achieved* and *Advanced* levels.
- There is need for gender-sensitive interventions in the teaching and learning of mathematics in schools in this district. The majority of boys are not likely to succeed in mathematics.

**Mathematics results by LOLT**

- 84% of learners who do not speak the LOLT, compared to 7% who do, functioned at the *Not Achieved* and *Partly Achieved* levels.
- Only 16% of learners who do not speak the LOLT, compared to 93% who do, functioned at the *Achieved* and *Advanced* levels.
- Language-related disparities in mathematics performance are too stark in this district. There is need to consider focused continuing professional development of teachers in schools that where LOLT is not a commonly-spoken language.

---

5 For specific skills and knowledge that learners at this PL can demonstrate see Table 2.
Figure 2: District mathematics performance by quintile

The results in Figure 2 indicate that:

- There are very similar patterns of performance in Quintile 1 and Quintile 2 schools across this district. Similarly, learners in Quintile 4 and Quintile 5 function in very similar patterns.

- 75% of learners in each of Quintile 1 and Quintile 2 functioned at the *Not Achieved* and *Partly Achieved* levels. These learners are unlikely to succeed in the next grade without significant support or at least some support.

- Approximately 40% of learners in Quintiles 4 and 5 function at the *Not Achieved* and *Partly Achieved* levels. These learners are unlikely to succeed in the next grade without significant support or at least some support.

- 25% of learners in each of Quintile 1 and Quintile 2 schools in this district functioned at the *Achieved* and *Advanced* levels. This is a remarkable observation. It has significant implications for improving the success rate of the poor and marginalised learners. It shows that:
  
  i. there is talent to be recognised and nurtured both among teachers and learners in these disadvantaged environments;
ii. teachers in these disadvantaged environments need to be supported by subject advisors to sustain and enhance their professional competencies; and,

iii. learners need to be motivated and provided with appropriate learning materials to develop their skills and knowledge.

Figure 3: District mathematics performance by sub-domains

As shown in Figure 3, the following refer:

- Generally, performance ranged from very low to low in all the mathematics sub-domains with all mean scores for all sub-domains falling below 50%.

- The mean score was 45% for Number and Operations and this was the highest performance among all the mathematics sub-domains.

- Performance in Probability and Measurement was a mean score of 12% and 15%, respectively. Implications for intervention suggest that urgent action is required from teachers first and also from learners.

- Subject advisors and curriculum support staff at the district level need to ensure that teachers have the requisite content knowledge in the subjects that they teach which seems to be a problem in this district.
The results in Figure 4 indicate that:

- There is a marked discrepancy in learner performance in Quintiles 1 and 2 on the one extreme and Quintiles 4 and 5 on the other extreme.

- In all the quintiles, the highest mean scores are in Number and Operations followed by Functions and Algebra.

- Except in Quintile 2, the lowest mean scores tend to be on the sub-domain of Probability.
# First Additional Language

## Table 3: Grade 6 FAL PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, read single words, short sentences &amp; pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL, read texts with simple sentences. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; interpret meaning of stories relating to real life experiences. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; understand complex texts in formal &amp; informal stories. She/he can:</td>
</tr>
<tr>
<td>- recognise words that were learnt previously</td>
<td>- skim &amp; scan some poetical elements</td>
<td>- identify characters &amp; plot setting</td>
<td>- read poetry effectively</td>
</tr>
<tr>
<td>- read forward &amp; answer short questions based on the text</td>
<td>- skim, scan &amp; summarise texts</td>
<td>- identify ethical issues such as cultural/ social diversity</td>
<td>- evaluate texts</td>
</tr>
<tr>
<td>- identify only the key character in a text</td>
<td>- use vocabulary of 1500 words.</td>
<td>- highlight the moral lesson behind a story</td>
<td>- identify formal and informal texts</td>
</tr>
<tr>
<td></td>
<td>- write friendly letters</td>
<td>- infer information from a complex text</td>
<td>- easily utilise vocabulary of at least 5000 words</td>
</tr>
<tr>
<td></td>
<td>- extract information directly from a short text</td>
<td>- identify key elements of poetry</td>
<td>- critique texts through book reviews &amp; reports.</td>
</tr>
<tr>
<td></td>
<td>- read short stanzas &amp; answer literal questions</td>
<td>- use dictionaries in building vocabulary of at least 3000 words</td>
<td>- analyse both formal &amp; informal documents</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- synthesise information from different parts of a text</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- demonstrate comprehension of inferred information from a text</td>
</tr>
<tr>
<td><strong>WRITING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, show understanding through drawings and pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL, write for social purposes. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write information in graphic &amp; descriptive forms. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write &amp; develop information from graphical &amp; descriptive forms to text. He/she can:</td>
</tr>
<tr>
<td>- write single words &amp; short sentences</td>
<td>- write personal letters, diaries, news reports</td>
<td>- write for social purposes, e.g. personal letters, diaries, dialogues &amp; simple news reports</td>
<td>- write extensively for social purposes</td>
</tr>
<tr>
<td>- represent ideas with drawings</td>
<td>- create a book cover</td>
<td>- design a book cover</td>
<td>- develop news reports</td>
</tr>
<tr>
<td>- spell few commonly used words correctly</td>
<td>- identify similar &amp; different texts</td>
<td>- develop &amp; edit key language structures.</td>
<td>- design questionnaires &amp; adverts.</td>
</tr>
<tr>
<td>- write brief diaries</td>
<td>- express cause &amp; effect relationships</td>
<td></td>
<td>- integrate ideas by classifying information</td>
</tr>
<tr>
<td></td>
<td>- write simple sentences &amp; draw pictures.</td>
<td></td>
<td>- solve problems</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- use relevant questioning styles to obtain information</td>
</tr>
</tbody>
</table>
| | | | - listen with effective comprehension.
From Figure 1, the district results for First Additional Language indicate that:

- 48% of learners functioned at the *Not Achieved* and *Partly Achieved*\(^6\) levels. These learners demonstrated very limited to partial understanding of the knowledge and skills required to function at this grade level, and are unlikely to succeed in the next grade without significant support;

- 52% of learners functioned at the *Achieved* and *Advanced*\(^7\) levels. They demonstrated sufficient to comprehensive understanding of the knowledge and skills required to function at this grade level and have a reasonable likelihood of success in the next grade.

### Table 4: District FAL performance by gender and LOLT

<table>
<thead>
<tr>
<th>Performance Levels</th>
<th>First Additional Language</th>
<th>First Additional Language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Girls</td>
<td>Boys</td>
</tr>
<tr>
<td>Not Achieved</td>
<td>55</td>
<td>41</td>
</tr>
<tr>
<td>Partly Achieved</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>Achieved</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Advanced</td>
<td>11</td>
<td>18</td>
</tr>
</tbody>
</table>

The results in Table 4 show that:

#### FAL results by gender

- 84% of girls, compared to 73% of boys, functioned at the *Not Achieved* and *Partly Achieved* levels.

#### FAL results by LOLT

- 70% of learners who do not speak the LOLT, compared to 25% who do, functioned at the *Not Achieved* and *Partly Achieved* levels.

---

\(^6\) For specific skills and knowledge that learners at this PL can demonstrate see Table 2.

\(^7\) For specific skills and knowledge that learners at this PL can demonstrate see Table 2.
16% of girls, compared to only 28% of boys, functioned at the Achieved and Advanced levels.

Performance in general is low but particularly worrying for girls. There is need for gender-sensitive interventions in the teaching and learning of FAL. The majority of girls are not likely to succeed in FAL.

30% of learners who do not speak the LOLT, compared to 75% who do, functioned at the Achieved and Advanced levels.

There are huge disparities in FAL performance with learners who do not speak the LOLT outside the teaching and learning environment being at great disadvantage. There is need to consider focused continuing professional development of teachers in schools where LOLT is not a commonly-spoken language.

**Figure 2: District FAL performance by quintile**

The results in Figure 2 indicate that:

- Between 75% and 73% of learners in each of Quintile 1 and Quintile 2 functioned at the Not Achieved and Partly Achieved levels. These learners are unlikely to succeed in the next grade without significant support or at least some support

- 54% of learners in Quintiles 4 and 30% in Quintile 5 functioned at the Not Achieved and Partly Achieved levels. These learners are unlikely to succeed in the next grade without significant support or at least some support.

- Overall, there is a marked disparity between performance in Quintile 1 and in Quintile 5.
Figure 3: District FAL performance by sub-domains

As shown in Figure 3:

- Generally, performance ranged from very low to low in all the FAL sub-domains.

- The mean score was 58% for Reading & Viewing and this was the highest performance among all the FAL sub-domains.

- Performance in Language Structure and Comprehension was a mean score of 30% and 12%, respectively. It is worrisome that Comprehension, a skill that applies across all learning areas, seems to be so low among the learners.

Figure 4: District FAL performance by quintile and sub-domains

The results in Figure 4 indicate that:

- There is a marked discrepancy in learner performance in Quintiles 1 and 2 on the one extreme and Quintiles 4 and 5 on the other extreme.
Q. Moloi

- In all the quintiles, the highest mean scores are in Reading & Viewing followed by Language Structure.

- The lowest mean scores are on Comprehension, including in Q5 schools where mean scores were only 40%.
SCHOOL REPORT

School name: City Primary School

Large-scale Assessment of Learners – 2013

Grade 6 mathematics and English First Additional Language

This report presents the results from the large-scale assessment survey (LSAS) of 2013 on the school’s levels of performance in mathematics and English First Additional Language (FAL) at the grade 6 level. Information is reported using performance standards. A “performance standard” describes performance in terms of the level at which a learner functions (Performance Level (PL)) shown in Table 1 and the set of knowledge and skills that the learner could or could not display (Performance Level Descriptors (PLDs) shown in Table 2 for mathematics and in Table 3 for FAL. It is important that the results be read in conjunction with Tables 1, 2 and 3.

Table 1: Definition and implications of performance standards

<table>
<thead>
<tr>
<th>Performance Levels</th>
<th>PL definition</th>
<th>Progression Implications</th>
<th>Intervention implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced</td>
<td>comprehensive understanding of the knowledge and skills required to function at this grade level.</td>
<td>The learner: has high likelihood of success in the next grade.</td>
<td>requires little or no academic intervention but needs to be provided with more challenging tasks to maximise their full potential.</td>
</tr>
<tr>
<td>Achieved</td>
<td>sufficient understanding of the knowledge and skills required to function at this grade level.</td>
<td>The learner: has a reasonable likelihood of success in the next grade.</td>
<td>may require some assistance with complex concepts to progress to the Advanced level.</td>
</tr>
<tr>
<td>Partly Achieved</td>
<td>partial understanding of the knowledge and skills required to function at this grade level.</td>
<td>The learner: is unlikely to succeed in the next grade without support.</td>
<td>requires specific intervention to address knowledge gaps, while also requiring additional and continued classroom support to progress to the required Achieved level.</td>
</tr>
<tr>
<td>Not Achieved</td>
<td>very limited understanding of the knowledge and skills required to function at this grade level.</td>
<td>The learner: unlikely to succeed in the next grade without significant support.</td>
<td>require specific intervention to address knowledge gaps, while also requiring additional teaching time and extensive and continued support within the classroom context, to progress to the required Achieved level.</td>
</tr>
</tbody>
</table>
Table 2: Grade 6 mathematics PLDs

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


The results in Figure 1 indicate that:

- 80% of learners in the school functioned at the *Not Achieved* (55%) and *Partly Achieved* (25%) levels. These are learners who are unlikely to succeed in the next grade without significant support or some support.

- 20% of learners functioned at *Achieved* and *Advanced* levels. These learners have a reasonable to high likelihood of success in the next grade.

- Overall, the school is underperforming and requires close monitoring and support from the district. Only five percent of learners functioned at the *Advanced* level.

### Table 3: School, district and provincial performance in mathematics

<table>
<thead>
<tr>
<th></th>
<th>% Not Achieved</th>
<th>% Partly Achieved</th>
<th>% Achieved</th>
<th>% Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>School</td>
<td>55</td>
<td>25</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>District</td>
<td>39</td>
<td>16</td>
<td>29</td>
<td>16</td>
</tr>
<tr>
<td>Province</td>
<td>45</td>
<td>35</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

The results in Table 3 show that the performance of the school, in comparison to the province and the district, is significantly low as indicated by the percentage of learners who functioned at the *Achieved* and *Advanced* levels. For the school, 15% of learners functioned at the *Achieved* level and 5% at the *Advanced* level. Corresponding percentages for the district, 29% and 16%, respectively, are significantly higher. Significantly higher percentages of learners functioned at the *Not Achieved* (55%) and *Partially Achieved* (25%) levels. The school is underperforming in relation to other schools in the district. It should be one of the priority schools for district improvement strategies.
The results in Table 4 show that there are gender-based disparities in performance in the school with more boys than girls functioning at the Not Achieved and Partly Achieved levels. In terms of the combined percentage of boys who functioned at the Achieved and Advanced levels (50%), the school boys’ performance exceeds that of boys in each of the district (45%) and the province (40%). However, the combined percentage of girls who functioned at the Achieved (40%) and Advanced (15%) levels in the school is lower than that of their district counterparts, 45% and 20%, respectively.

The results in Table 5 indicate that at the Achieved and Advanced levels mean score percentages were the highest in the sub-domain of Number & Operations and ranged between 45% and 50%, while the lowest performance was in Probability where mean score percentages ranged between 15% and 20%. Performance in the rest of the sub-domains fell between these extremes. Generally, the levels and pattern of performance indicate that there are serious mathematics content knowledge deficiencies in the school. As has been noted, underperformance is markedly serious in the critical sub-domains of Measurement, Data Handling and Probability.

The results have serious teaching and learning implications and require a multi-pronged intervention strategy which must include:
• The School Head and the Senior Management Teams (SMT) must lead a skills audit to identify areas of content knowledge deficiency among the teachers of mathematics and initiate targeted support;

• The SMT, supported by subject advisors, must conduct continuing professional training of teachers in effective methods of teaching mathematics in general and teaching specific topics;

• The school must introduce and maintain a sustained system of regular assessments for timeous identification of learner knowledge and skill gaps to be followed by appropriate interventions at individual learner level;

• The school must popularise the value of learning mathematics among parents, provide them with regular and meaningful feedback on the performance of their children so that parents can provide necessary motivation and support to children to learn mathematics.

Table 6: Mean score (%) by PL, content sub-domain and gender

<table>
<thead>
<tr>
<th></th>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boys</td>
<td>Girls</td>
<td>Boys</td>
<td>Girls</td>
</tr>
<tr>
<td>Number &amp; Operations</td>
<td>35</td>
<td>40</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>30</td>
<td>35</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Measurement</td>
<td>20</td>
<td>25</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>Data Handling</td>
<td>15</td>
<td>20</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Probability</td>
<td>10</td>
<td>15</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>

As indicated in Table 6, at the Achieved and Advanced levels:

• The highest mean scores (%) were in Numbers & Operations for both boys and girls.

• The lowest mean scores were in Data Handling and Probability, again for both boys and girls.

• Generally, more girls than boys function at the Not Achieved level in all the sub-domains.

In addition to the strategies that have been recommended for improving performance in
mathematics and all its sub-domains, the school needs to address the gender-related disparities in performance. Some of the interventions to consider include:

i. Engaging the services of female role models in mathematics from among subject advisors and teachers to address learners on the value of learning mathematics;

ii. Consciously removing male stereotypes from the teaching and learning of mathematics; and,

iii. Effective teaching of mathematics to all learners, boys and girls alike, to succeed in proportion to their capabilities to eliminate fear of the subject by learners.
First Additional Language

Table 3: Grade 6 FAL PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, read single words, short sentences &amp; pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL, read texts with simple sentences. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; interpret meaning of stories relating to real life experiences. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; understand complex texts in formal &amp; informal stories. She/he can:</td>
</tr>
<tr>
<td>- recognise words that were learnt previously</td>
<td>- skim &amp; scan some poetical elements;</td>
<td>- identify characters &amp; plot setting;</td>
<td>- read poetry effectively evaluate texts;</td>
</tr>
<tr>
<td>- read forward &amp; answer short questions based on the text; and,</td>
<td>- skim, scan &amp; summarise texts;</td>
<td>- identify ethical issues such as cultural/social diversity;</td>
<td>- identify formal and informal texts;</td>
</tr>
<tr>
<td>- identify only the key character in a text.</td>
<td>- use vocabulary of 1500 words;</td>
<td>- highlight the moral lesson behind a story;</td>
<td>- easily utilise vocabulary of at least 5000 words;</td>
</tr>
<tr>
<td></td>
<td>- write friendly letters;</td>
<td>- infer information from a complex text;</td>
<td>- critique texts through book reviews &amp; reports;</td>
</tr>
<tr>
<td></td>
<td>- extract information directly from a short text; and,</td>
<td>- identify key elements of poetry;</td>
<td>- analyse both formal &amp; informal documents;</td>
</tr>
<tr>
<td></td>
<td>- read short stanzas &amp; answer literal questions.</td>
<td>- use dictionaries in building vocabulary of at least 3000 words; and,</td>
<td>- synthesise information from different parts of a text;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- understand factual information from non-fictional documents.</td>
<td>- demonstrate comprehension of inferred information from a text.</td>
</tr>
<tr>
<td><strong>WRITING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, show understanding through drawings and pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL, write for social purposes. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write information in graphic &amp; descriptive forms. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write &amp; develop information from graphical &amp; descriptive forms to text. He/she can:</td>
</tr>
<tr>
<td>- write single words &amp; short sentences;</td>
<td>- write personal letters, diaries, news reports;</td>
<td>- write for social purposes, e.g. personal letters, diaries, dialogues &amp; simple news reports;</td>
<td>- write extensively for social purposes;</td>
</tr>
<tr>
<td>- represent ideas with drawings;</td>
<td>- create a book cover;</td>
<td>- design a book cover; and,</td>
<td>- develop news reports;</td>
</tr>
<tr>
<td>- spell few commonly used words correctly; and,</td>
<td>- identify similar &amp; different texts;</td>
<td>- develop &amp; edit key language structures.</td>
<td>- design questionnaires &amp; adverts;</td>
</tr>
<tr>
<td>- write brief diaries.</td>
<td>- express cause &amp; effect relationships; and,</td>
<td></td>
<td>- integrate ideas by classifying information;</td>
</tr>
<tr>
<td></td>
<td>- write simple sentences &amp; draw pictures.</td>
<td></td>
<td>- solve problems;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- use relevant questioning styles to obtain information;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- listen with effective comprehension.</td>
</tr>
</tbody>
</table>
The results in Figure 1 indicate that:

- 60% of learners functioned at the Not Achieved and Partly Achieved levels. These are learners who are unlikely to succeed in the next grade without significant support or some support.

- 40% of learners functioned at Achieved and Advanced levels and thus have a reasonable to high likelihood of success in the next grade.

Table 4: School, district and provincial performance in FAL

<table>
<thead>
<tr>
<th></th>
<th>% Not Achieved</th>
<th>% Partially Achieved</th>
<th>% Achieved</th>
<th>% Advance</th>
</tr>
</thead>
<tbody>
<tr>
<td>School</td>
<td>20</td>
<td>40</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>District</td>
<td>23</td>
<td>25</td>
<td>35</td>
<td>17</td>
</tr>
<tr>
<td>Province</td>
<td>42</td>
<td>27</td>
<td>21</td>
<td>10</td>
</tr>
</tbody>
</table>

The results in Table 4 show that while the school performance seems comparable to the province in terms of learners who functioned at the Advanced level, the school is underperforming at the other PLs in relation to other schools in the district. The school must be included in the list of schools targeted for support and improvement in the district.

Table 5: School performance in FAL by gender

<table>
<thead>
<tr>
<th></th>
<th>% Not Achieved</th>
<th>% Partially Achieved</th>
<th>% Achieved</th>
<th>% Advance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girls</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>45</td>
</tr>
<tr>
<td>District</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>Province</td>
<td>30</td>
<td>30</td>
<td>25</td>
<td>30</td>
</tr>
</tbody>
</table>
The results show that, in terms of the percentage of learners who function at the *Advanced* level, the school boys’ performance exceeds that of boys in the district but is the same as for boys in the province. However, the school also had a higher percentage of boys who functioned at the lowest level (*Not Achieved*) than the district. The percentage of school girls who functioned at the *Advanced* level is the lowest (10%) compared to the district (15%) and the province (30%). It is worth noting that a higher percentage of school girls (25%) functioned at the *Not Achieved* level than their district counterparts (15%). Overall, there are no significant differences between the performances of boys and girls in FAL in this school but general performance is not satisfactory with 45% of each of boys and girls functioning at the two lower PLs, *Not Achieved* and *Partly Achieved*. Implications are that the school requires closer monitoring and support from the district.

**Table 6: Mean score (%) by PL and content sub-domain**

<table>
<thead>
<tr>
<th></th>
<th>Not Achieved</th>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading &amp; Viewing</td>
<td>20</td>
<td>25</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>Language Structure</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Comprehension</td>
<td>5</td>
<td>10</td>
<td>25</td>
<td>35</td>
</tr>
</tbody>
</table>

The results in Table 6 indicate that at the *Achieved* and *Advanced* levels mean score percentages were the highest in Reading and Viewing and ranged between 55% and 60% whilst the lowest mean scores were for Comprehension ranging between 25% and 35%. It is of grave concern that in the more cognitively demanding language skills, Language Structure and Comprehension, mean score percentages at all the PLs, including the Advanced level, were below 50%. This observation has significant implications for the teaching and learning of language:

i. The School Head and the Senior Management Team (SMT) must prioritise the teaching of the FAL particularly to learners who are not speakers of the language outside the teaching and learning environment.

ii. Subject advisors must initiate appropriate training courses for teachers of languages, such as English, which are used for teaching and learning but are not indigenous languages.
iii. The school must adopt a policy of teaching language and comprehension skills across all subjects.

iv. The school must introduce and maintain a sustained programme of regular assessments in order to identify and track learning deficiencies in language and follow with individualised interventions for improvement.

Table 7: Mean score (%) by PL, content sub-domain and gender

<table>
<thead>
<tr>
<th></th>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boys</td>
<td>Girls</td>
<td>Boys</td>
<td>Girls</td>
</tr>
<tr>
<td>Reading &amp; Viewing</td>
<td>35</td>
<td>40</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Language Structure</td>
<td>20</td>
<td>25</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>Comprehension</td>
<td>10</td>
<td>15</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>

As indicated in Table 7, at the *Achieved* and *Advanced* levels:

- the highest mean scores (%) were in Reading & Viewing for both boys and girls; and,

- the lowest mean scores were in Comprehension, again for both boys and girls, although girls generally obtained higher mean score percentages in all the sub-domains and at all the PLs. In addition to the generic recommendations to teaching FAL, specific measures need to be considered to narrow gender-related gaps in performance in FAL:

  i. The school must initiate and sustain a culture of love for reading for all boys and girls with teachers leading as role models.

  ii. Regular assessment and comprehensive feedback will help communicate expected standards of performance to learners.

  iii. Eliminate gender-stereotypes to reading by making teachers model reading to learners of the opposite sex.
LEARNER REPORT

Learner name: - John Blues: School name: - City Primary School

Large-scale Assessment of Learners - 2013

Grade 6 Mathematics and First Additional Language

This report presents the results from the large-scale assessment survey (LSAS) of 2013 on learner levels and quality of performance in mathematics and First Additional Language (FAL) at the grade 6 level. Information is reported using performance standards. A “performance standard” describes performance in terms of the level at which a learner functions (Performance Level (PL)) shown in Table 1 and the set of knowledge and skills that the learner could or could not display (Performance Level Descriptors (PLDs) shown in Table 2 for mathematics and in Table 3 for FAL. The results must be read against Tables 1, 2 and 3.

Table 1: Definition and implications of performance standards

<table>
<thead>
<tr>
<th>Performance Level</th>
<th>Performance Level definition</th>
<th>Progression Implications</th>
<th>Intervention implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced</td>
<td>comprehensive understanding of the knowledge and skills required to function at this grade level</td>
<td>has high likelihood of success in the next grade</td>
<td>requires little or no academic intervention but needs to be provided with more challenging tasks to maximise their full potential</td>
</tr>
<tr>
<td>Achieved</td>
<td>sufficient understanding of the knowledge and skills required to function at this grade level</td>
<td>has a reasonable likelihood of success in the next grade</td>
<td>may require some assistance with complex concepts to progress to the Advanced level</td>
</tr>
<tr>
<td>Partly Achieved</td>
<td>partial understanding of the knowledge and skills required to function at this grade level</td>
<td>is unlikely to succeed in the next grade without support</td>
<td>requires specific intervention to address knowledge gaps, while also requiring additional and continued classroom support to progress to the required Achieved level</td>
</tr>
<tr>
<td>Not Achieved</td>
<td>very limited understanding of the knowledge and skills required to function at this grade level</td>
<td>unlikely to succeed in the next grade without significant support</td>
<td>require specific intervention to address knowledge gaps, while also requiring additional teaching time and extensive and continued support within the classroom context, to progress to the required Achieved level</td>
</tr>
</tbody>
</table>
### Table 2: Grade 6 mathematics PLs and PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>A learner at this level will, with intensive support, recognise basic number systems. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs:</td>
</tr>
<tr>
<td>- count forward only with whole numbers;</td>
<td>- count forward and backwards in decimals;</td>
<td>- critically read, interpret &amp; analyse with awareness of sources of error &amp; manipulation to draw conclusions &amp; make predictions;</td>
<td></td>
</tr>
<tr>
<td>- count objects not exceeding 10;</td>
<td>- recognise place value up to 9 digits;</td>
<td>- list possible outcomes for simple experiments including tossing a coin, rolling a die &amp; spinning a spinner;</td>
<td></td>
</tr>
<tr>
<td>- add whole numbers up to 10;</td>
<td>- round off number up to 1000;</td>
<td>- distinguish between volume, surface area &amp; dimensions of rectangular prisms;</td>
<td></td>
</tr>
<tr>
<td>- draw simple pictures of objects;</td>
<td>- add and subtract up to 9 digits;</td>
<td>- solve problems involving finances and measurement;</td>
<td></td>
</tr>
<tr>
<td>- Measure length of lines;</td>
<td>- do simple calculations using ordinary fractions and decimals;</td>
<td>- compare rate and ratio;</td>
<td></td>
</tr>
<tr>
<td>- Name few SI units;</td>
<td>- read digital and analogue time;</td>
<td>- identify &amp; describe numeric and geometric patterns;</td>
<td></td>
</tr>
<tr>
<td>- measure area &amp; perimeter of objects; and,</td>
<td>- measure using basic SI units; and,</td>
<td>- use and describe transformations;</td>
<td></td>
</tr>
<tr>
<td>- draw simple bar graphs.</td>
<td>- draw pictographs</td>
<td>- locate and describe movement on a grid;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- differentiate between sample &amp; population; and,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- draw &amp; interpreting bar and pictographs</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

222
Table 3: Grade 6 FAL PLs and PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, read single words, short sentences &amp; pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL, read texts with simple sentences. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; interpret meaning of stories relating to real life experiences. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; understand complex texts in formal &amp; informal stories. She/he can:</td>
</tr>
<tr>
<td>• recognise words that were learnt previously;</td>
<td>• skim &amp; scan some poetical elements;</td>
<td>• identify characters &amp; plot setting;</td>
<td>• read poetry effectively;</td>
</tr>
<tr>
<td>• read forward &amp; answer short questions based on the text; and,</td>
<td>• read forward &amp; answer short questions based on the text; and,</td>
<td>• identify ethical issues such as cultural/social diversity;</td>
<td>• evaluate texts;</td>
</tr>
<tr>
<td>• identify only the key character in a text.</td>
<td>• identify only the key character in a text.</td>
<td>• highlight the moral lesson behind a story;</td>
<td>• identify formal and informal texts;</td>
</tr>
<tr>
<td><strong>WRITING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, show understanding through drawings and pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL, write for social purposes. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write information in graphic &amp; descriptive forms. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write &amp; develop information from graphical &amp; descriptive forms to text. He/she can:</td>
</tr>
<tr>
<td>• write single words &amp; short sentences;</td>
<td>• write personal letters, diaries, news reports;</td>
<td>• write for social purposes; e.g. personal letters;</td>
<td>• write extensively for social purposes;</td>
</tr>
<tr>
<td>• represent ideas with drawings;</td>
<td>• create a book cover;</td>
<td>• diaries, dialogues &amp; simple news reports;</td>
<td>• develop news reports;</td>
</tr>
<tr>
<td>• spell few commonly used words correctly; and</td>
<td>• identify similar &amp; different texts;</td>
<td>• design questionnaires &amp; adverts;</td>
<td>• design questionnaires &amp; adverts;</td>
</tr>
<tr>
<td>• write brief diaries.</td>
<td>• express cause &amp; effect relationships;</td>
<td>• integrate ideas by classifying information;</td>
<td>• solve problems;</td>
</tr>
<tr>
<td>• write simple sentences &amp; draw pictures.</td>
<td>• write simple sentences &amp; draw pictures.</td>
<td>• use relevant questioning styles to obtain information; and</td>
<td>• use relevant questioning styles to obtain information; and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>listen with effective comprehension.</td>
<td>listen with effective comprehension.</td>
</tr>
</tbody>
</table>
LEARNER REPORT

City Primary School

Grade 6 National Assessment

Year: 2013

Learner: John Blues

<table>
<thead>
<tr>
<th>Sub-domains</th>
<th>Learner score (%)</th>
<th>Class score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading &amp; Viewing</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Language Structure</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td>Comprehension</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>

- John functions at the Achieved level and has a reasonable likelihood of success in the next grade
- He must spend at least 30 mins a day reading English fiction of various levels of complexity and summarise each piece to improve his comprehension skills

<table>
<thead>
<tr>
<th>Sub-domains</th>
<th>Learner score (%)</th>
<th>Class score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number &amp; Operations</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>Measurement</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>Data Handling</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Probability</td>
<td>10</td>
<td>35</td>
</tr>
</tbody>
</table>

Legend:

NA = Not Achieved; PA = Partly Achieved; A = Achieved; Adv = Advanced

General comments

For details of what John knows and can do in mathematics and First Additional Language, refer to the relevant tables that accompany this report.

Issued by the Department of Basic Education

November 2013

Signed
6.3 Reporting trends over time

An important question that LSAS need to provide answers to and, therefore, must be part of the NRF, is whether performance trends over time show improvement, remain unchanged or decline. Whilst historical information is important for evaluating the impact of interventions in education, making comparisons in learner performance from year to year requires appropriate planning and design of the assessment processes. The best design for producing test results that are comparable is to administer the same test from year to year. But for that to happen the test must be kept confidential to avoid possible artificial inflation of scores due to teachers teaching to the test. Test equating is a common technique that is used to link and compare the results of different tests that were administered on separate occasions. One commonly used design feature for test equating is to administer common items on the separate occasions. Since trend information is important for tracking changes in performance over time, test equating must be a mandatory element of the NRF.

Provided that the test design allows comparisons of performance over time, the following illustrations show how trends over time can be reported. Changes in performance could provide useful lessons on what interventions work well or do not.

Grade 6 trends in learner performance: mathematics

Changes in performance at national level since the previous ANA are presented in bar graphs in Figure 1. Changes in performance could provide useful lessons on what interventions work well or do not.

![Figure 18: National grade 6 trends in performance in mathematics](image-url)
From Figure 1, the percentage of learners at the *Not Achieved* level who are unlikely to proceed to the next grade, has decreased by five percent since the last assessment. This is a good sign of improvement at the lower end of the competency continuum. However, the percentage of learners who functioned at the *Advanced* level and are certain to proceed to the next grade remained unchanged at 10%. This could mean that all provinces focused interventions at schools that were considered to be underperforming and relaxed monitoring schools that were considered to be functioning at the *Advanced* level and, consequently, these schools may have relaxed. The implication is that interventions must address school need at all levels.

Trends over time in performance in mathematics at provincial level can be presented as illustrated in Figure 2.

![Figure 19: Provincial grade 6 trends in performance in mathematics](image)

**Figure 19: Provincial grade 6 trends in performance in mathematics**

From Figure 2, the percentage of learners at the *Not Achieved* level who are unlikely to proceed to the next grade, decreased by five percent since the last assessment. At the same time, the percentage of learners functioning at the *Partly Achieved* level increased by five percent from 45% to 50%. However, the percentage of learners who functioned at the *Advanced* level and are certain to proceed to the next grade remained unchanged at 25%. The implication is that all district interventions were focused to underperforming schools and
schools that were considered to be performing above expectations were left unattended. Schools need to be supported and motivated to improve continuously.

District level changes in performance since the previous ANA are presented in bar graphs in Figure 3.

![Bar graph showing district grade 6 trends in performance in mathematics]

**Figure 20: District grade 6 trends in performance in mathematics**

From Figure 3, the percentage of learners at the *Not Achieved* level who are unlikely to proceed to the next grade, increased by 10% since the last assessment. This is a sign of regression; the district is performing worse than in the previous assessment. At the same time, the percentage of learners functioning at the *Partly Achieved* level decreased by 12% from 28% to 16%. The percentage of learners who functioned at the *Achieved* level remained unchanged at 29%. However, the percentage of learners who functioned at the Advanced Level (PL4) and are certain to proceed to the next grade increased from 14% to 16%. Generally, well-performing schools have improved while underperforming schools have gone even further down. Improvement strategies must seek to reverse the situation at the lower end of performance.

Changes in performance since the previous assessment at school level are presented in bar graphs in Figure 4.
From Figure 4, the percentage of learners at the Not Achieved level who are unlikely to proceed to the next grade, increased by four percent since the last assessment. At the same time, the percentage of learners functioning at the Partly Achieved level and Advanced level remained unchanged at 15% each level. However, the percentage of learners who functioned at the Achieved level and are highly to progress to the next grade decreased by four percent from 34% to 30%. Generally, there has been a decline in performance. Drastic measures to improve are required at all levels, at the level of management, at professional services by teachers and commitment by learners.

National grade 6 trends in learner performance: English FAL

Changes in English FAL performance at national level since the previous assessment are presented in bar graphs in Figure 5.
From Figure 5, the percentage of learners at the *Not Achieved* level who are unlikely to proceed to the next grade, has decreased by five percent since the last assessment. This is a good and positive sign of improvement. At the same time, the percentage of learners functioning at the *Partly Achieved* level increased by 10% from 15% to 25%, again a good sign of improvement. Similarly, the percentage of learners who function at the *Achieved* level increased from 10% to 25%. However, the percentage of learners who functioned at the *Advanced* level and are highly to progress to the next grade decreased by 10% from 25% to 15%. So, while there has been improvement at the lower end, there has been a decline at the top end of the performance continuum. There is need to focus on the whole system with appropriate interventions and not in specific sectors.

**Provincial grade 6 trends in learner performance: FAL**

At provincial level changes in English FAL performance since the previous assessment are presented in bar graphs in Figure 6.
From Figure 6, the percentage of learners at the *Not Achieved* who are unlikely to proceed to the next grade, increased by five percent since the last assessment. At the same time, the percentage of learners functioning at the *Partly Achieved* level increased by 12% from 15% to 27%. Similarly, the percentage of learners who functioned at the *Achieved* level increased from 20% to 21%. However, the percentage of learners who functioned at the *Advanced* level and are highly to progress to the next grade decreased by 15% from 25% to 10%. Generally, there has been a decline in performance at all levels. Significantly drastic measures are required to improve.

At district level changes in performance since the previous assessment can also be presented as in the bar graphs in Figure 7.
From Figure 7, the percentage of learners at the *Not Achieved* level who are unlikely to proceed to the next grade, has increased by 10% since the last assessment. This is an indication of a decline in performance and will call for definite measures to reverse the situation. At the same time, the percentage of learners functioning at the *Partly Achieved* level increased by 10% from 30% to 40%, again an indication that performance has taken a downward spiral. The percentage of learners who functioned at the *Achieved* level also increased by five percent from 20% to 25% which was an appreciable improvement at this level. However, the percentage of learners who function at the *Advanced* level (PL3) and are certain to progress to the next grade decreased by 10% from 20% to 15%. These trends indicate that there was no district-wide focused monitoring of performance since the last assessment. Where some improvement was realised, this could have been driven by individuals in schools that are generally functional. Definite measures will need to be taken to reverse the situation.

### 6.3 Decision-making and interventions

Following each assessment, the national department must conduct orientation sessions on how to read, interpret and utilise the information in the standards-based reports. This is a
critical short-term intervention until a culture of using data has taken root in the system. Upon receiving the report, leaders and all role-players at each level need to interact with the report, make informed decisions and plan appropriate interventions that apply to the roles and responsibilities of the level. For instance, the national Department may need to review policies that have a direct bearing on teaching and learning like revising norms and standards for the provision of subject advisors for special subjects such as mathematics and English. A province may need to provide special attention in terms of capacity building to a district which, according to the report, is performing particularly poorly in one or both subjects that were assessed.

An illustration of possible decisions and interventions that can be made at district level is given to clarify the essential aspects of the NRF. Districts provide critical support to schools. At district level there are curriculum specialists and subject advisors whose responsibility is to provide continuing professional support and development to teachers. This set of professionals should be the first to receive and interact with the report from LSAS. In the absence of inadequacy of subject advisors, the district management should facilitate support groups of teachers whereby senior or lead teachers provide the necessary support.

The leader of the professional support team or an official charged with this task at district level must convene a working session where the results are presented. The leader could summarise the key findings from the assessment, preferably on a PowerPoint presentation to help focus attention and discussion. Copies of the performance standards which must accompany every report must be made available in this session. The purpose of the work session is to:

- Discuss identified weaknesses and strengths from the assessment;
- Identify priority areas for intervention; and,
- Draw up an action plan on how improvements shall be implemented.

An example of a summary that could be presented is shown below. The example presentation is for a district but similar presentations can be prepared and presented at national, provincial and school levels.

In the given district example presentation, the pie chart gives an overview of performance in the district in terms of percentages of learners who achieved various PL in the assessments. It
will be important to discuss what each PL represents in terms of the knowledge and skills displayed by learners. The bar graphs provide historical information that gives context to ensuing discussions. The table that presents results by gender is important for identifying possible elements of exclusion or biases in the delivery of education. The table that summarises performance in terms of sub-domains is critical for targeted planning according to identified areas of challenge. Bar graphs are also used to depict overall baseline results and targets as well as baseline and targets across school quintiles.

5.9 PowerPoint presentation of summary district results

Purpose of presentation

The purpose of the PowerPoint presentation is to:

- Present the results of the latest LSAS;
- Identify areas of strength and weakness in learner performance; and,
- Discuss and plan appropriate interventions following the information that emanates from the assessment.

Aim

To improve the teaching and learning of mathematics and language in all schools in the district so that at least 60% of learners achieve acceptable levels (PL3 and PL4) in LSAS.

Situation analysis

What has changed since the last assessment in terms of:

- the profile of the learner population;
- teacher supply;
- teacher content knowledge;
- teaching and assessment methods; and,
- availability and utilisation of appropriate teaching and learning materials?
Interpretation of performance standards

Presentation of results must always be preceded by the presentation and interpretation of performance levels and performance level descriptors. Tables 19, 20 and 21 form a key element of the NRF and must be attached to every report.

Table 20: Policy definitions and their practical significance per Performance Level

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

8 For details of what learners at each PL can or cannot do refer to supplied Table of Performance Standards
### Table 20: Grade 6 mathematics PLs and PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>A learner at this level will, with intensive support, recognise basic number systems. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs:</td>
</tr>
<tr>
<td>• count forward only with whole numbers;</td>
<td>• count forward and backwards in decimals;</td>
<td>• count, recognise and do calculations using fractions &amp; decimals;</td>
<td>• critically read, interpret &amp; analyse with awareness of sources of error &amp; manipulation to draw conclusions &amp; make predictions;</td>
</tr>
<tr>
<td>• count objects not exceeding 10;</td>
<td>• recognise place value up to 9 digits;</td>
<td>• represent multiples, factors &amp; prime numbers;</td>
<td>• list possible outcomes for simple experiments including tossing a coin, rolling a die &amp; spinning a spinner;</td>
</tr>
<tr>
<td>• add whole numbers up to 10;</td>
<td>• round off number up to 1000;</td>
<td>• find percentages of whole numbers;</td>
<td>• distinguish between volume, surface area &amp; dimensions of rectangular prisms;</td>
</tr>
<tr>
<td>• draw simple pictures of objects;</td>
<td>• add and subtract up to 9 digits;</td>
<td>• solve problems involving finances and measurement</td>
<td>• solve problems involving different time zones;</td>
</tr>
<tr>
<td>• Measure length of lines;</td>
<td>• do simple calculations using ordinary fractions and decimals;</td>
<td>• compare rate and ratio;</td>
<td>• estimate, record, compare &amp; convert between SI units (including mass, temperature, distance and capacity);</td>
</tr>
<tr>
<td>• Name few SI units;</td>
<td>• read digital and analogue time;</td>
<td>• identify &amp; describe numeric and geometric patterns;</td>
<td>• organise &amp; record data;</td>
</tr>
<tr>
<td>• measure area &amp; perimeter of objects; and,</td>
<td>• measure using basic SI units; and,</td>
<td>• use and describe transformations;</td>
<td>• calculate the median &amp; mode of data; and,</td>
</tr>
<tr>
<td>• draw simple bar graphs.</td>
<td>• draw pictographs.</td>
<td>• locate and describe movement on a grid;</td>
<td>• list possible outcomes &amp; predict “likelihood” of events.</td>
</tr>
</tbody>
</table>
Table 21: Grade 6 FAL PLs and PLDs

<table>
<thead>
<tr>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, read single words, short sentences &amp; pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL, read texts with simple sentences. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; interpret meaning of stories relating to real life experiences. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, read &amp; understand complex texts in formal &amp; informal stories. She/he can:</td>
</tr>
<tr>
<td>- recognise words that were learnt previously;</td>
<td>- skim &amp; scan some poetical elements;</td>
<td>- identify characters &amp; plot setting;</td>
<td>- read poetry effectively;</td>
</tr>
<tr>
<td>- read forward &amp; answer short questions based on the text; and,</td>
<td>- use vocabulary of 1500 words;</td>
<td>- identify ethical issues such as cultural/social diversity;</td>
<td>- evaluate texts;</td>
</tr>
<tr>
<td>- identify only the key character in a text.</td>
<td>- write friendly letters;</td>
<td>- highlight the moral lesson behind a story;</td>
<td>- identify formal and informal texts;</td>
</tr>
<tr>
<td></td>
<td>- extract information directly from a short text; and,</td>
<td>- infer information from a complex text;</td>
<td>- easily utilise vocabulary of at least 5000 words;</td>
</tr>
<tr>
<td></td>
<td>- read short stanzas &amp; answer literal questions.</td>
<td>- identify key elements of poetry;</td>
<td>- critique texts through book reviews &amp; reports;</td>
</tr>
<tr>
<td><strong>WRITING</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A learner at this level will, with intensive support, show understanding through drawings and pictures. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PL, write for social purposes. He/she can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write information in graphic &amp; descriptive forms. She/he can:</td>
<td>A learner at this level can, in addition to skills &amp; knowledge in the lower PLs, write &amp; develop information from graphical &amp; descriptive forms to text. He/she can:</td>
</tr>
<tr>
<td>- write single words &amp; short sentences;</td>
<td>- write personal letters, diaries, news reports;</td>
<td>- write for social purposes, e.g. personal letters, diaries, dialogues &amp; simple news reports;</td>
<td>- write extensively for social purposes;</td>
</tr>
<tr>
<td>- represent ideas with drawings;</td>
<td>- create a book cover;</td>
<td>- design a book cover; and,</td>
<td>- develop news reports;</td>
</tr>
<tr>
<td>- spell few commonly used words correctly; and,</td>
<td>- identify similar &amp; different texts;</td>
<td>- develop &amp; edit key language structures.</td>
<td>- design questionnaires &amp; adverts;</td>
</tr>
<tr>
<td>- write brief diaries.</td>
<td>- express cause &amp; effect relationships; and,</td>
<td></td>
<td>- integrate ideas by classifying information;</td>
</tr>
<tr>
<td></td>
<td>- write simple sentences &amp; draw pictures.</td>
<td></td>
<td>- solve problems;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- use relevant questioning styles to obtain information; and,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- listen with effective comprehension.</td>
</tr>
</tbody>
</table>
Overall performance of the District is shown in Figure 1. The following observations can be made:

**Mathematics**

- More than half of the grade 6 learners (55%), 39% at the *Not Achieved* level and 16% at the *Partly Achieved* level, are unlikely to succeed in the next grade without either some or significant support in mathematics. This category of learners will require at least specific intervention to address the identified knowledge gaps, while also requiring additional and continued classroom support.

- The proportion of learners who have a reasonable to high likelihood of success in the next grade is 45% which includes 29% who are at the *Achieved* level and 16% at the *Advanced* level.
The following key observations can be made from Figure 2:

- Overall, there has been a drop in grade 6 mathematics performance in the district in the period under review.

- The proportion of grade 6 learners performing at the *Not Achieved* level has increased from 29% in 2012 to 39% in 2013. This means that there are 10% more learners who require at least specific intervention to address the identified knowledge gaps, while also requiring additional and continued classroom support.

- Performance at the *Achieved* and *Advanced* levels has almost remained unchanged in this period except for a slight upward movement at the *Advanced* level. There is need to maintain and increase performance at these two levels so that overall district performance can also increase.

- On the whole, the district needs to attend to the apparent equity issues that are signalled by the disparate changes in performance levels during the period under review. Performance in well-performing schools remained either unchanged or at least increased a little whereas in the underperforming schools the situation seems to have gone worse. There is need to investigate whether these disparities are not linked to equitable distribution of human and other resources and, if so, take steps to redress the situation.
Table 1: District grade 6 performance in mathematics (mean score %)

<table>
<thead>
<tr>
<th>Sub-domains</th>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numbers &amp; Relationships</td>
<td>25</td>
<td>30</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>20</td>
<td>25</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>Measurement</td>
<td>15</td>
<td>20</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>Data Handling</td>
<td>10</td>
<td>15</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>Probability</td>
<td>5</td>
<td>10</td>
<td>20</td>
<td>40</td>
</tr>
</tbody>
</table>

The following key observations can be made from Table 1:

- The highest mean score percentages, ranging between 40% and 55%, were in two sub-domains, ‘Numbers and Relationships’ and ‘Functions and Algebra’, at the *Achieved* and *Advanced* levels. Even though this was the highest performance, it was too low for a district with schools that span all five quintiles.

- In the rest of the sub-domains mean score percentages ranged from 20% in Probability at the *Achieved* level to 40% in Measurement at the *Advanced* level. This is a matter of concern because these competencies are critical to learning mathematics beyond grade 6. In these three sub-domains there is need to address the identified knowledge gaps, while also requiring additional classroom support to teachers by SMTs and district officials.

- The district needs to mobilise a monitoring campaign and conduct a skills audit to identify teachers who require training in specific areas of mathematics.
Table 2: District grade 6 performance in mathematics by quintile (mean score %)

<table>
<thead>
<tr>
<th>Sub-domains</th>
<th>Not Achieved</th>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quintile 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number &amp; Operations</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Measurement</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>Data Handling</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>Probability</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td><strong>Quintile 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number &amp; Operations</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Measurement</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Data Handling</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td>Probability</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td><strong>Quintile 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number &amp; Operations</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Measurement</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Data Handling</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td>Probability</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
</tr>
<tr>
<td><strong>Quintile 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number &amp; Operations</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Measurement</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Data Handling</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Probability</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td><strong>Quintile 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number &amp; Operations</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
</tr>
<tr>
<td>Functions &amp; Algebra</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>Measurement</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Data Handling</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Probability</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
</tr>
</tbody>
</table>

From Table 2 the following key observations can be made:

- District performance in terms of mean score percentages was the highest in the sub-domains of ‘Numbers and Operations’ and ‘Functions and Algebra’ in all quintiles. It would appear that these two sub-domains are taught well and understood better by learners in all the schools although in Quintile 1 schools there is room for improvement.

- Performance is the lowest in the sub-domains of ‘Data Handling’ and ‘Probability’ and this applies to all quintiles although the situation is the worst in Quintile 1 schools.
The district requires a multi-pronged strategy of dealing with underperformance in mathematics:

i. There is need to address content knowledge deficiencies among teachers across the system regardless of the quintile level of schools where they teach. Content areas such as ‘Data Handling’ and ‘Probability’ require training of teachers in both content knowledge as well as in the methods of teaching these topics. Provision of subject advisors of a high calibre is critical here;

ii. There is need to redress inequities in the distribution of human and material resources in schools in the lowest three quintiles. Provision of appropriate learning materials and assignment of properly qualified teachers and subject advisors to these schools are critical; and,

iii. Schools must be assisted to establish assessment regimes that ensure there is feedback that is meaningful to model expected performance standards to learners and parents.
**First Additional Language**

Key observations from Figure 1:

- The proportion of learners who are unlikely to succeed in the next grade without either significant support, *Not Achieved* level (23%), or some support, *Partly Achieved* level (25%), is 48%. These are learners who require specific intervention to address identified knowledge gaps, while also requiring additional teaching time and extensive and continued support within the classroom context;

- 52% of learners, 35% at the *Achieved* level and 17% at the *Advanced* level, have reasonable to high likelihood of succeeding in the next grade.

- Overall just over half of the learners in the district function at acceptable levels, viz. *Achieved* and *Advanced* levels. This has heavy implications for improvement:
  
  i. The majority of the underperforming learners are in schools where they are taught in a LOLT that they do not speak outside the teaching and learning environments. There is a need for either appropriately qualified teachers or continuing professional training of the existing personnel on how to teach a language that is not a mother tongue but is used for teaching and learning.

  ii. The majority of schools are under-resourced in terms of learning and teaching materials. There is need to supply materials but also to equip teachers to optimise the use of available materials.
Figure 3: District performance trends in grade 6 FAL

From Figure 3 key observations are:

- Overall, there has been a drop in grade 6 FAL performance in the district in the period under review.

- The proportion of grade 6 learners who perform at the Not Achieved level has increased from 29% in 2012 to 39% in 2013. This means that there are 10% more learners who require at least specific intervention to address the identified knowledge gaps, while also requiring additional and continued classroom support.

- There has been a very small improvement of two percent at the Advanced level.

- The district needs to address issues of equity as manifested in the disparate performance of learners in disparate school environments.

Table 3: District performance in FAL by quintile and sub-domain (mean score %)

<table>
<thead>
<tr>
<th>Sub-domains</th>
<th>Not Achieved</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading &amp; Viewing</td>
<td>30</td>
<td>45</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>Language Structure</td>
<td>35</td>
<td>42</td>
<td>48</td>
<td>55</td>
</tr>
<tr>
<td>Comprehension</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
</tr>
</tbody>
</table>

From Table 3 key observations are:
• Mean score percentages are the highest in the ‘Reading and Viewing’ sub-domain and increase from the lowest to the highest PL, from the Not Achieved to Advanced level.
• The lowest means scores are in the sub-domain of ‘Comprehension’.
• These patterns of performance have serious implications because ‘Comprehension’ skills are required across the curriculum.

**Table 4: District grade 6 performance in FAL by quintile (mean score %)**

<table>
<thead>
<tr>
<th>Sub-domains</th>
<th>Not Achieved</th>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quintile 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading &amp; Viewing</td>
<td>10</td>
<td>15</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Language Structure</td>
<td>15</td>
<td>20</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Comprehension</td>
<td>20</td>
<td>25</td>
<td>35</td>
<td>40</td>
</tr>
<tr>
<td><strong>Quintile 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading &amp; Viewing</td>
<td>15</td>
<td>20</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Language Structure</td>
<td>20</td>
<td>25</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td>Comprehension</td>
<td>25</td>
<td>30</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td><strong>Quintile 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading &amp; Viewing</td>
<td>20</td>
<td>25</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>Language Structure</td>
<td>25</td>
<td>30</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Comprehension</td>
<td>30</td>
<td>35</td>
<td>45</td>
<td>50</td>
</tr>
<tr>
<td><strong>Quintile 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading &amp; Viewing</td>
<td>25</td>
<td>30</td>
<td>60</td>
<td>65</td>
</tr>
<tr>
<td>Language Structure</td>
<td>30</td>
<td>35</td>
<td>55</td>
<td>60</td>
</tr>
<tr>
<td>Comprehension</td>
<td>35</td>
<td>36</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td><strong>Quintile 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading &amp; Viewing</td>
<td>30</td>
<td>34</td>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td>Language Structure</td>
<td>33</td>
<td>35</td>
<td>60</td>
<td>65</td>
</tr>
<tr>
<td>Comprehension</td>
<td>36</td>
<td>38</td>
<td>55</td>
<td>60</td>
</tr>
</tbody>
</table>

Key observations from Table 4 indicate that:
• Although performance in Quintile 1 and Quintile 2 schools was very low, as low as 10% at the Not Achieved level in Quintile 1, learners in these two quintiles managed to obtain mean scores up to 55% at the Advanced level. This points to potential that the district needs to nurture and develop in the disadvantaged schools. It also shows that given opportunity, all learners have the capability to succeed.
Performance in ‘Reading and Viewing’ as shown in mean score percentages is the highest, followed by ‘Language Structure’ and performance in ‘Comprehension’ is the lowest in all the quintiles.

The district needs to address the problem of relatively low performance in the teaching and learning of the critical competencies of ‘Language Structure’ and ‘Comprehension’ in all schools but also focus on dealing with equity and redress issues as the disparate performances across quintiles point out.

A multi-pronged strategy is required so that all schools are assisted to improve performance but the disadvantaged schools are particularly assisted to overcome the impact of poverty and disadvantage on effective teaching and learning.

**Target setting**

What change does the district want to see in the next round of assessment? Targets must include overall performance, issues of equity (school quintiles), gender, sub-domains, etc. Targets need to be negotiated with relevant role-players unless national or provincial guidelines have already been given. For instance, the Presidential injunction in 2010 put improvement targets at a minimum of 60% of grade 3 and 6 learners to achieve acceptable levels of performance in literacy and numeracy by 2014. Using that injunction to illustrate target setting, a district could set targets as shown in Figure 4.
**Figure 4: District grade 6 mathematics improvement targets: 2013-2018**

From Figure 4, the baseline performance in grade 6 mathematics in 2013 was 39% of learners functioning *Not Achieved* level, 16% at *Partly Achieved* level, 29% at the *Achieved* and 16% at the *Advanced* level. Improvements in this scenario must involve reducing the percentage of learners who function at the lowest ends of the performance continuum and increasing the percentage of those who function at the top end. In the illustration, the suggestion is to reduce the percentage of learners at the lowest end from 39% to 5% and also reduce the percentage of learners at the *Partially Achieved* level from 16% to 15% while simultaneously increasing the percentage of learners at the *Achieved* level from 29% to 40% and from 16% to 20% at the *Advanced* level.

Similar targets could be set for grade 6 FAL as illustrated in Figure 5.

**Figure 5: District grade 6 FAL improvement targets: 2013-2018**

From Figure 5, baseline grade 6 FAL performance is 23% learners at the *Not Achieved* level, 25% at the *Partly Achieved* level, 35% at the *Achieved* level and 17% at the *Advanced* level. Improvement targets could be set at reducing the percentage of learners at the *Not Achieved* level to 5% and those at the *Partly Achieved* level to 15% while increasing the percentages at the *Achieved* and *Advanced* levels to 40% and 20%, respectively.
Addressing equity problems

Improvement of performance must take into cognisance and address issues pertaining to equity. In particular, improvement targets must be directed to schools and learners across all the poverty quintiles. Examples of targets that address equity issues in mathematics and FAL are presented in Illustrative Figures 4 & 5, respectively. In each case, the baseline percentages are indicated with a “B” following the quintile number and a “T” for the target percentage. For example, “Q1B” refers to “Quintile 1 baseline percentage” while “Q1T” refers to “Quintile 1 target percentage”.

From each illustrative figure, the target is to make sure that at least 60% of the learners in each quintile achieve Achieved and Advanced levels, 40% and 20%, respectively, as guided by the Presidential injunction.

Figure 6: District grade 6 mathematics targets by quintile

From Figure 6, the mathematics target in Quintile 1 schools is mainly to reduce the baseline percentage at the Not Achieved level from Q1B=60% to Q1T=15% and increase the corresponding baseline percentages at the Achieved and Advanced levels to Q1T=40% and
QIT=20%, respectively. In Quintile 5, schools which have relatively lower percentages of learners at the lower levels (Not Achieved and Partly Achieved levels), the main target is to increase mathematics baseline percentages at the Achieved and Advanced levels to Q5T=35% and Q5T=40%, respectively.

An example of setting targets in language across the school quintiles is given in Figure 7.

![Figure 7: District grade 6 FAL targets by school quintile](image)

From Figure 7, schools in Quintiles 1-3 are characterised by majority of learners at the lower achievement levels, Not Achieved and Partly Achieved levels, while schools in Quintiles 4 and 5 have already met the target of at least 60% of learners functioning at Achieved and Advanced levels. The targets in Quintiles 1-3 schools are to reduce the percentages of learners who function at the Not Achieved and Partly Achieved levels while in Quintiles 4 and 5 schools the targets are to increase the percentages of learners who function at the Achieved and Advanced levels.
Plan of Action

Any dissemination and presentation of LSAS results must culminate in a definite Plan of Action at each of the national, provincial, district and school levels. The Plan of Action must seek to find clear and specific answers to the question: What must be done by whom, how, etc. to achieve targets in the short term (One year), medium term (Two years) and long term (Three years and beyond)?

The participants at each level may decide on various courses of action depending on what the results show. The NRF, in terms of the performance standards, provides guidance on where action should begin and be focused. From the district results in the example presentation, for instance, the following general observations would indicate where district action should begin and focus:

- The majority of learners in this district require direct teaching of basic mathematical skills and knowledge (Not Achieved and Partly Achieved levels).
- The fact that trends show little to no improvement in the percentage of learners who function at the Achieved and Advanced levels suggests that learners may not be exposed to a variety of exercises to extend their capabilities and independent investigations that require high-order mathematical and language skills.

Sample decision-making template

It will be important that at each level interventions are considered according to priority. Easy-to-use templates can be developed for recording priorities. The example of priorities below will be appropriate in the first three-year cycle following the adoption of the NRF. It could be modified according to need. What are the TOP 3 “things-to-do-differently” arising from the current LSAS results?
<table>
<thead>
<tr>
<th>No.</th>
<th><strong>Short-term (1 yr.)</strong></th>
<th><strong>Mid-term (2 yrs.)</strong></th>
<th><strong>Long-term (3 or more yrs.)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Orientation</strong></td>
<td><strong>Capacity Building</strong></td>
<td><strong>Institutionalisation</strong></td>
</tr>
<tr>
<td>1</td>
<td>Train teachers on how to use standards-based data to inform their teaching and enhance learning.</td>
<td>Offer short weekend courses to lead teachers at Circuit Level on how performance standards are developed and how to use standards-based data to improve learning.</td>
<td>Collaborate with local universities to offer specialised courses on assessment, standard setting and effective utilisation of assessment data.</td>
</tr>
<tr>
<td>2</td>
<td>Provide guidelines on how to use available Workbooks to teach basic skills and knowledge identified to be deficient in the standard-based reports.</td>
<td>Train teachers on how to develop tests that are aligned to national performance standards, administer the tests in their classes and utilise the results to improve learning.</td>
<td>Conduct district-level standards-based common tests that are set by teams of teachers and are used to monitor performance in the district.</td>
</tr>
<tr>
<td>3</td>
<td>Provide teachers with relevant tools (e.g. simple programmed Excel spreadsheets) for recording, analysing and reporting school-based tests. Provide appropriate support.</td>
<td>Short courses for teachers on basic analysis skills; training on how to read technical assessment reports.</td>
<td>Ensure each school has a core team of staff members who are competent in basic skills on how to analyse, report and utilise data from standards-based assessments.</td>
</tr>
</tbody>
</table>
What requisite skills/knowledge shall we need to achieve our priorities?

### Table 6: Requisite skills/knowledge

<table>
<thead>
<tr>
<th>No.</th>
<th>Short-term (1 yr.)</th>
<th>Mid-term (2 yrs.)</th>
<th>Long-term (3 or more yrs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Elementary theoretical training on assessment, its purpose and uses</td>
<td>Moderate to advanced knowledge of assessment types; ability to plan and implement class lessons on the basis of evidence collected from assessment</td>
<td>Data-driven decision-making and instructional leadership that promotes data use in all schools in the district</td>
</tr>
</tbody>
</table>

What are our targets in terms of learner performance?

### Table 7: Targets

<table>
<thead>
<tr>
<th>No.</th>
<th>Short-term (1 yr.)</th>
<th>Mid-term (2 yrs.)</th>
<th>Long-term (3 or more yrs.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Quintile 1-3 schools</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reduce the percentage of grade 6 learners who function at the <em>Not Achieved</em> level in mathematics to 20% or fewer</td>
<td>Reduce the percentage of grade 6 learners who function at <em>Not Achieved</em> level in mathematics to 15% or fewer</td>
<td>Increase the percentage of grade 6 learners who function at <em>Achieved</em> and <em>Advanced</em> levels in mathematics to at least 60%</td>
</tr>
<tr>
<td></td>
<td>Reduce the percentage of grade 6 learners who function at <em>Not Achieved</em> level in FAL to 15% or fewer</td>
<td>Reduce the percentage of grade 6 learners who function at <em>Not Achieved</em> level in FAL to 5% or fewer</td>
<td>Increase the percentage of grade 6 learners who function at <em>Achieved</em> and <em>Advanced</em> levels in FAL to at least 60%</td>
</tr>
<tr>
<td></td>
<td><strong>Quintile 4&amp;5 schools</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>At least 55% of primary school learners in the district function at <em>Achieved</em> and <em>Advanced</em> levels in mathematics in national assessments, including ANA</td>
<td>At least 60% of primary school learners in the district display <em>Achieved</em> and <em>Advanced</em> level knowledge and skills in mathematics in national assessments, including ANA</td>
<td>At least 70% of primary school learners in the district display <em>Achieved</em> and <em>Advanced</em> level knowledge and skills in mathematics in national assessments, including ANA</td>
</tr>
</tbody>
</table>
7. CONCLUSION AND RECOMMENDATIONS

7.1 Introduction

This study set out to develop a National Reporting Framework (NRF) that could be used in recording, reporting, disseminating and using data from large-scale assessment survey (LSAS) in South Africa. The study was prompted by observations of continued under-performance in basic education in spite of increasing attempts by the sector to assess, monitor performance and provide necessary feedback with the intention of improving learner performance.

Some of the initiatives that South Africa has undertaken to improve performance in the sector overall and achievement in schools in particular, have been acknowledged in this study as follows:

a) Development of a national Action Plan that specifies targets, use indicators and milestones that must be achieved in specified timeframes. By fostering a culture of evidence-based planning, the Action Plan intends to improve overall efficiency in the sector and raise levels of achievement in schools.

b) Providing schools with workbooks that are designed to promote efficiency in teaching and learning. There is a potential for positive impact in this initiative provided the workbooks are utilised regularly and necessary assessment is carried out to monitor progress.

c) Introduction of Annual National Assessment (ANA) and using the results of the assessment to measure progress and provide meaningful feedback to schools to inform interventions for improved teaching and learning.

The study also recognised that assessment and the effective use of results from these assessments did not seem to have the desired impact on the quality of teaching and learning and the levels of achievement in schools. In particular, cognisance was taken of continued learner underperformance in the areas of mathematics and language. Some of the observations that the study made in relation to assessment and the use of assessment data indicated that:
a) Assessment reports, specifically ANA reports, communicated feedback in statistical formats that are embedded in the norm-referenced paradigm of ranking schools and learners, e.g. results are reported in raw scores that do not communicate anything on what learners know or can do.

b) Guidelines on how to use assessment information were limited in terms of addressing issues of capacity for data use, curriculum relevance and how to establish a culture of data use in provinces, districts and schools.

7.2 A theory of action on data use

Drawing from literature and international practice on how assessment has been successfully utilised to improve the quality of educational outcomes, this study posited a theory of action that holds that, providing a standards-based NFR within a criterion-referenced paradigm, presenting assessment results in formats that are user-friendly, meaningful and information-rich accompanied with relevant capacity building in a sustained culture of data use, will lead to effective data use at all levels and enhanced evidence-based decision-making regarding the implementation of relevant interventions for improving learning, which will result in improved performance in schools.

Building on this theory of action, the study set out to develop an NRF that will provide guidelines on how assessment data should be analysed, recorded, reported and utilised effectively to impact on achievement. The key tenets underpinning the NRF are:

a) a shift into a criterion-referenced paradigm in assessment;

b) using performance standards to report results in meaningful ways to improve practice; and

c) building capacity for data use at all levels of the system and particularly in the teaching and learning space.

7.3 Summary of the study design and methodology

The design and methodology of the study followed an exploratory approach whereby practices, technical requirements and capacity issues that must be part of the NRF to enhance
use of assessment data were identified. The study followed chronological steps that culminated in findings that became part of the NRF.

The first step involved comparing three standard setting methods to identify the most appropriate method of standard-setting (SS) for the South African context. The SS methods were selected for comparison, viz. the Angoff, the Objective Standard Setting (OSS) method and the Objective Method (OBM). While the Angoff method has been used and modified over more than fifty years, the other two methods are fairly new and have made promising claims of efficacy elsewhere (Stone, 2001; Shulruf et al, 2014). To contextualise all the three SS methods, a validation process which involved data simulations was used to check if the new methods yielded similar or different cutscores under different contexts represented by real and simulated datasets.

The second step in the study involved identifying conditions that were found to be most conducive to data use in empirical studies on data use for improving teaching and learning. The purpose of this exercise was to inform possible decisions on guidelines for data use that could be included as part of the NRF.

In the third step a survey of Department of Basic Education (DBE) and provincial education officials in assessment, curriculum and teacher development was conducted on their views regarding the use of either the Curriculum and Assessment Policy Statement (CAPS)-based or standards-based reporting templates for ANA. The CAPS-based template uses raw scores to report performance whilst the standards-based template uses Performance Levels (PLs) and Performance Level Descriptors (PLDs).

7.4 Summary of key research findings

The key research findings of this study are summarised in this section.

7.4.1 Identifying appropriate SS method

A significant finding in this study was that, in developing an NRF that is based on performance standards, careful selection of standard setting methods that enhance the quality of the information content of performance standards is critical. Standard setting methods that generate valid, content-relevant and curriculum-specific performance standards enhance the
meaningfulness and usefulness of reports on what learners know and can do in the subjects in which they are assessed.

The OSS method was found to satisfy the requirements of meaningful, relevant and user-friendly performance standards. The Angoff method was found to be limited in the extent to which it was able to yield meaningful performance standards while the OBM generated standards that were devoid of any curriculum content but the method produced cut scores of very high precision. Inconclusive previous research findings on the best method of standard setting led Morgan & Perie (2004) to conclude that, rather than looking for the best SS method, focus should rather be on looking for the best use of different methods. From this study, and in the context of education in South Africa, the OSS method holds good prospects of the “best use method”.

The significance of this finding lies in the realisation that introducing a criterion-referenced and standards-based NRF in a system that has been steeped in a norm-referenced paradigm could pose major paradigm shifts. One such shift will be in acceptance by officials in education and the public at large to move away from reporting performance in league tables. Instead, users will have to be informed that, data use that can transform schools and improve performance must be provided in ways that enable stakeholders to raise questions, identify issues and be able to make evidence-based decisions (Schmoker, 2008).

This finding has important implications. One implication is that the system of education will have to distinguish between assessments that are meant to improve teaching and learning and examinations that produce data for use in accountability purposes. The benefit of a standards-based NRF is that the use of PLDs presents opportunity to intervene uniquely at individual learner and school levels. Not only do PLDs enable schools to have longitudinal data on cohorts of learners but the cumulative repertoire of knowledge and skills that individual learners accumulate from grade to grade can be tracked and monitored by developing performance standards that are articulated across grades.

Many of the relevant programs of education stand to benefit from the use of performance standards, especially when the PLDs are concrete and clear. Designing and developing teacher training programs will draw relevant information from PLDs on content that needs to be included, more relevantly in-service programs that need to be responsive to current issues.
in teaching. Developers of learning, teaching and support materials (LTSM) have a reliable and relevant source of information in PLDs. All these are programs that require a sound knowledge of the skills and knowledge that learners have already acquired and what are the next levels of knowledge and skills that they need to learn.

Adopting a standards-based NRF will also have capacity implications as well as implications for relevant tools. The DBE needs to develop a strategy for building the required capacity for standard setting, use of performance standards and establishing a culture of data use in provinces, districts and schools. In the short term, schools can be provided with simple pre-programmed Excel spread sheets to enter and analyse data from periodic assessments during the course of the year. They should be encouraged and assisted to use performance standards and cutscores that are set at national level to categorise learners into appropriate PLs. In the medium- to long-term, the DBE should develop and enhance a centralised data system to be accessible to districts and schools so that these levels of the education system can execute analyses of own data within the framework of a standard-based paradigm.

7.4.2 Guidelines and support to provinces and districts

This study explored guidelines and caveats that need to be taken into consideration in adopting a standards-based NRF that will enhance teaching and learning and improve learner achievement. Published research is unanimous on key lessons learnt and emerging guidelines from successful school reform projects and other initiatives that focused on data use to improve teaching and learning. Reporting data in meaningful and easy-to-interpret formats, ensuring that assessment data is relevant to what is being taught in schools, building the capacity of teachers on data use, providing supportive instructional leadership and promoting a culture of reflective collaboration on data use were some of the common lessons shared among all data-driven school improvement projects. A direct implication of this finding is that the NRF needs to be underpinned by these important lessons and guidelines.

7.4.3 Views of education officials on standards-based reporting

This study solicited the perceptions of national and provincial education officials who are responsible for delivery of curriculum, assessment and teacher development on whether they would prefer a raw score-based or standards-based reporting framework to communicate
results of national assessments. The respondents expressed overwhelming support for a standards-based reporting frame, citing meaningfulness, information-richness and user-friendliness of the reports that are based on a standards-based framework. One would expect that introducing a NRF that is based on performance standards would receive prompt support from these critical role players.

7.5 Conclusion

The key objectives of this study were specified in Chapter 1. The main objective was to develop a standards-based NRF for meaningful reporting of the results of LSAS in South Africa, using ANA as a specific example. The proposed NRF has been presented in Chapter 5. Partly the NRF was informed by evidence emanating from reviewed research literature from wide-ranging successful initiatives on effective use of data to improve teaching and learning. More importantly, the essential features of the NRF are underpinned by empirical evidence on requisite processes of setting performance standards and key outputs which are standards-based reports for learners, schools, districts, provinces and the national Department of Basic Education.

The second objective of the study was to investigate and identify the most appropriate method of setting performance standards that will enhance the usefulness of the proposed NRF. The investigation, which involved research experts and experienced teachers of language and mathematics from different contexts in South Africa, indicated that the OSS method was the most appropriate of the three methods that were compared in this study.

A particularly distinguishing feature of the OSS method included the fact that the specific “essential” content and skills that characterise learners at a particular performance level were identified not only from the hierarchy of competencies in the curriculum, but were also corroborated by teachers who taught the relevant subjects at the affected grades. As such, the OSS method was found to contribute significantly to ensuring that LSAS reports meet the curriculum needs of teachers.

Other advantages of the OSS over the Angoff method which also, to a lesser degree, met the curriculum needs of teachers, were that the former had a relatively very short turnaround time and cost less. It has the potential to reduce turnaround times for providing feedback to a
minimum. The efficacy of the OSS method has also been noted in other studies (Stone, 1995, 2001 & 2008; Stone et al, 2011; & Khatimin et al, 2013).

The third objective was to provide exemplars of standards-based reports as envisaged in the NRF. Exemplars of standards-based reports for learners, schools, districts, provinces and the national DBE are presented in Chapter 6. What is important is that each report is accompanied by a comprehensive description of what a learner(s) at a particular performance level can do and should know as well as recommended interventions for improvement. The rest of the report presents, in graphs and tables that are easy to read and understand, the distribution of learners across the performance levels and involves little to no demand for complicated data processing and decoding skills on the part of teachers and other users.

The categorisation of learners according to the knowledge and skills that they either possess or need makes it easy to plan differentiated interventions rather than common one-size-fits-all strategies. As can be observed in Chapter 6, once the data has been prepared through standard setting processes, the compilation and dissemination of reports can be completed in relatively short periods thus lending a valuable feature of cost-effectiveness to the use of a standards-based NRF.

7.6 Recommendations

In addition to the key findings, there were also a few constraints and incidental findings that were made in this study. Some of these have implications for improving data use in the education system, others have implications for future standard setting initiatives and use of performance standards to improve teaching and learning. Specific recommendations are made for various possibilities.

a) Use of performance standards to improve test development

In test-centred standard setting the design of tests is critical to setting meaningful performance standards. In this study there were some limitations in the design of the tests that were used for setting performance standards. The tests suffered serious ceiling effects. Consequently, the resulting cutscores were more stable and robust at the lower end of learner ability where there were adequate test items to estimate learner ability with minimum error.
Cutscores at the upper end of the ability continuum tended to be variable and imprecise most probably because there were very few test items that properly targeted higher ability learners.

b) Need for replication studies

One of the limitations of this study was in the use of very few panel members in the SS methods such as the OSS and the Angoff. Given the role that panels play in the validity of cutscores generated by these methods, it is recommended that the SS exercises using these two methods be replicated with adequate numbers of panel members, between 15 and 20 in each panel (Hambleton, 2001).

Although definite measures were taken to ensure that the findings of this research are valid e.g. by subjecting real and simulated datasets, representing different contexts, to the same standard setting methods in the cases of OBM and OSS methods, other appropriate methodological triangulation techniques (Cohen et al, 2011:196-197) could be used in replication studies to validate the observations that were made from the other research questions. For example, a replication study should involve collecting empirical data on what South African educators perceive are necessary conditions for data-driven decisions and verify if their perceptions are in tandem with what has been reported in research from elsewhere.

c) Focused pilot study of the NRF

This study recommended a NRF that is based on performance standards, underpinned by guidelines from successful studies on effective use of data to improve teaching and learning and seemed to enjoy the support of national and provincial education officials. However, nothing is known about possible constraints that the implementation of the NRF could face in the real contexts of provinces, districts and schools.

It is recommended that the DBE facilitate the piloting of NRF in select districts and schools that represent different contexts in which education is delivered in South Africa. The sample of districts and schools could represent urban and rural contexts, different quintiles and school sizes.
REFERENCES


California Department of Education, 2007. Development of performance level descriptors for the California standards tests (CSTs) and high school exit exam (CAASBEE).


Q. Moloi


Q. Moloi


Nzomo, J. & Makuwa, D. 2006. How can countries use cross-national research results to address ‘the big issues’?: case studies from Francophone Africa. In: Ross, K.N. and


Shulruf, B. 2015. E-mail message. [Notes in possession of author].


Q. Moloi


Van der Berg, S. 2007. How effective are poor schools: poverty and educational outcomes in South Africa. (Paper presented at the Centre for European, Governance and Economic
Q. Moloi


### Appendix A: Extract from a FAL rater sheet used by panellists in Angoff SS

<table>
<thead>
<tr>
<th>Item ID</th>
<th>Item description</th>
<th>Maximum marks</th>
<th>Type of question</th>
<th>Rating 1</th>
<th>Rating 2</th>
<th>Rating 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Basic</td>
<td>Proficient</td>
<td>Advanced</td>
<td>Basic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Proficient</td>
<td>Advanced</td>
<td>Basic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Proficient</td>
<td>Advanced</td>
<td>Basic</td>
</tr>
<tr>
<td>6F1</td>
<td>Discusses character and setting in a story</td>
<td>1</td>
<td>MC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6F2</td>
<td>Expresses cause and effect relations</td>
<td>1</td>
<td>MC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6F6</td>
<td>Responds to a direct question on a story</td>
<td>2</td>
<td>OE</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>6F62</td>
<td>Punctuates a sentence correctly</td>
<td>3</td>
<td>OE</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
</tbody>
</table>

**Mean Estimate** the mean score of all just proficient learners.

0; 0.5; 1; 1.5; 2; 2.5; 3; 3.5; 4; 4.5; 5; 5.5; 6 etc.

(Allocate one of these where the word "Mean" appears in a box)
# Appendix B: Grade 6 mathematics rating sheet and summary (Angoff SS)

<table>
<thead>
<tr>
<th>Item</th>
<th>Max</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>R1  R2  R3  R4</td>
<td>R1  R2  R3  R4</td>
<td>R1  R2  R3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Min  Max _rng</td>
<td>Min  Max _rng</td>
<td>Min  Max _rng</td>
</tr>
<tr>
<td>Q1_1</td>
<td>1</td>
<td>30   40   40</td>
<td>70   70   80</td>
<td>80   80   100</td>
</tr>
<tr>
<td>Q1_2</td>
<td>1</td>
<td>10   30   30</td>
<td>50   50   80</td>
<td>80   80   80</td>
</tr>
<tr>
<td>Q1_3</td>
<td>1</td>
<td>5    20   20</td>
<td>50   50   60</td>
<td>60   60   100</td>
</tr>
<tr>
<td>Q2</td>
<td>1</td>
<td>20   20   20</td>
<td>80   80   80</td>
<td>80   80   80</td>
</tr>
<tr>
<td>Q3</td>
<td>1</td>
<td>15   10   10</td>
<td>70   70   80</td>
<td>80   80   80</td>
</tr>
<tr>
<td>Q4_1</td>
<td>1</td>
<td>20   30   30</td>
<td>80   80   90</td>
<td>90   90   90</td>
</tr>
<tr>
<td>Q4_2</td>
<td>1</td>
<td>40   30   40</td>
<td>80   80   90</td>
<td>90   90   90</td>
</tr>
<tr>
<td>Q5</td>
<td>2</td>
<td>0    0    0</td>
<td>1.5   1   1</td>
<td>1   1   1</td>
</tr>
<tr>
<td>Q6_1</td>
<td>1</td>
<td>15   20   20</td>
<td>80   80   90</td>
<td>90   90   90</td>
</tr>
<tr>
<td>Q6_2</td>
<td>1</td>
<td>15   20   20</td>
<td>80   80   90</td>
<td>90   90   90</td>
</tr>
<tr>
<td>Q7</td>
<td>2</td>
<td>0    0    0</td>
<td>11.5  11.5  12</td>
<td>11  1   2</td>
</tr>
<tr>
<td>Q8_1</td>
<td>2</td>
<td>1    0    0</td>
<td>2     1   1</td>
<td>2   2   2</td>
</tr>
<tr>
<td>Q8_2</td>
<td>2</td>
<td>1    0    0</td>
<td>1     1   1</td>
<td>1   1   1</td>
</tr>
<tr>
<td>Q8_3</td>
<td>1</td>
<td>60   40   40</td>
<td>80   80   80</td>
<td>80   80   80</td>
</tr>
<tr>
<td>Q8_4</td>
<td>1</td>
<td>60   20   20</td>
<td>80   80   79</td>
<td>90   90   90</td>
</tr>
<tr>
<td>Q8_5</td>
<td>2</td>
<td>0    0    0</td>
<td>2     1   1</td>
<td>2   2   2</td>
</tr>
<tr>
<td>Q8_6</td>
<td>4</td>
<td>0    1    1</td>
<td>1     1   1</td>
<td>2   2   2</td>
</tr>
<tr>
<td>Q9_1</td>
<td>3</td>
<td>0    0    0</td>
<td>0     1   1</td>
<td>1   1   1</td>
</tr>
<tr>
<td>Q9_2</td>
<td>2</td>
<td>0    0    0</td>
<td>0     1   1</td>
<td>1   1   1</td>
</tr>
</tbody>
</table>

*Minimum (Min), Maximum (Max) and Range (Rng) values are not shown for Advanced level

*R1=Round 1; R2=Round 2, etc*
<table>
<thead>
<tr>
<th>Item</th>
<th>Max</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q10_2</td>
<td>4</td>
<td>0  1  1  0  0  1</td>
<td>1  1  2  2  1  1</td>
<td>2  3  3</td>
</tr>
<tr>
<td>Q11</td>
<td>3</td>
<td>0  0  0  0  0  0</td>
<td>0  1  1  1  1  1</td>
<td>0  1.5  2  2</td>
</tr>
<tr>
<td>Q12</td>
<td>3</td>
<td>0  0  0  0  0  0</td>
<td>0  1  1  1  1  1</td>
<td>0  2  2  2</td>
</tr>
<tr>
<td>Q13</td>
<td>2</td>
<td>0.5 0  0  0.5  0  1</td>
<td>1  1  1  1  1  1</td>
<td>0  1  1</td>
</tr>
<tr>
<td>Q14</td>
<td>1</td>
<td>10  20  20  10  20  10</td>
<td>60  60  80  80  60  80</td>
<td>20  80  90  90</td>
</tr>
<tr>
<td>Q15</td>
<td>1</td>
<td>10  20  20  10  20  10</td>
<td>1.5  1.5  60  60  2  60</td>
<td>59  80  80  80</td>
</tr>
<tr>
<td>Q16</td>
<td>2</td>
<td>0  0  0  0  0  0</td>
<td>0  1  1  1  1  1</td>
<td>0  2  2  2</td>
</tr>
<tr>
<td>Q17</td>
<td>2</td>
<td>0  0  0  0  0  0</td>
<td>0  1  1  1  1  1</td>
<td>0  1  2  2</td>
</tr>
<tr>
<td>Q18_1</td>
<td>1</td>
<td>30  40  40  30  30  40</td>
<td>10  70  70  80  70  80</td>
<td>10  80  100  100</td>
</tr>
<tr>
<td>Q18_2</td>
<td>1</td>
<td>40  30  30  40  30  40</td>
<td>10  80  80  80  80  80</td>
<td>0  80  90  90</td>
</tr>
<tr>
<td>Q19_1</td>
<td>1</td>
<td>40  20  20  40  20  40</td>
<td>20  80  80  70  70  70</td>
<td>80  10  90  90</td>
</tr>
<tr>
<td>Q19_2</td>
<td>1</td>
<td>10  20  20  10  20  10</td>
<td>10  70  70  80  70  80</td>
<td>10  90  90  90</td>
</tr>
<tr>
<td>Q19-3</td>
<td>1</td>
<td>20  20  20  20  20  20</td>
<td>0  80  80  80  80  80</td>
<td>0  80  100  100</td>
</tr>
<tr>
<td>Q20-1</td>
<td>1</td>
<td>5  20  20  5  5  20</td>
<td>15  70  70  80  70  80</td>
<td>10  80  90  90</td>
</tr>
<tr>
<td>20-2</td>
<td>1</td>
<td>10  30  30  10  30  20</td>
<td>20  80  80  80  80  80</td>
<td>0  80  90  90</td>
</tr>
<tr>
<td>20-3</td>
<td>1</td>
<td>10  20  20  10  20  10</td>
<td>10  80  80  70  70  70</td>
<td>10  90  90  90</td>
</tr>
<tr>
<td>20-4</td>
<td>1</td>
<td>0  20  20  0  0  20</td>
<td>20  80  80  70  70  70</td>
<td>10  90  90  90</td>
</tr>
<tr>
<td>21-1</td>
<td>2</td>
<td>30  0  0  0  0  30</td>
<td>30  1  1  1  1  1</td>
<td>0  1.5  2  2</td>
</tr>
<tr>
<td>21-2</td>
<td>1</td>
<td>50  10  10  30  10  50</td>
<td>40  80  80  60  60  60</td>
<td>20  90  80  80</td>
</tr>
<tr>
<td>22-1</td>
<td>1</td>
<td>30  50  50  30  30  50</td>
<td>20  90  90  90  90  90</td>
<td>0  100  100  100</td>
</tr>
<tr>
<td>22-2</td>
<td>1</td>
<td>0  40  40  30  40  40</td>
<td>0  80  80  80  80  80</td>
<td>0  90  90  90</td>
</tr>
<tr>
<td>23-1</td>
<td>2</td>
<td>0  0  0  0  0  0</td>
<td>0  0.5  0.5  1  1  1  1</td>
<td>1  1  2  2</td>
</tr>
<tr>
<td>23-2</td>
<td>2</td>
<td>20  0  0  1  0  1</td>
<td>1  1  1  1  1  1</td>
<td>0  1  2  2</td>
</tr>
<tr>
<td>23-3</td>
<td>1</td>
<td>10  50  50  10  10  50</td>
<td>40  80  80  70  70  70</td>
<td>10  90  90  90</td>
</tr>
</tbody>
</table>
Appendix C: Extract from a FAL rating form used for the OSS

<table>
<thead>
<tr>
<th>Item ID</th>
<th>Item description</th>
<th>Skills</th>
<th>Skills (Abbrev.)</th>
<th>Basic</th>
<th>Proficien</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>6F1</td>
<td>Discusses character and setting in a story</td>
<td>Reading &amp; Viewing</td>
<td>R&amp;V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6F2</td>
<td>Expresses cause and effect relations</td>
<td>Reading &amp; Viewing</td>
<td>R&amp;V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6F6</td>
<td>Responds to a direct question on a story</td>
<td>Reading &amp; Viewing</td>
<td>R&amp;V</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6F23_2</td>
<td>Punctuates a sentence correctly</td>
<td>Language Structures &amp; Conventions</td>
<td>LS&amp;C</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix D: Standard-setting Evaluation Instrument

<table>
<thead>
<tr>
<th>No.</th>
<th>Criteria</th>
<th>Description</th>
<th>Research reference</th>
<th>Possible impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Resource intensity</td>
<td>To what extent does the SS method incur costs, human resource time, etc?</td>
<td>Hambleton (2000:7)</td>
<td>High resource intensity may not be sustainable</td>
</tr>
<tr>
<td>2</td>
<td>Technical demands</td>
<td>How much technical expertise is required to complete the SS process and prepare feedback?</td>
<td>Hambleton (2000:6)</td>
<td>Demand for high (scarce) skills may cause delay or discontinuity of SS</td>
</tr>
<tr>
<td>3</td>
<td>Precision of cutscores</td>
<td>To what extent does the SS method account for and reduce measurement error in the cutscores?</td>
<td>Hambleton (2000:11)</td>
<td>Large standard errors in cutscores may lead to loss of confidence in the performance standards</td>
</tr>
<tr>
<td>4</td>
<td>Consistency of cutscores</td>
<td>Would different panels with similar characteristics arrive at same cutscores using this SS method?</td>
<td>Hambleton (2000:6)</td>
<td>Variations of cutscores depending on panellists may invalidate the performance standards</td>
</tr>
<tr>
<td>5</td>
<td>Content validity</td>
<td>Are the resulting performance categories from this SS method relevant to the subject or users?</td>
<td>Hambleton (2000:6), TIMPERLY (2009:1), Kane (2001)</td>
<td>Irrelevant performance standards may lead to non- or under-utilisation of assessment feedback</td>
</tr>
<tr>
<td>6</td>
<td>Need for specialised software</td>
<td>Does the SS method require specialised software that may not be easily available?</td>
<td>Hambleton (2000:7), Ricker (2006,</td>
<td>Specialised software may be costly to purchase and may require high levels of technical expertise to effectively use</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Turnaround time</td>
<td>Can SS process be completed in less time using this method or another method?</td>
<td>Hambleton (2000:7)</td>
<td>SS method(s) that take long may be costly, delay feedback to data users and miss critical time requirements of data users</td>
</tr>
</tbody>
</table>
Appendix E: Questionnaire and templates used for survey

Dear Colleague

Request for objective views on how to report ANA results meaningfully and effectively

The purpose of reporting assessment results is to provide useful and meaningful feedback that will enable users of reports to take appropriate action(s) to address matters of concern arising out of the assessment. For instance, following an assessment, teachers need to know what appropriate teaching strategies to take in order to help learners learn better. Subject advisors also need to know specific curriculum areas where teachers need support in order to teach better.

Arguing from this premise, I am exploring a meaningful, information-rich and user-friendly template for reporting ANA results at national and provincial levels. Kindly make an objective assessment of the two reporting templates, Template 1 and Template 2 below and give an honest, candid and well-thought-out option that would make reporting of ANA results more meaningful and effective. There are seven (7) levels of performance in Template 1 and four (4) in Template 2. In each template the symbols used under “% Learners at this level” are only indicative (fictitious). They have no specific value attached to them and may not necessarily be used to compare whether there are more or fewer learners in one level compared to another. Any views or fears on the complexity of preparing or compiling the report for either template must not deter you from evaluating the templates. Relevant training and expertise will be provided in each case.

Criteria for comparing the reporting templates are described below. In each case make a cross (X) just behind the number of the template you select:

i. User-friendliness: Which template makes reading of the results more user-friendly? [ ] or [ ]

ii. Information-richness: Which template provides a report that communicates more information to end-users, e.g. teachers, subject advisors? [ ] or [ ]

iii. Clarity: Which template provides clearer information on what must be done to improve weaknesses identified in the assessment? [ ] or [ ]

iv. Meaningfulness: Which template provides a report that is more meaningful in terms of what needs to be done from the results? [ ] or [ ]

v. Curriculum-specificity: Which template provides more specific curriculum-related information? [ ] or [ ]

vi. Knowledge & skill deficiency: Which template will better help identify learner i) knowledge deficiencies [ ] or [ ], ii) skills deficiencies [ ] or [ ]?

vii. Personal choice: Which template would you prefer if you were i) a teacher? [ ] or [ ], ii) a subject advisor? [ ] or [ ], iii) a parent? [ ] or [ ]?

viii. Use with confidence: Which template would present results that you could use with more confidence, [ ] or [ ]?
Reporting Template 1
Summary of Provincial Learner Performance: ANA 2014
MANDELA PROVINCE
Mathematics Grade 6

<table>
<thead>
<tr>
<th>Level No.</th>
<th>Level Definition</th>
<th>% Learners at this level</th>
<th>Score Range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Not Achieved</td>
<td>X</td>
<td>0–29</td>
</tr>
<tr>
<td>2</td>
<td>Elementary Achievement</td>
<td>Y</td>
<td>30–39</td>
</tr>
<tr>
<td>3</td>
<td>Moderate achievement</td>
<td>Z</td>
<td>40–49</td>
</tr>
<tr>
<td>4</td>
<td>Adequate achievement</td>
<td>XY</td>
<td>50–59</td>
</tr>
<tr>
<td>5</td>
<td>Substantial Achievement</td>
<td>XZ</td>
<td>60–69</td>
</tr>
<tr>
<td>6</td>
<td>Meritorious Achievement</td>
<td>YY</td>
<td>70–79</td>
</tr>
<tr>
<td>7</td>
<td>Outstanding Achievement</td>
<td>YZ</td>
<td>80–100</td>
</tr>
</tbody>
</table>

Reporting Template 2
Summary of Provincial Learner Performance: ANA 2014
MANDELA PROVINCE
Mathematics Grade 6

<table>
<thead>
<tr>
<th>Level No.</th>
<th>Level Definition</th>
<th>% Learners at this level</th>
<th>Level Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Not Achieved</td>
<td>A</td>
<td>Performance at this level is indicated largely by recognition and manipulation of iconic/pictorial forms of information</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Skills</strong> demonstrated by learners who function at this level include ability to:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• <strong>Recognise</strong> shapes and numbers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• <strong>Count</strong> forwards</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• <strong>Perform</strong> single operations of addition and, to a less extent, subtraction, on whole numbers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• <strong>Match</strong> words with shapes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Progression Implications</th>
<th>Intervention Implications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners at this level are highly unlikely to progress to the next grade</td>
<td>Learners at this level require basic direct teaching with concrete materials to scaffold their learning</td>
</tr>
<tr>
<td>Level</td>
<td>Partially Achieved</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Achieved</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Advanced</td>
</tr>
<tr>
<td></td>
<td>D</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other comments and suggestions for improving the quality and meaningfulness of assessment results
### Appendix F: Grade 3 PLs and PLDs

#### Grade 3 Home Language

<table>
<thead>
<tr>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3 learners performing at the Partially Achieved level should have knowledge of basic phonics, read a text alone, and recognise and comment on key details in a text.</td>
<td>Grade 3 learners performing at the Achieved level should be able to have knowledge of complex patterns of phonics, read, comment and recognise key details in a text, state the cause and effects of the story, and use self correcting methods when reading.</td>
<td>Grade 3 learners performing at the advanced level should be able to read a variety of complex texts (e.g. fiction, non-fiction), identify, find and use information from different sources in an appropriate way, and analyse texts for socio-cultural values, attitudes and assumptions.</td>
</tr>
<tr>
<td><strong>WRITING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3 learners performing at the Partially Achieved level should be able to build vocabulary, spell words, and use appropriate grammar and punctuation, write a selection of short texts with ease and increased speed, and be able to complete the written tasks within an appropriate set time.</td>
<td>Grade 3 learners performing at the Achieved level uses pre-writing strategies (brainstorming, visual images), begin to plan their writing, write different kinds of factual and imaginative texts, write a title that reflects the content and for different purposes and audiences, and are able to confidently share their written work with others.</td>
<td>Grade 3 learners performing at the advanced level should be able to edit their own writing, review and give feedback to the writing of others; must be able to process information in different ways and develop critical language awareness.</td>
</tr>
</tbody>
</table>

#### Grade 3 Mathematics

<table>
<thead>
<tr>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Partially Achieved</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 3 learners performing at the Partially Achieved level in numeracy should be able to know, read and write number symbols and names from 1 to 1000; count (forwards and backwards) and in multiples (of up to 1000); convert between rands and cents; add and subtract two digit numbers; identify and name two dimensional objects; determine lines of symmetry;</td>
<td>Grade 3 learners performing at the Achieved level should be able to add and subtract at least three digit, multiply and divide 2 digits by one digits; do problem solving; sequencing and extending patterns; identify and name three dimensional objects; read analogue and digital clocks; sort, read, draw and interpret given data with 1 to 1 correspondence.</td>
<td>Grade 3 learners performing at the advanced level should be able to master all the Achieved level skills and work with bigger number ranges and measure mass, lengths and capacity; determine the area of a two dimensional shape; read, draw and interpret given data with more than 1 to 1 correspondence; estimate answers and round off to 10. The learners should also be able to explain solutions to problems.</td>
</tr>
</tbody>
</table>

---

11 Level descriptors developed at a workshop in 2012 convened by DBE involving district and provincial officials as well as teachers.
### Appendix G: Grade 6 PLs and PLDs

#### Grade 6 English Home Language

<table>
<thead>
<tr>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The learner finds it difficult to read and to view for information or enjoyment, because his/her foundation of phonetics is not developed. Hence the learner is discouraged to read aloud and to adjust to reading strategies to suit the purpose and audience. Because of his shortcoming e.g. poor phonetic development, he cannot explain and interpret overall response to text and cannot give reasons based on text or own experience. The learner cannot explain themes, plots and characters because of a problem of cognitive retention.</td>
<td>The learner is able to read for information and enjoyment a variety of fiction and non-fiction texts for different purposes e.g. magazines and poems. The learner reads aloud and silently adjusts reading strategies to suit the purpose and audience. He/she also uses appropriate reading strategies like skimming, scanning, predictions etc. The learner can view and discuss various visual and multimedia texts such as TV, internet, and videos/films etc. The learner can explain interpretations and overall response to text, giving reasons based on text or own experience and shows understanding of the text in relation to her/his life and how she/he functions. He/she can also explain themes, plot, setting and characterisation. The learner can identify and critically discuss cultural and social values in texts.</td>
<td>Learner adds his own opinion and perspective to variety of texts read. She can confidently read aloud and can silently adjust reading strategies to suit the purpose and audience. The learner can creatively discuss, view various visual and multimedia texts and also ask own questions for deeper understanding. The learner cannot only interpret, but can also explain overall response giving own reasons based not only on text, but also on own experience. She can also explain themes, thought, settings and characterisation, adding her own perspective and opinions. This holds especially with regards to cultural and social values and texts.</td>
</tr>
<tr>
<td><strong>WRITING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Due to poor phonetic development the learner cannot write different kinds of factual and imaginative text for a wide range of purposes. The learner can only use language in a limited way to think and reason and as well as to access, process and use information for learning because the learner has a poor phonetic development. The learner cannot construct sentences because of lack of knowledge of punctuation and parts of speech.</td>
<td>The learner can write different kinds of factual and imaginative texts for a wide range and purposes. The learner can use language to think and reason as well as to access, process and use information for learning. He also knows how to use the sounds, words and grammar of the language to create and interpret text. The learner can write for personal, exploratory, playful, imaginative and creative purposes. He/she can also organise and develop ideas through the writing process e.g. brainstorm ideas and topics, and develop ideas by consulting a variety of sources.</td>
<td>The learner can independently create own personal spelling list and always explores new words for vocabulary and enrichment. The learner constructs complex sentences, because of sound knowledge of punctuation and principles of parts of speech. The learner writes creatively because she reads a variety of texts and also reads extensively.</td>
</tr>
</tbody>
</table>

#### Grade 6 English First Additional Language

<table>
<thead>
<tr>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The learner reads text with simple sentences. She/he needs assistance with scanning and skimming of poetical elements. Has a vocabulary of 1500 words.</td>
<td>The learner can read in order to interpret meaning of stories, relating it to real life experiences. She/he can identify the basic elements of a story e.g. characters, plot setting. She/he can focus on and develop ethical issues such as cultural/social diversity. She/he can highlight the moral lesson behind the story. She/he can identify elements in poetry e.g. rhyme, alliteration, onomatopoeia and personification. She/he can skim, scan and summarise text. She/he can write friendly letters. The learner can compile book reports: fiction and non-fiction. The learner uses dictionaries in striving at building a vocabulary of at least 3000 words.</td>
<td>The learner reads and understands complex texts in formal and informal stories. She/he reads poetry, evaluates text and is able to reflect on the information that is read. She/he identifies formal and informal text and builds a vocabulary of 5000 words and uses the words to critique text through book reviews and reports.</td>
</tr>
<tr>
<td><strong>WRITING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>She/he writes for social purposes with assistance e.g. personal letters, diaries, news reports and the creating of a book cover. When drafting text</td>
<td>The learner writes information in graphic and descriptive form. She/he writes for social purposes, e.g. frames, personal letters, keeping a diary (reflection). The learner writes dialogues, simple news reports and friendly letters. She/he</td>
<td>She/he writes and develops information from graphical and descriptive form to text. The learner does extensive writing for social purposes e.g. writing letters, which are reflective, developing news reports and designing questionnaires and</td>
</tr>
</tbody>
</table>

---

12 Level descriptors developed at a workshop in 2012 convened by DBE involving district and provincial officials as well as teachers.
the learner is able to identify similarities and differences as well as express cause and effect among things. Pictures are used to write simple sentences. can design a book cover. She/he develops and edits language structures. The learner focuses on sentence construction, grammar, spelling and punctuation. She/he practices handwriting skills. adverts. She/he writes extensively to develop sentences that are constructed grammatically. The learner writes and integrates ideas by classifying information, solving problems and questioning. The learner develops auditory memory through listening comprehension.

**Grade 6 Mathematics**

<table>
<thead>
<tr>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>A learner at this level should be able to recognise various number systems and count forward and backwards in decimals. They should be able to recognise place value up to 9 digits. Round off up to 1000. The learner should be able to add and subtract up to 9 digits as well as multiplying and dividing 4 digit by 3 digit numbers without remainders. They should recognise and do simple calculations using fractions and decimals. The learner should be able to identify and describe basic geometric and numeric patterns. They should recognise 2D shapes and 3D objects. They should identify and manipulate different objects to describe transformations. To read and tell digital and analogue time is essential as well as estimate and measure SI units. They should have the ability to pose simple questions, collect and record data. Draw and interpret simple bar and pictographs.</td>
<td>A learner at this level should be able to describe and illustrate various number systems, they should count, recognise and do calculations using fractions and decimals. They have the ability to represent multiples, factors and prime numbers and find percentages of whole numbers and solve problems involving finances and measurement; and to compare rate and ratio, and to identify, formulate and describe numeric and geometric patterns. It is necessary for learners to use and describe transformations, symmetry, draw enlargements and reductions of these shapes and objects. The ability to locate and describe movement on a grid is required. The learner must read and tell analogue time and solve problems involving different time zones. They need to estimate, record, compare and convert between SI units (including mass, temperature, distance and capacity) Measuring area, perimeter and volume of shapes is of utmost importance. They should be able to recognise, measure and describe different angle sizes. Learners should differentiate between sample and population, organise and record data, and be able to calculate the median and mode, draw and interpreting bar and pictographs is necessary. They should be able to list possible outcomes and predict “likelihood” of events.</td>
<td>Learners at this level should be able to demonstrate and surpass all the knowledge of a learner at the Partially Achieved and Achieved level as well as show advanced mathematical skills in the following. Critically read, interpret and analyse with awareness of sources of error and manipulation to draw conclusions and make predictions. The learners should list possible outcomes for simple experiments including tossing a coin, rolling a die and spinning a spinner. They should have the ability to distinguish between volume, surface area and the dimensions of rectangular prisms.</td>
</tr>
</tbody>
</table>
**Appendix H: Grade 9 PLs and PLDs**

### Grade 9 English Home Language

<table>
<thead>
<tr>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advance level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The learners at this level are able to read a limited range of texts. He/she is able to select and record relevant information and texts for personal needs.</td>
<td>The learner is able to read independently a wide range of texts. She/he is able to explain the purpose, audience and context of a text. She/he shows an understanding of and is able to identify and evaluate the main ideas of the texts.</td>
<td>The learner is able to read spontaneously and extensively a wide range of text types. She/he is able to compare personal responses and makes motivated recommendations. She/he demonstrates an ability to analyse an interpret critically and relate to own life experiences with regards to plot, themes, values, characters and setting. The learner is easily able to generate complex texts independently for a wide range of purposes.</td>
</tr>
<tr>
<td><strong>WRITING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The learner writes a limited range of texts using simple sentences although not always grammatically correct. There are gaps in the development and execution of ideas and a lack of cohesion of the final effort. He/she seldom uses appropriate language to think and reason. She/he writes an speaks using simple language and grammatical structures. The learner is able to use sounds, words and grammar, however she/he experiences difficulty in creating and interpreting texts.</td>
<td>The learner demonstrates an ability to write different kinds of factual and imaginative texts for a wide range of purposes. She/he produces a wide range of written multimedia texts for various purposes. The learner demonstrates an ability to analyse, evaluate and make recommendations. She/he is able to use language to think and reason, as well as access, process and use information for learning. She/he applies thinking and reasoning skills in a variety of contexts across the curriculum. She/he is able to use language to investigate and explore by using questions on cross-cultural issues. The learner will know and be able to use the sounds, words and grammar of the language to create and interpret texts. She/he develops awareness and use of language style.</td>
<td>She/he demonstrates advanced skills in a range of features of writing appropriate to the text type. She/he is able to use increasingly complex texts and analyses multiple drafts considering purpose, audience, point of view, bias and complex organisation and varied elements of style. The learner uses a wide and varied range of strategies in grammar to create and interpret texts. She/he critically analyses the grammatical difference between statements, questions, commands, instructions and exclamations.</td>
</tr>
</tbody>
</table>

### Grade 9 English First Additional

<table>
<thead>
<tr>
<th>Partially Achieved</th>
<th>Achieved</th>
<th>Advanced level</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>READING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The learner is able to read, interpret, analyse and compare uncomplicated texts that consist of simple sentences. The learner may be able to find main ideas in a simple text. The learner may be able to identify some characters and a straightforward plot setting. The learner can answer basic questions and give simple examples.</td>
<td>The learner must be able to read, interpret, analyse and compare ordinary texts that consist of simple and compound sentences. The learner can identify the topic sentence and main ideas in complicated texts. The learner can identify some characters and the complicated plot. The learner can answer and ask simple questions and give examples and evidence.</td>
<td>The learner is able to read critically and interpret, analyse and compare complicated and difficult texts that consist of simple, compound and complex sentences. The learner can identify the topic sentence, main ideas, register and style in difficult texts. The learner can identify all characters and very complicated plots. The learner is able to ask and answer complex questions and give many examples and evidence.</td>
</tr>
<tr>
<td><strong>WRITING</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The learner can write different types of text using simple language and some planning skills. The learner can select and rewrite information from simple text. The learner can use the present; past and future tense correctly and use some of the grammatical structures correctly. The learner may be able to summarise a short text of about 100 words and make simple sentences because of a limited vocabulary of about 900 words.</td>
<td>The learner can write different types of texts using creative and some figurative language by planning and editing their texts. The learner can use most of the tenses and grammar structures correctly. The learner can summarise complex texts of 200 words and make compound sentences because they have a vocabulary rating of 1200 words.</td>
<td>The learner can write a variety of complex texts by being highly creative and using a variety of figurative language. By brainstorming, planning, editing and rewriting the learner can use all tenses and grammatical structures correctly. The learner is able to summarise difficult texts of 300 words and make complex sentences of a vocabulary of 1500 words.</td>
</tr>
</tbody>
</table>

### Grade 9 Mathematics

13 Level descriptors developed at a workshop in 2012 convened by DBE involving district and provincial officials as well as teachers
Partially Achieved | Achieved | Advance level
--- | --- | ---
At this level the learner should be able to analyse budgets and calculate profit and loss. They should be able to apply at least the first two laws of exponents and solve basic equations. Identifying input and output values, drawing of graphs on a Cartesian plane and presenting a variety of data in different forms of graphs is essential. They should name geometric figures and solids and be able to substitute values in a given formula and determine probability using given examples.

A learner at this level should be able to analyse budgets and calculate simple as well as compound interest, they should have the ability to calculate time, speed and distance and be able to use and apply the laws of exponents. They should have the ability to factorise and simplify algebraic expressions in calculations and investigate a variety of numeric and geometric patterns as well as solve equations by using algebraic processors. Drawing and analysing of graphs on a cartesian plane is a priority for a learner at this level. They should have the ability to draw and interpret a variety of graphs (Histograms, bar graphs, pie charts, broken line, line and scatter plots), recognise visualise and name geometric figures and solids. They must use transformation, congruency and similarity to investigate, describe and justify properties of geometric figures. Selecting and using appropriate formulae and measurements to calculate perimeter, area and volume as well as probability using a 2 way table and tree diagrams is a prerequisite.

Learners at this level should be able to demonstrate all the knowledge of a learner at the Partially Achieved and Achieved level as well as show advanced mathematical skills in the following. Critically read, interpret and analyse with awareness of sources of error and manipulation to draw conclusions and make predictions. The learners should discuss the differences between the probability of outcomes and their relative frequency and make a sound judgement.
Appendix I: Grade 6 mathematics item difficulties from OSS method\textsuperscript{14}

<table>
<thead>
<tr>
<th>Item ID\textsuperscript{15}</th>
<th>Item description</th>
<th>Content Area (Abbr.)</th>
<th>Ratings (in logits) by individual Panel Members</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Terence</td>
<td>Pieter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Partly Achieved</td>
<td>Achieved</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Advanced</td>
<td>Partly Achieved</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Achieved</td>
<td>Advanced</td>
</tr>
<tr>
<td>6M1_1</td>
<td>Investigate and extend numeric patterns</td>
<td>PFA\textsuperscript{16}</td>
<td>-2.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Terence</td>
<td>Pieter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Partly Achieved</td>
<td>Achieved</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Advanced</td>
<td>Partly Achieved</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Achieved</td>
<td>Advanced</td>
</tr>
<tr>
<td>6M1_2</td>
<td>Count backwards in decimal fraction</td>
<td>NOR\textsuperscript{17}</td>
<td>-1.84</td>
</tr>
<tr>
<td>6M1_3</td>
<td>Practical measuring of 2-D shapes</td>
<td>S&amp;S</td>
<td>-1.01</td>
</tr>
<tr>
<td>6M1_4</td>
<td>Solve number sentences by inspection</td>
<td>NOR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Terence</td>
<td>Pieter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Partly Achieved</td>
<td>Achieved</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Advanced</td>
<td>Partly Achieved</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Achieved</td>
<td>Advanced</td>
</tr>
<tr>
<td>6M1_5</td>
<td>Determine the middlemost score</td>
<td>DH\textsuperscript{18}</td>
<td>-1.14</td>
</tr>
<tr>
<td>6M1_6</td>
<td>Round off to the nearest 100 000</td>
<td>NOR</td>
<td>-0.22</td>
</tr>
<tr>
<td>6M1_7</td>
<td>Factors of 2 digit whole numbers</td>
<td>NOR</td>
<td>-0.18</td>
</tr>
<tr>
<td>6M1_8</td>
<td>Practical measuring of 3-D objects</td>
<td>ME\textsuperscript{19}</td>
<td>0.03</td>
</tr>
<tr>
<td>6M1_9</td>
<td>Common fraction</td>
<td>NOR</td>
<td>0.17</td>
</tr>
<tr>
<td>6M1_10</td>
<td>Views of geometric objects</td>
<td>S&amp;S\textsuperscript{20}</td>
<td>0.06</td>
</tr>
<tr>
<td>6M2</td>
<td>Represent numbers up to at least 9 digit number</td>
<td>NOR</td>
<td>-0.27</td>
</tr>
<tr>
<td>6M3</td>
<td>Represent prime numbers</td>
<td>NOR</td>
<td>1</td>
</tr>
<tr>
<td>6M4</td>
<td>Recognize place value to at least 9-digit number</td>
<td>NOR</td>
<td>-1.1</td>
</tr>
</tbody>
</table>

\textsuperscript{14} Two of six raters shown
\textsuperscript{15} Grade_Subject_Item Number, e.g. 6M1_1=Grade 6 Math Item 1.1
\textsuperscript{16} Patterns, Functions & Algebra
\textsuperscript{17} Number, Operations & Relations
\textsuperscript{18} Data Handling
\textsuperscript{19} Measurement
\textsuperscript{20} Structure & Shape
<table>
<thead>
<tr>
<th>Item ID&lt;sup&gt;11&lt;/sup&gt;</th>
<th>Item description</th>
<th>Content Area (Abbr.)</th>
<th>Ratings (in logits) by individual Panel Members</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Terence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Partly Achieved</td>
</tr>
<tr>
<td>6M5_1</td>
<td>Addition whole number</td>
<td>NOR</td>
<td>-0.96</td>
</tr>
<tr>
<td>6M5_2</td>
<td>Subtraction of whole number</td>
<td>NOR</td>
<td>-0.53</td>
</tr>
<tr>
<td>6M5_3</td>
<td>Multiplication</td>
<td>NOR</td>
<td>0.43</td>
</tr>
<tr>
<td>6M5_4</td>
<td>Long division</td>
<td>NOR</td>
<td>0.23</td>
</tr>
<tr>
<td>6M5_5</td>
<td>Addition of mixed numbers</td>
<td>NOR</td>
<td>-0.89</td>
</tr>
<tr>
<td>6M5_6</td>
<td>Find a percentage of whole numbers</td>
<td>NOR</td>
<td>0.28</td>
</tr>
<tr>
<td>6M5_7</td>
<td>Addition and subtraction of decimal fractions</td>
<td>NOR</td>
<td>-0.5</td>
</tr>
<tr>
<td>6M6</td>
<td>Multiple operations on whole numbers</td>
<td>NOR</td>
<td>1.12</td>
</tr>
<tr>
<td>6M7</td>
<td>Recognize and use the distributive property</td>
<td>NOR</td>
<td>-0.13</td>
</tr>
<tr>
<td>6M8</td>
<td>Solve problems involving financial contexts</td>
<td>NOR</td>
<td>-1.27</td>
</tr>
<tr>
<td>6M9</td>
<td>Use addition and subtraction as inverse operations</td>
<td>NOR</td>
<td>-0.66</td>
</tr>
<tr>
<td>6M10</td>
<td>Multiples of 3-digit numbers</td>
<td>NOR</td>
<td>0.99</td>
</tr>
<tr>
<td>6M12</td>
<td>Equivalence common fraction, decimal fraction</td>
<td>NOR</td>
<td>-0.57</td>
</tr>
<tr>
<td>6M13</td>
<td>Compare and order decimal fractions</td>
<td>NOR</td>
<td>-0.88</td>
</tr>
<tr>
<td>6M14</td>
<td>Extend numeric patterns</td>
<td>PFA</td>
<td>1.73</td>
</tr>
<tr>
<td>6M15</td>
<td>Determine input values</td>
<td>PFA</td>
<td>0.02</td>
</tr>
</tbody>
</table>

<sup>11</sup> Item ID: 6M5_1, 6M5_2, 6M5_3, 6M5_4, 6M5_5, 6M5_6, 6M5_7, 6M6, 6M7, 6M8, 6M9, 6M10, 6M11, 6M12, 6M13, 6M14, 6M15
<table>
<thead>
<tr>
<th>Item ID</th>
<th>Item description</th>
<th>Content Area (Abbr.)</th>
<th>Ratings (in logits) by individual Panel Members</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Terence</td>
</tr>
<tr>
<td>6M16</td>
<td>Different descriptions of the same relationship</td>
<td>PFA</td>
<td>-0.11</td>
</tr>
<tr>
<td>6M17_1</td>
<td>Recognise size of angles</td>
<td>S&amp;S</td>
<td>-0.71</td>
</tr>
<tr>
<td>6M17_2</td>
<td>Recognise size of angles</td>
<td>S&amp;S</td>
<td>-1.24</td>
</tr>
<tr>
<td>6M18</td>
<td>Different descriptions of the same relationship</td>
<td>PFA</td>
<td>-0.35</td>
</tr>
<tr>
<td>6M19</td>
<td>Compare objects - 3-D objects in terms of shapes</td>
<td>PFA</td>
<td>-0.34</td>
</tr>
<tr>
<td>6M20</td>
<td>Draw lines of symmetry</td>
<td>S&amp;S</td>
<td>3.89</td>
</tr>
<tr>
<td>6M21</td>
<td>Draw reduction</td>
<td>S&amp;S</td>
<td>0.35</td>
</tr>
<tr>
<td>6M22_1</td>
<td>Read time zone map</td>
<td>ME</td>
<td>2.16</td>
</tr>
<tr>
<td>6M22_2</td>
<td>Read time zone map</td>
<td>ME</td>
<td>2.18</td>
</tr>
<tr>
<td>6M23_1</td>
<td>Convert units</td>
<td>ME</td>
<td>0.39</td>
</tr>
<tr>
<td>6M23_2</td>
<td>Convert units</td>
<td>ME</td>
<td>1.92</td>
</tr>
<tr>
<td>6M24</td>
<td>Solve problems in context using mass</td>
<td>ME</td>
<td>-0.22</td>
</tr>
<tr>
<td>6M25</td>
<td>Practical measuring of temperature</td>
<td>ME</td>
<td>2</td>
</tr>
<tr>
<td>6M26</td>
<td>Problems related to Length-perimeter</td>
<td>ME</td>
<td>1.54</td>
</tr>
<tr>
<td>6M27_1</td>
<td>Interpret chart</td>
<td>DH</td>
<td>-1.87</td>
</tr>
<tr>
<td>6M27_2</td>
<td>Interpret chart</td>
<td>DH</td>
<td>0.91</td>
</tr>
<tr>
<td>6M28</td>
<td>Find mode</td>
<td>DH</td>
<td></td>
</tr>
<tr>
<td>6M29</td>
<td>Recognise and name 2-D SHAPE</td>
<td>S&amp;S</td>
<td>-1.47</td>
</tr>
<tr>
<td>Item ID&lt;sup&gt;15&lt;/sup&gt;</td>
<td>Item description</td>
<td>Content Area (Abbr.)</td>
<td>Ratings (in logits) by individual Panel Members</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Terence Partly Achieved</td>
</tr>
<tr>
<td>6M30_1</td>
<td>Interpret chart</td>
<td>DH</td>
<td>-0.11</td>
</tr>
<tr>
<td>6M30_2</td>
<td>Interpret chart</td>
<td>DH</td>
<td>-0.03</td>
</tr>
<tr>
<td>Mean (Individual)</td>
<td>Whole Test</td>
<td></td>
<td>-0.86 0.21 0.39 -0.59 0.41 0.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level</th>
<th>Partly Achieved</th>
<th>Achieved</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (Group)</td>
<td>-0.58</td>
<td>0.00</td>
<td>0.27</td>
</tr>
<tr>
<td>1 Criterion Point</td>
<td>-0.58</td>
<td>0.00</td>
<td>0.27</td>
</tr>
<tr>
<td>2 Mastery (A)</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>3 Confidence Interval</td>
<td>1.96</td>
<td>1.96</td>
<td>1.96</td>
</tr>
<tr>
<td>4 RMSE/SEM</td>
<td>0.27</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>5 (Confidence Level) x SEM</td>
<td>0.53</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>6 Final Standard (1+2 - 5)</td>
<td>-0.26</td>
<td>0.32</td>
<td>0.59</td>
</tr>
</tbody>
</table>
Appendix J: Relationship between “range of error” and cutscores in OBM

Example of relationship between "Range of error" and OBM cutscores

"Range of error"
Appendix K: MS-DOS commands executed in data simulation

“rem - replace 2 with the number of text lines to skip before the text line you want in a Winsteps output file

@echo off
setlocal EnableDelayedExpansion
if exist result.csv del result.csv
for %%f in (*.txt) do (  
    echo %%f
    set i=a
    for /F "skip=2 delims=" %%l in (%%f) do (  
        if "%i!" == "a" echo %%%f, %%l >> result.csv & set i=b
    )
)
notepad result.csv”