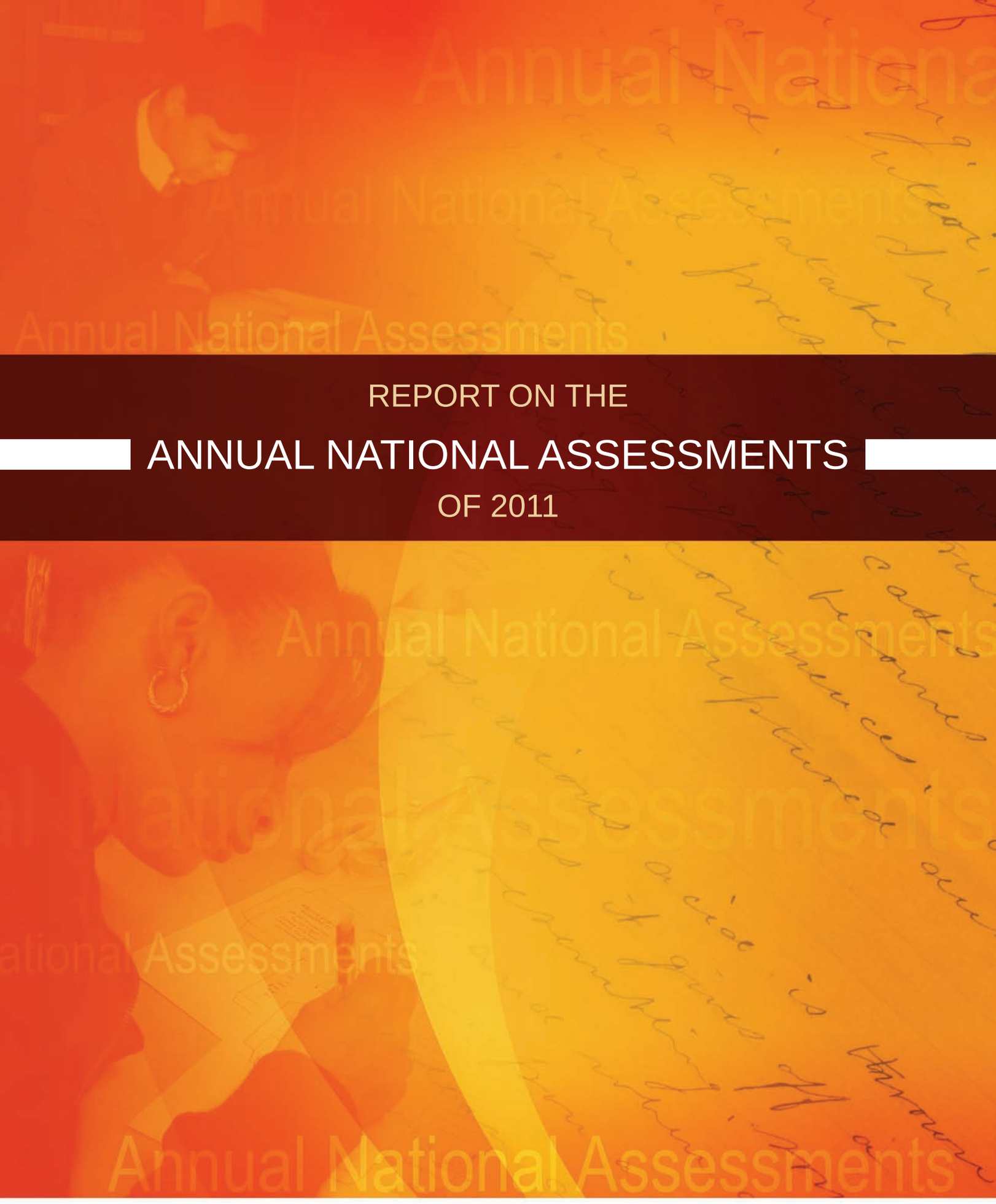


REPORT ON THE ANNUAL NATIONAL ASSESSMENTS OF 2011

basic education
Department:
Basic Education
REPUBLIC OF SOUTH AFRICA

REPORT ON THE

ANNUAL NATIONAL ASSESSMENTS

OF 2011

Grades 1-6

Published by the Department of Basic Education

222 Struben Street
Private Bag X895, Pretoria, 0001
Telephone: 012 357 3000 Fax: 012 323 0601
Website: http://www.education.gov.za

© Department of Basic Education, 2010

TABLE OF CONTENTS

FOREWORD BY MINISTER 4

EXECUTIVE SUMMARY 5

1 INTRODUCTION 8

2 BACKGROUND 8

3 THE PURPOSE OF THE ANNUAL NATIONAL ASSESSMENTS PROGRAMME 10

4 THE STEPS INVOLVED IN THE 2011 WAVE OF ANA 11

5 LIMITATIONS OF THE DATA 17

6 ANALYSIS OF THE 2011 DATA 18

7 THE WAY FORWARD FOR ANA BEYOND 2011 35

8 CONCLUSION 36

FOREWORD BY MINISTER

The improvement of the quality and levels of educational outcomes in the schooling system is a top priority of both Government and the Department of Basic Education (DBE). The extent to which these outcomes are achieved will be monitored through the administration of the Annual National Assessment (ANA).

The ANA written in February 2011 involving almost six million learners in primary schools throughout South Africa represents one of the most significant proactive interventions by Government to strengthen the foundational skills of Literacy and Numeracy among South African learners.

This important intervention forms one of the key strategies that the Department has put into place to annually measure progress on learner achievement towards the 2014 target of 60% achievement rate articulated in the *Action Plan to 2014: Towards the Realisation of Schooling 2025*.

ANA is expected to have four key effects on schools: to expose teachers to better assessment practices, make it easier for districts to identify schools in most need of assistance, encourage schools to celebrate outstanding performance and empower parents with important information about their children's performance.

The ANA 2011 Report provides the first national baseline to benchmark annual targets and achievement towards realising the desired 60% threshold of learners mastering the minimum Literacy and Numeracy competencies by the end of Grade 3, 6 and 9 respectively.

The Report includes the results of learners in Grades 3 and 6 which were independently moderated by the Human Sciences Research Council (HSRC). From 2012 onwards, the results of Grade 9 learners will also be reported.

The release of the Report on ANA performance is also an opportunity to highlight: the positive impact of the Foundations for Learning Programme introduced in 2008 on learner achievement; the massive unprecedented intervention by Government in the development and distribution of quality learning and teaching support material through the Workbook programme; and other key deliverables already underway including the strengthening of the curriculum (CAPS) and improving infrastructural support (ASIDI).

It is critical that the ANA data be effectively utilised at all levels of the system to sustain Government's solid electoral mandate and the hopes and aspirations of all South Africans in the medium to long term and to demonstrate that measurable delivery is taking place in the basic education sector.

With the release of this Report I invite all education stakeholders and the broader South African public to view the ANA results with a sense of ownership and involvement characteristic of our *Hands Up 4 Education Campaign* to support the projects, programmes and efforts of the Department in our mission to deliver quality basic education.

Thank you for your committed and dedicated support.

MRS ANGELINA MATSIE MOTSHEKGA, MP
MINISTER OF BASIC EDUCATION
28 JUNE 2011

EXECUTIVE SUMMARY

2011 is likely to be considered a watershed year in the struggle to ensure that the poor in South Africa enjoy quality basic education. This year has seen the implementation of two significant national interventions: the introduction of standardised national workbooks for Grades 1 to 6 aimed at improving classroom practices and the first fully-fledged application of the Annual National Assessments (ANA) programme with a focus on learning in Grades 1 to 6. This report deals with the second of these two interventions. Specifically, the report describes the successes and challenges experienced in the implementation of the 2011 wave of ANA and provides key findings emerging from the data collected. Whilst this report provides important information that must inform the focus of current actions by the education departments, school principals, teachers, parents and other stakeholders, this report constitutes only a part of the overall 2011 ANA process, which includes additional reports and activities emerging from the data that have been collected and the experiences gained.

The 2011 wave of ANA is the outcome of many years of capacity building and learning in the area of standardised assessments. ANA is far more ambitious and is designed to have a much greater impact than its predecessor, the Systemic Evaluation programme, run in 2001, 2004 and 2007 and involving in each run just one grade and a sample of between 35 000 and 55 000 learners. In contrast, ANA 2011 involved the testing of all learners in public schools in Grades 2 to 7 during February 2011, the focus being on the levels of learner performance in the previous year, in other words in Grades 1 to 6. This means that almost six million learners were tested. In 2008 and 2009 trial runs of ANA were conducted, largely with a focus on exposing teachers to better assessment practices. However, 2011 is the first year in which ANA produced sufficiently standardised data in order to allow for the analysis provided in this report and the generation of tools that will enable provinces and districts to target the right support to schools at the Foundation and Intermediate Phases (Grades 1 to 6) in a more effective manner.

ANA 2011 moreover draws from experiences in a number of international assessment programmes in which South Africa has actively participated in during the last decade. These include the regional Southern and East Africa Consortium for Monitoring Education Quality (SACMEQ) programme and the global Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and Science Studies (TIMSS) programmes.

ANA 2011 involved both ‘universal ANA’ and ‘verification ANA’. In universal ANA all learners in Grades 2 to 7 were tested in both languages and mathematics. Verification ANA involved applying more rigorous procedures to a sample of around 1 800 schools offering Grades 3 or 6 in order to verify results emerging from universal ANA. Specifically, in verification ANA external controls in the test administration process were more rigorous and test scripts from each school were re-marked after the initial marking by teachers.

The purpose of ANA is to make a decisive contribution towards better learning in schools. Under-performance in schools, especially schools serving the poorest communities, is a widely acknowledged problem. Clearly, ANA cannot bring about improvements on its own and should be seen as part of the wider range of interventions undertaken by Government to promote quality schooling. Part of the purpose of ANA is to provide the necessary information to planners, from the Minister all the way to teachers who need to plan their work in the classroom. At the national level, ANA is a vital instrument intended to measure progress towards the targets set by President Zuma in his 2009 State of the Nation address. These targets state that by 2014, 60% of learners in Grades 3, 6 and 9 should perform at an acceptable level in languages and mathematics. However, the information obtained

from ANA is needed for many other purposes at the national level. For instance, it is needed to diagnose in which specific areas teachers need most support and how the learning materials used by learners need to be improved.

International and local experiences point to the importance of ensuring that programmes such as ANA are viewed not only as measurement activities, but also as programmes that encourage action that will lead to better practices. *Action Plan to 2014: Towards the realisation of Schooling 2025,* released by the Minister of Basic Education in 2010, highlights four such areas of action:

- ANA should encourage teachers to assess learners using appropriate standards and methods. This has been a focus of ANA since the trial runs of 2008 and 2009. Evidence indicates that ANA has indeed brought about better assessment practices in the classroom, partly by encouraging district offices and provincial departments to review their own initiatives aimed at supporting teachers in this area.

- ANA should encourage better targeting of support to schools. District offices, which have been integrally involved in the 2008 and 2009 trial runs, have used ANA results to produce a better picture of what support to provide to which schools. The 2011 ANA data allow the Department of Basic Education (DBE) to assist districts and provinces in a more direct way in this area. Specifically, the 2011 data will be used to produce standard reports for districts that will encourage a more effective approach to the targeting of schools for support purposes.

- ANA should encourage the celebration of success in schools. By providing schools with a clearer picture of how well they perform in comparison to schools facing similar socio-economic challenges, schools that perform well will know when this is the case and schools which do not will have a clearer idea of what is possible and who they could learn from. Government does not support the use of ANA for the purposes of 'naming and shaming' those who do not perform well. At the same time, good performance should be recognised and lauded.

- ANA should encourage greater parent involvement in improving the learning process. During 2011 some schools have used ANA as an opportunity to get parents more involved in academic improvement. Specifically, ANA can provide parents on the School Governing Body, as well as parents in general, with a better picture of the grades and subjects where special attention is needed. This can assist both efforts in the school and efforts in the home aimed at ensuring that learning occurs as it should.

The data analysis provided in this report is based on the data collected from the approximately 1 800 schools forming part of verification ANA. Key findings discussed in this report include the following:

- Above all, the quality of basic education is still well below what it should be. The percentage of learners reaching at least a 'partially achieved' level of performance varies from 30% to 47%, depending on the grade and subject considered. The percentage of learners reaching the 'achieved' level of performance varies from 12% to 31%. Even the best provincial figure in this regard, 46% for Grade 3 literacy in Western Cape, is well below what can be considered acceptable. These figures reflect the magnitude of the challenge still facing the sector. However, they also reflect the high standards that we have set for ourselves. The figures are not very different from those of other countries with similar assessment systems and similar aspirations.

- Contrary to the expectations of many, the standards that teachers set when they mark are *on average* the correct ones, meaning that provincial and national results based on marks given by teachers do not differ greatly from results based on external marking. However, there is a significant minority of teachers who give marks that are either too high or too low. This emphasises the need to strengthen support to teachers in the use of appropriate standards in assessing learners.

- The analysis confirms that the greatest need for support lies in quintiles 1 and 2, quintiles which cover largely rural and the poorest communities. On the positive side, in all provinces there are schools within these quintiles which can be considered to be showing promise and which can provide guidance both to other schools and to district officials in understanding what practices contribute to better teaching and learning.

The results provided in this report do not point towards the presence of the critical upward turn needed to attain the 60% targets referred to above. There are however some indications that efforts that have gone into supporting schools are succeeding. For instance, recent provincial efforts in Eastern Cape and KwaZulu-Natal to strengthen teacher capacity to assess learners are likely to explain some of the apparent improvements in these provinces. However, as explained in the report, it would be dangerous to read too much into the differences between the 2011 ANA results and results from earlier assessments, or even the differences between provinces in 2011 ANA. Importantly, this report focuses on aggregate trends only and does not include an analysis of trends at the level of individual test items, or test questions. The DBE is currently conducting this latter analysis and the results from this will provide an indication of whether upward or downward trends over time can be observed with respect to specific skills, such as multiplication or fractions in the case of mathematics.

Though the challenges for the schooling system remain great, the 2011 wave of ANA provides a basis for optimism. Both the process of 2011 ANA and the information obtained from this process represent a basis for improvement that did not exist previously. Schools and the education departments have gained important experiences in better assessment and, through this, a better focus on what must improve. The unprecedented step of providing all Grades 1 to 6 learners with national workbooks in 2011 has, according to preliminary reports, shifted classroom practices in the right direction. The 2012 wave of ANA, to be conducted early in the 2012 school year, will serve as a critical instrument with which to monitor the degree to which national workbooks and other interventions, such as the streamlining of the national curriculum, have had an impact on learning. The 2011 wave of ANA has provided a wealth of experience in how to conduct a programme of this nature in a way that contributes to quality education. Lessons learnt in 2011, which are discussed in this report, will inform the implementation of the 2012 wave of ANA.

1 INTRODUCTION

This report describes the 2011 Annual National Assessments (ANA) process and discusses the pointers provided by this programme that should guide the actions of those involved in improving learning and teaching in the Foundation and Intermediate Phases (Grades 1 to 6) of South Africa's schools. The analysis in the report should be regarded as initial analysis aimed at identifying broad patterns. Beyond this report, more detailed and specialised reports will emerge from the 2011 wave of ANA focussing on, for instance, learner performance patterns within individual districts and trends with respect to specific competencies described in the curriculum.

Section 2 of the report briefly discusses the nature of the educational quality challenge in our schools as well as the recent history of national assessments in South Africa. Section 3 describes the purpose of ANA and specifically how ANA is expected to result in improvements in the quality of teaching and learning. Section 4 outlines the planning and implementation steps of the 2011 wave of ANA, including problems encountered and how these should inform future waves of the programme. Steps beyond this report that form part of the 2011 wave of ANA are also described. Section 5 sums up the key limitations of the data analysed for this report. Section 6 provides the results of the analysis of the 2011 ANA data. Section 7 sums up what the lessons emerging from ANA 2011 are and how ANA can be improved in coming years. Section 8 provides a conclusion.

The Foundation Phase (Grades 1 to 3) subjects assessed in ANA are literacy and numeracy. In the Intermediate Phase (Grades 4 to 6) the subjects are languages and mathematics. For the sake of brevity, in this report the term 'languages' may refer to both literacy and languages. Similarly, 'mathematics' may refer to both numeracy and mathematics.

2 BACKGROUND

Of Government's twelve development priorities announced in 2010, the first priority is improving the quality of basic education. The Minister of Basic Education's Delivery Agreement, concluded with the President in 2010, says the following about the importance of improving the quality of schooling[1]:

It is widely recognised that the country's schooling system performs well below its potential and that improving basic education outcomes is a prerequisite for the country's long-range development goals. Hence the 2008 election manifesto refers to the need for a major renewal of South Africa's schools. In the 2010 State of the Nation Address, the President referred to the vital role of the education system in improving productivity and competitiveness in the economy. The President also underlined that 'our education targets are simple but critical'. The need is fairly straightforward as far as the basic education sector is concerned. Our children and youths need to be better prepared by their schools to read, write, think critically and solve numerical problems. These skills are the foundations on which further studies, job satisfaction, productivity and meaningful citizenship are based.

1 This is from the complete version of the Outcome 1 Delivery Agreement, available on the DBE website.

The first three of 27 goals in *Action Plan to 2014: Towards the realisation of Schooling 2025*, published by the Department of Basic Education (DBE) in 2010[2], refer to the need to improve learner performance in languages and mathematics at the Grades 3, 6 and 9 levels, in other words at the final grades of three curriculum phases. Whilst quality education is far more than just adequate competencies in languages and mathematics, it is common knowledge that without competencies in these two subjects, quality education in a broad sense is unachievable. The 'gateway' nature of these subjects lies behind the strong focus on the two subjects not just in South Africa, but around the world. In his 2009 State of the Nation address, President Zuma set the target that by 2014, 60% of learners in Grades 3, 6 and 9 should perform at an acceptable level in languages and mathematics.

There is a strong tradition in South Africa of focussing on the Grade 12 examination results and improvements in these results. During the last decade Government and society generally have moreover placed a growing emphasis on monitoring learner performance at the lower grades, in particular within the Foundation and Intermediate Phases. The Systemic Evaluation programme involved sampling between 35 000 and 55 000 learners in 2001 (Grade 3), 2004 (Grade 6) and again in 2007 (Grade 3). In each wave of the Systemic Evaluation, learners wrote languages and mathematics tests, which were externally marked. Moreover, school principals, teachers, learners and parents filled in background questionnaires which provided important information that would contribute towards understanding the factors affecting performance. The Systemic Evaluation programme followed internationally established principles for these kinds of programmes. The results could therefore be considered highly representative of the performance of learners across all public schools in the grades concerned. The Systemic Evaluation played an important role in clarifying where the areas of weakness were and the data collected have been used to inform a variety of policy changes, including recent changes to the curriculum. At the same time, the Systemic Evaluation tended to be regarded as a 'backroom' exercise which was not well understood by schools. There was a need for a more comprehensive national learner assessment programme in the early grades that would have a more direct impact on practices in schools.

In 2008, Government introduced the Foundations for Learning Campaign[3], which focussed on the Foundation and Intermediate Phases and included clearer specifications of the materials learners should have access to and the time needed for different learning activities in a week. The campaign involved a number of teacher training and materials distribution initiatives and trial runs in 2008 and 2009 of a new national assessment system known as Annual National Assessments. This new system was to cover all schools and would allow teachers themselves to be involved in the marking of tests in order to facilitate teacher training and the exposure of teachers to better assessment and marking practices. In 2008 and 2009 tests were distributed across the country and most schools participated in the programme. The decision was taken to conduct a major wave of ANA at the beginning of the 2011 school year where participation by public schools would be compulsory and a sample of schools would be subject to more rigorous controls in the test administration and marking stages in order to assess the validity of results obtained from the school system as a whole. An important aspect of the 2011 wave of ANA was that data from the testing should be collected into a national database, which would be used to inform planning and to produce reports that provinces and districts could use to target schools for different kinds of support.

2 Government Notice 752 of 2010.

3 Government Notice 306 of 2008.

South Africa's active participation in a number of international testing programmes over the last decade has provided important experiences that have informed the design of the Systemic Evaluation and ANA programmes. Specifically, South Africa has participated, or will be participating, in the following international programmes:

- Southern and East Africa Consortium for Monitoring Education Quality (SACMEQ) in 2001 and 2007, with a focus on languages and mathematics in Grade 6. The next wave of SACMEQ is currently being planned, though a year has not been set.
- Trends in International Mathematics and Science Studies (TIMSS) in 2003, with a focus on mathematics and science in Grade 8. Testing for the 2011 wave of TIMSS is occurring in South Africa in the current year.
- Progress in International Reading Literacy Study (PIRLS) in 2006, with a focus on literacy in Grade 5. Testing for the 2011 wave of PIRLS is occurring in South Africa in the current year.

Results from international assessment programmes have played an important role in confirming the trends emerging from national assessments, in particular the Systemic Evaluation programme.

3 THE PURPOSE OF THE ANNUAL NATIONAL ASSESSMENTS PROGRAMME

The main purpose of the Annual National Assessments (ANA) programme is to make a decisive contribution towards better learning in schools. Four key areas of impact at the school and district levels have been identified:

- **ANA should encourage teachers to assess learners using appropriate standards and methods.** This has been a focus of ANA since the trial runs of 2008 and 2009. Evidence indicates that ANA has indeed brought about better assessment practices in the classroom, by exposing teachers to well constructed tests and marking memoranda but also by encouraging district offices and provincial departments to review their own initiatives aimed at supporting teachers in this area. For instance, in Eastern Cape newly developed 'common tests' in languages and mathematics for Grades 3, 6 and 9 were conducted across the province in the middle and the end of the 2010 school year. Having ANA at the start of the school year provides teachers with an additional indication of where their own learners will require extra support during the course of the year.
- **ANA should encourage better targeting of support to schools.** District offices, which have been integrally involved in the 2008 and 2009 trial runs, have used ANA results to produce a better picture of what support to provide to which schools. Before ANA, in most provinces there was no reliable way of knowing which schools performed better than others because assessments were not sufficiently standardised across schools. ANA will help to ensure that schools receive the support they should receive and that schools are not obliged to participate in development activities which they have no need for. Moreover, ANA makes it easier for districts to identify which schools perform well and therefore which schools the district as a whole can learn from. The 2011 ANA data allow the Department of Basic Education (DBE) to assist districts and provinces in a more direct way in this area. Specifically, the 2011 data will be used to produce standard reports for districts that will encourage a more effective approach to the targeting of schools for support purposes.

- **ANA should encourage the celebration of success in schools.** By providing schools with a clearer picture of how well they perform in comparison to schools facing similar socio-economic challenges, schools that perform well will know when this is the case and schools which do not will have a clearer idea of what is possible and who they could learn from. Government does not support the use of ANA for the purposes of 'naming and shaming' those who do not perform well. At the same time, good performance should be recognised and lauded. ANA should become a key indicator in the school development plan and in fact in some districts ANA targets for future years are being negotiated between support personnel from district offices and schools.

- **ANA should encourage greater parent involvement in improving the learning process.** During 2011 some schools have used ANA as an opportunity to get parents more involved in academic improvement. Specifically, ANA can provide parents on the School Governing Body, as well as parents in general, with a better picture of the grades and subjects where special attention is needed. This can assist both efforts in the school and efforts in the home aimed at ensuring that learning occurs as it should.

The collection of data from the ANA process, including marks, into a national database is an important aspect of ANA. This allows the Department of Basic Education (DBE) to analyse important patterns, for instance, how frequently schools perform poorly in both languages and mathematics, as opposed to just one of these subjects, or how often learners within the same school provide the same incorrect responses to the same question, something which could point towards problems in the way teachers teach. National reports on these matters need to guide those working on teacher development, those developing workbooks and textbooks and those involved in providing management support to schools. Moreover, the national database allows for the generation of standard reports for provinces and districts which can inform the targeting of support to schools.

4 THE STEPS INVOLVED IN THE 2011 WAVE OF ANA

This section describes the planning and implementation steps of the 2011 wave of ANA, including steps that will be taken following this report.

4.1 Initial planning of universal ANA and verification ANA

Early in 2010 the leadership of the nine provincial education departments and the national department agreed to proceed with the testing of all Grades 1 to 6 learners at the end of the 2010 year as part of 'universal ANA' and at the same time to apply more rigorous controls to a sample of schools as part of 'verification ANA'. The timing of the testing was subsequently shifted to early 2011, largely to ensure that district officials would have sufficient time available for the management and monitoring of ANA (at the end of the school year, district officials are deeply involved in the Grade 12 examinations processes). This meant that Grades 2 to 7 learners would be tested with respect to what they should have learnt by the end of the previous year. Thus a learner in Grade 7 would be tested on what he or she should have learnt by the end of Grade 6. The tests were thus applicable to what had to be learnt in Grades 1 to 6. Each learner would sit for two tests, one in languages and the other in mathematics. The literacy tests covering the three Foundation Phase grades tested competencies

in the language taught as the home language in the school, meaning different versions of the test in all eleven official languages were developed. The languages tests covering Grades 4 to 6 were in English and Afrikaans only. Similarly, the Grades 1 to 3 numeracy tests had versions in all eleven official languages, whilst the Grades 4 to 6 mathematics tests were available only in English or Afrikaans.

Verification ANA would focus on verifying universal ANA results at the Grades 3 and 6 levels (verification ANA also focussed on Grade 9, as explained in section 4.11 below). Verification ANA introduced exceptional controls with respect to two aspects of ANA, namely the administration of tests and the marking of tests. An independent agency, the Human Sciences Research Council (HSRC), was appointed to manage verification ANA.

It was decided to treat learners repeating their current grade at the start of 2011 no differently to learners who were not repeating. This decision was taken largely to avoid making repeaters feel marginalised. This meant that, for instance, a learner in Grade 4 at the start of 2011 who was repeating Grade 4, would write the tests focussing on competencies that should have been acquired by the end of Grade 3.

4.2 Selection of verification ANA schools

For verification ANA, 900 schools offering Grade 3 were selected and a further 900 schools offering Grade 6 were selected. Each set of 900 schools was divided equally across the nine provinces, meaning there were 100 schools per province and per grade. The reason why an equal number of schools per province were chosen, despite the fact that the school populations of provinces vary greatly, is that it was important for verification ANA to produce provincial average scores that were from a statistical point of view, equally reliable. A two stage sampling design was employed. The first stage entailed sampling schools using province, district and quintile as stratification variables. The second stage involved the sampling of 25 test scripts per school and per subject on the basis of random selection rules determined by the Department of Basic Education (DBE). The test scripts analysed for verification ANA were those submitted by learners enrolled in Grade 4 (most of whom had been in Grade 3 in 2010) and learners enrolled in Grade 7 (most of whom had been in Grade 6 in 2010).

4.3 The design of the assessment frameworks and tests

In line with standard practices, before tests were developed, assessment frameworks outlining such details as the outcomes and assessment standards to be tested was drawn up. Moreover blueprints for each test specifying, for instance, how many test items (or questions) to have at varying levels of difficulty, were produced. The basis for this work was the National Curriculum Statement (NCS), as well as the achievement milestones that had been established as part of the Foundations for Learning Campaign.

In both the languages and mathematics tests a design was chosen whereby 20% of items could be considered easy, 60% moderate and 20% difficult. This was to enable a moderately performing learner to succeed without denying outstanding learners the opportunity to demonstrate their performance. Questions included in the languages tests were spread across the three cognitive areas: knowledge of basic language concepts, comprehension and application of language concepts.

In mathematics cognitive areas covered were knowledge of basic mathematical concepts (20% of items), application of concepts (60%) and non-routine problem solving (20%).

To obtain a detailed insight into the kinds and levels of competency displayed by learners, the DBE analysed a random sample of marked scripts from the 2009 trial run of ANA from almost 1 000 schools spread across all provinces. This analysis informed the design of the assessment framework, blueprints and actual tests used in 2011 ANA.

Draft tests in English were first developed. This work was undertaken by departmental officials from the national and provincial levels and external experts. Thereafter piloting of the draft tests occurred in order to establish whether the test items made sense to learners, whether the assumptions around the difficulty levels of items were correct and whether each test as a whole was appropriate as a measure of the intended outcomes. In the piloting, two versions of each test per grade and subject were used in order to create sufficient opportunities to discard items. Piloting occurred in eight schools. Apart from testing learners, teachers from the schools were interviewed to obtain their views on the relevance of the items in the light of their experiences and teaching practices. The test responses of learners were captured and analysed, partly through the calculation of difficulty indices (reflecting the proportion of learners who responded correctly) and discrimination indices (reflecting the ability of test items to discriminate between learners whose overall performance was better and learners whose overall performance was worse).

Final tests in English were produced using items that appeared appropriate and in line with the relevant blueprints. This process included a final round of verification to ensure that tests met a number of key criteria. Experts at the national and provincial levels checked to see, amongst other things, that instructions were clear and simple to read, that there was no bias towards any religion, culture or gender, and that the assessment framework had been properly followed.

Tests in English were then versioned into the remaining ten official languages (in the case of the Grades 1 to 3 tests) and into Afrikaans (in the case of the Grades 4 to 6 tests). Versioning is different from translating in the sense that versioning takes into account a wider range of socio-linguistic factors and thus enhances the comparability of the tests across different languages. Finally, all tests were proofread and suitable layouts developed.

At the Foundation Phase, a total of 66 tests were developed in literacy and numeracy. At the Intermediate Phase, a total of 12 tests were developed in languages and mathematics. Thus altogether a total of 78 tests were developed for Grades 1 to 6.

Tests were adapted for blind, short-sighted and deaf learners, for instance through translation to Braille and the use of larger font sizes.

4.4 Training of school and district personnel

Schools were informed about ANA and every effort was made to ensure that schools were ready. The DBE in collaboration with the HSRC trained provincial ANA coordinators who then cascaded the same training to district officials and principals responsible for administering ANA. DBE officials monitored the training in a sample of districts. The training included how the ANA process should be managed, from the distribution of

tests to schools, up to the reporting of ANA results to parents. School principals were informed about minimum requirements in terms of facilities such as desks and chairs. Since learners in different grades wrote the test on different days, it was possible to rotate the use of classrooms in such a way that the minimum requirements would be met.

4.5 Distribution of universal ANA materials to schools

The DBE appointed service providers to print, package and distribute copies of all the required tests to district offices. The target population was 5 841 562 learners from 19 619 schools. Data collected previously through the Annual Survey of Schools were used to ensure that packages for schools included tests in the correct languages and the correct numbers. School principals collected their packages of tests and marking memoranda from their district office on each test day, if possible, or less frequently if the distance between the district office and the school was great.

A total of 2 076 learners from 20 special schools were provided with adapted tests.

4.6 Test administration

A comprehensive set of instructions had been sent to every school specifying how test administration should proceed. These instructions indicated that testing should occur on the four days 8 to 11 February 2011. However, due to various logistical problems such as tests not being available on time, this period had to be extended in parts of the country to 4 March 2011. Invigilation of tests had to be undertaken by teachers who were not the teachers of the tested learners in either 2010 or 2011. In most schools, the school principal was ultimately responsible for ensuring that test administration proceeded as it should. In this sense, the arrangement was similar to the arrangement applicable to the Grade 12 examinations. However, in the case of verification ANA schools, district officials who had been trained by the HSRC were present in the school on the test days and were ultimately responsible for assuring the rigour of the process. Items in the test venue such as wallcharts that could assist learners in responding to test questions had to be removed beforehand.

A variety of monitoring of the test administration process by people external to the school occurred. This varied from district to district, depending on the availability of staff and transport in the district office and the perceived need for external moderation. Schools comprising the verification ANA set of schools were subject to more stringent external monitoring during the test administration phase. In general, the reports from those involved in external monitoring were satisfied with the procedures followed within schools.

In all test booklets, the grade, language and subject were clearly indicated on the front cover. Learners responded to questions by writing in the printed test booklet. At the start of each test was at least one practice question, not counted for the final result, that invigilators guided learners through in order to ensure that learners understood the methodology of the test. Invigilators could respond to learners' questions about the methodology before the start of the test itself. For the Grades 4 to 6 tests, however, invigilators had to use only English or Afrikaans, given that these were the languages used in the tests. For the Grades 1 and 2 tests (written by Grades 2 and 3 learners), invigilators had to guide learners through the test by reading each

question out aloud before allowing the learner to write down a response. This approach was considered suitable for these grades given the low level of exposure of learners to tests of this nature. However, for the Grades 3 to 6 tests, learners had to understand the questions in the test by reading the questions themselves, without any assistance from the invigilator.

One of the advantages of having the ANA at the start and not the end of the year was that learners were more relaxed about the tests than they might have been at the end of the year, when the school conducts summative assessments which determine whether learners can be promoted to the following grade.

4.7 Marking of tests by teachers

Marking was done by the teacher teaching the subject to the learner in the current year. For instance, the Grade 6 mathematics tests, written by Grade 7 learners, would be marked by the teacher teaching the Grade 7 learners mathematics in 2011. Marking occurred in accordance with national marking memoranda that had been distributed to all schools. School principals, in conjunction with the other members of the School Management Team (SMT), were responsible for quality assuring the marking process through moderation procedures explained in a standard moderation guide provided to all schools.

The approach followed in the national marking memoranda were those that teachers would be have been exposed to previously in the ANA trial runs of 2008 and 2009 and in various in-service training programmes. The memoranda gave possible options for open-ended questions in the languages tests. Similarly, the memoranda for the mathematics tests provided alternative correct methods or techniques of answering open-ended questions. Moreover, in mathematics mark allocations were included for every step of the more complex questions. Teachers were requested to discuss the memoranda with their colleagues before commencement of marking at the school.

In some provinces, districts officials organised meetings with teachers to further standardise the marking of learner responses and collected samples of scripts from schools in order to conduct a district-level verification of the standard of marking separately from the national verification ANA process.

4.8 Communication of ANA results to parents

A short report, following a nationally determined format, had to be produced by schools for all parents and guardians of learners who participated in ANA 2011. This report included the marks obtained by the learner in the two tests he or she took, as well as the school average.

4.9 Collection of universal ANA results

An ambitious feature of the 2011 wave of ANA was the collection of every learner's subject mark into a national database in order to facilitate analysis of the results and the production of standard management tools for

provinces and districts. After the marking process was complete, schools had three options through which to submit results by learner to the Department. One was to fill in a standard paper-based mark schedule. The second was to enter the marks onto a standard Excel spreadsheet. The third was to capture the data onto the SA-SAMS system, where schools had this system installed on a computer. The State Information Technology Agency (SITA), which maintains the system used to capture Grade 12 examination results, created a new facility for the storage of ANA results. Provincial departments entered data into the SITA system either through the transfer of data files obtained from schools (where schools were able to perform the data capturing themselves) or by means of manual capturing off the paper mark schedules.

A variety of quality control measures were implemented. Districts and provinces followed up cases of missing or clearly incorrect information. In the case of several provinces which encountered problems or had insufficient capacity, the DBE became directly involved in the data capturing process.

4.10 Collection and re-marking of verification ANA test scripts

District offices were requested to collect test scripts from the just under 1 800 schools identified as part of that component of verification ANA focussing on Grades 3 and 6. In the case of each school, 25 test scripts per grade and per subject were randomly selected using a method specified by the DBE. Scripts were collected after marking by teachers in the school had occurred and schools had recorded the results. The overall return rate was 93%, meaning scripts were received from 93% of the schools in the sample. Whilst this is a feature that must improve in future years, the return rate for most provinces can be considered sufficiently high to avoid any substantial distortion of the provincial statistics. The possible exception is Eastern Cape, where the return rate was relatively low, at 84%.

Table 1: Percentage of successful returns in verification ANA

	% of schools
EC	84
FS	88
GP	96
KN	100
LP	92
MP	92
NC	96
NW	95
WC	97
SA	93

Scripts from the sample of schools were sent to the HSRC, which oversaw the re-marking of all scripts in a highly controlled environment. Teachers from schools considered as being good at the assessment of learners were selected to perform the re-marking. Close attention went towards ensuring that sufficient numbers of

teachers with competencies in each of the official languages were recruited, given the use of all the official languages in the tests. The training of markers and the actual marking process occurred at a central venue. The training of teachers included practice marking sessions to ensure that the marking would be sufficiently standardised and accurate. During the re-marking process, there was ongoing moderation of marks allocated by the recruited teachers.

The marks originally assigned to tests by teachers in their schools, as well the new marks allocated by teachers recruited and training by the HSRC were captured into a database established by the HSRC and which was completely separate from the SITA database. The HSRC data capturing process included the capturing of the new marks allocated to every question in each test. This will allow for important analysis of the competencies displayed by learners with respect to specific areas in the curriculum.

4.11 The Grade 9 ANA pilot

In preparation for the inclusion of Grade 9 in universal ANA in 2012, the piloting of Grade 9 tests in a sample of schools was undertaken by the DBE and the HSRC. Twelve tests were developed for the Grade 9 level in languages and mathematics (eleven languages tests corresponding to the eleven official languages as well as two mathematics tests, one in English and one in Afrikaans). The test design steps followed were similar to those of the Grades 1 to 6 tests.

The HSRC sampled 450 schools offering Grades 9 and 10 to pilot the Grade 9 tests. The procedures followed with respect to test administration and marking were similar to those applicable to schools with Grades 3 and 6 participating in verification ANA.

5 LIMITATIONS OF THE DATA

ANA 2011 has been an ambitious exercise which has exposed teachers and the nation to more standardised and rigorous assessments. It has also produced data that greatly enhance the ability of the national department, the provincial departments and districts to plan effectively. At the same time, it is important to realise that there are limitations regarding what ANA can do and what the data can tell us. Some of these limitations are inherent in this kind of programme and are likely to remain in the coming years. Other limitations should be resolved in coming years as ANA is strengthened.

Four main limitations can be identified.

Firstly, though the ANA 2011 tests were put through a standard piloting exercise and editing by experts, these tests are not perfect. Certain problems only arise after the entire assessment exercise has run its course. Though comparability with previous assessments, in particular the 2004 and 2007 Systemic Evaluation runs (which covered Grades 6 and 3 respectively) was aimed for, the comparability is not perfect with respect to the tests as a whole. However, at the level of individual test questions more accurate comparison is possible. This level of comparison is still under way and is not presented in the current report but will be made available in future. Apart from comparability problems, problems with the formulation and translation of certain questions

have been noted and underline the importance of strengthening the test design processes in future years.

Secondly, it was decided to place most of the responsibility for the correct administration of the ANA tests in the hands of school principals. This is the procedure followed in schools participating in the Grade 12 examinations. Clearly, these responsibilities are new for most school principals working in schools offering Grades 1 to 6. It proved difficult to apply exactly the same test administration procedures across all schools in ANA 2011, in particular in Grades 1 and 2, where teachers were asked to read questions to learners in order to facilitate the assessment process. This points to a need for better training of school principals in future, instructions that are in certain respects clearer and stronger external controls, in particular as far as verification ANA schools are concerned, and schools with a pattern of irregular marks in ANA 2011.

Thirdly, whilst allowing teachers themselves to mark the ANA tests was advantageous, both in terms of teacher development and insofar as on average teachers marked mostly at the appropriate level, there were problems here too (the analysis that follows provides details). A substantial minority of teachers have problems assessing correctly, though under-valuation and over-valuation of learner responses tend to cancel each other out. This means that although at a high level, for instance nationally, average results emerging from teacher-marked scripts can be considered accurate, this becomes less so the lower down the level. There is thus the possibility that individual school average results will be well above or below what they should be, due to poor marking practices by teachers.

Fourthly, verification ANA is based on a sample of schools. Sampling is widely used across the world in order to gauge the level of performance in schools. Samples have the advantage that more rigorous quality assurance becomes possible than would be possible if all schools were considered. However, results emerging from samples must always be interpreted keeping in mind that each statistic has a 'confidence interval', or a margin of error. The sizes of these confidence intervals in the case of verification ANA are explained below.

6 ANALYSIS OF THE 2011 DATA

The analysis in this section mostly used data from verification ANA of 2011. Those data covered 19 470 Grade 3 learners and 19 397 Grade 6 learners in 827 and 840 schools respectively. Grade 3 and Grade 6 data were mostly collected from different schools. Only in the case of 21 schools, mostly in Limpopo Province, were both Grade 3 and Grade 6 data collected.

The aim was to have an equal number of schools per province in order to provide similar levels of statistical reliability across all provinces. Rates of return below 100% (as explained above) meant that the total number of schools per province with data available for the analysis varied from 164 in the case of Eastern Cape to 193 in the case of Western Cape.

Schools were randomly selected in each province, with small schools with fewer than 25 learners in a grade excluded. This approach of excluding small schools is taken to improve the size of the sample and is an approach followed in, for instance, SACMEQ[4]. The sample was stratified by quintile, meaning that the spread of schools across quintiles in the sample and in the entire province was designed to be the same.

4 Southern and Eastern Africa Consortium for Monitoring Educational Quality.

In the analysis that follows, no weighting of the data within a province occurred. However, in the calculation of national statistics the total level of enrolment in Grades 3 and 6 per province during 2010 was used to ensure that larger provinces were weighted more. Unless otherwise stated, the marks analysed were those obtained from the verification ANA re-marking process.

This report presents a first picture of what ANA 2011 tells us about the state of learning in the Foundation Phase and the Intermediate Phase in schools across the country. Importantly, what is presented here should be seen as the start of a series of processes and reports in which ANA 2011, the most ambitious assessment initiative of its kind ever to be undertaken in South Africa, gets to inform people, from planners to teachers to parents, about the best ways of tackling under-performance in schools. The findings from ANA 2011 will moreover guide future improvements to the design and implementation of ANA to ensure that this programme becomes a world class assessment system.

The analysis provided below provides important pointers for improvements and confirms that many of the support initiatives currently being run by government are focussing on the right things.

Test scores converted to percentages are used in the analysis. The following table indicates what the maximum possible mark per test was before conversion to percentages.

Table 2: Maximum possible mark per test

Grade 3 literacy	30
Grade 3 numeracy	40
Grade 6 languages	50
Grade 6 mathematics	75

6.1 Provincial ANA averages in a historical context

If one examines how well provinces performed relative to each other in 2011 ANA and compares this to the provincial averages derived from earlier assessments in the 2004 to 2007 period, there is considerable consistency, in particular as far as the relative positions of Free State, Gauteng, Northern Cape and Western Cape are concerned. Eastern Cape's Grade 3 performance according to ANA 2011 is considerably higher than one would expect, given this province's performance in earlier assaessments. To some extent, the same can be said of KwaZulu-Natal. The results suggest that whilst ANA provides relatively accurate provincial pictures of learner performance with respect to at least half of the tests, there is room for more standardisation in the way ANA is implemented.

As has been indicated above in this report, verification ANA in 2011 focussed extensively on verifying the degree of accuracy in the marking practices of teachers. In the test administration phase within the school, verification ANA schools were monitored more closely than other schools. However, responsibility for a proper test administration process still rested with the school principal. In this respect, the approach was the same

as that followed in the Grade 12 examinations. As indicated previously, district offices were responsible for randomly selecting the 25 test scripts per grade and subject required for verification ANA. The approach followed in the Systemic Evaluation of allowing people who were completely external to the school randomly select learners within the school, administer the tests and leave the school with the test scripts without letting teachers participate in any way was not followed in verification ANA. This was to ensure that teachers in verification ANA schools engaged fully with the ANA programme. Whilst this approach had obvious advantages, it also reduced certain controls relative to those of the Systemic Evaluation. The patterns in the results presented below need to be interpreted in this context. Clearly a key consideration in future waves of ANA will be how best to balance the need for teacher participation with the need for standardisation within verification ANA schools.

The next table provides the average percentage scores per subject, grade and province emerging after the verification ANA re-marking process. Because the averages in the table are derived from a sample, each comes with a margin of error. We can be 95% certain that the margin of error is around 6 percentage points at the provincial level and around 2 percentage points at the national level. This means that the Free State mean of 37 in Grade 3 literacy, for instance, has a margin of error of around 34 to 40. We can thus not be certain that the Free State value for Grade 3 literacy is better than the Gauteng value as the values are too close to each other and the margins of error overlap. However, we can be certain that the Free State value is higher than the North West value, as here the gap is large and the margins of error would not overlap.

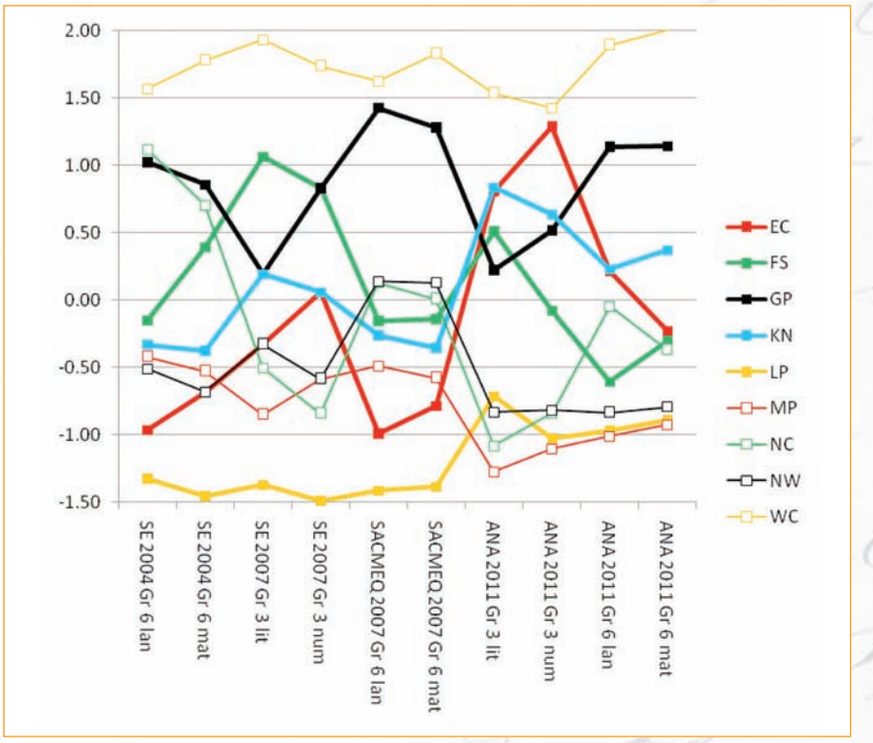
Table 3: Average percentage scores after re-marking

	Grade 3		Grade 6	
	Literacy	**Numeracy**	**Languages**	**Mathematics**
EC	39	35	29	29
FS	37	26	23	28
GP	35	30	35	37
KN	39	31	29	32
LP	30	20	21	25
MP	27	19	20	25
NC	28	21	27	28
NW	30	21	22	26
WC	43	36	40	41
SA	35	28	28	30

In the following graph, the ANA values from the previous table are converted to values that would allow comparison to results from previous assessments. Basically values are converted in such a way that the average across provinces becomes zero and the standard deviation 1,0. In some ways, performance by provinces has been consistent across different tests. For example, Western Cape has consistently performed better than other provinces, regardless of grade or subject. Gauteng, more than any other province, has taken the second position. Limpopo has, before ANA, performed at a considerably lower level than any other province. The high average scores in Grade 3 in the case of Eastern Cape in ANA suggest that different standards were applicable in these tests and in this province. It is unlikely that Eastern Cape would have seen

such a marked improvement in Grade 3 (but not Grade 6). The graph suggests that the same provisos may apply to the KwaZulu-Natal Grade 3 ANA results. At the same time, the possibility cannot be discarded that the recent introduction of new provincial systems of standardised tests in Eastern Cape and KwaZulu-Natal lie behind some of the improvement seen in the graph.

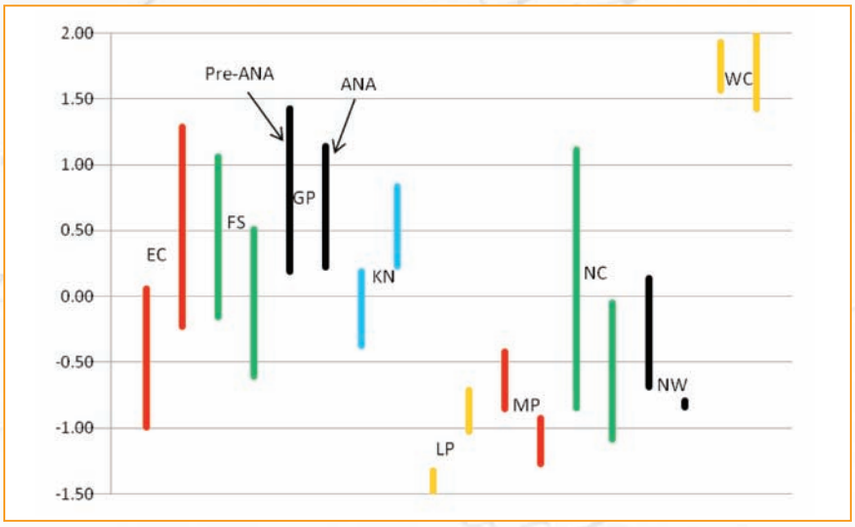
Figure 1: Relative positions of provincial performance in recent years



The above graph highlights the point that comparisons across tests must be done with much caution, even as far as comparisons across the same grade and subject are concerned. Sample-based values come with margins of error, the comparability of the tests themselves is not always perfect and test administration procedures may differ slightly from test to test.

The next graph illustrates the minimum and maximum values from the previous graph, for the pre-ANA tests and the ANA tests. The left-hand vertical bar for each province represents pre-ANA performance whilst the right-hand bar represents performance in ANA. The very large difference in the performance of Eastern Cape in the pre-ANA tests and the ANA tests is clear. For at least four provinces, Free State, Gauteng, Northern Cape and Western Cape, performance in ANA 2011 can be considered more or less consistent with performance in previous tests.

Figure 2: ANA against pre-ANA comparison



6.2 Distribution of ANA scores by province

The distribution of ANA scores by province, grade and subject emphasise the great challenge that the country faces in breaking the legacy of illiteracy and innumeracy in the first six grades of school. In most provinces and most tests, the most common percentage score achieved by learners (the 'mode') is below 20%. It is vital that current interventions aimed at accelerating improvements, such as the provision of standardised workbooks to all learners in Grades 1 to 6, should focus strongly on what learners actually learn.

The distribution of scores in the Grade 3 tests in Eastern Cape appear to confirm that national standards with respect to assessment practices need to be reinforced in this province.

The next four graphs illustrate the distribution of ANA percentage scores by province. The peak, or mode, for each curve, which are known as a kernel density curve, indicates the most common score for the test and province in question. The values at the bottom of each graph refer to the original score converted to a percentage, after the re-marking process. What stands out is that the mode is mostly below 20. The average scores seen previously, which are mostly above 20, hide the fact that scores below 20 are very common. The fact that in the two Grade 3 graphs the Eastern Cape curves should be so unlike the curves of other provinces facing similar socio-economic challenges seems to confirm a different application of the verification ANA process in this province.

Figure 3: Distribution of Grade 3 literacy scores

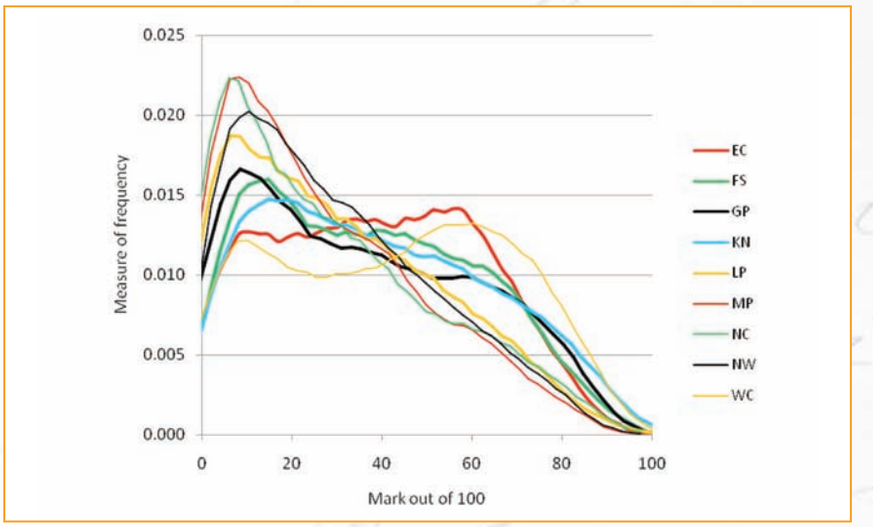


Figure 4: Distribution of Grade 3 numeracy scores

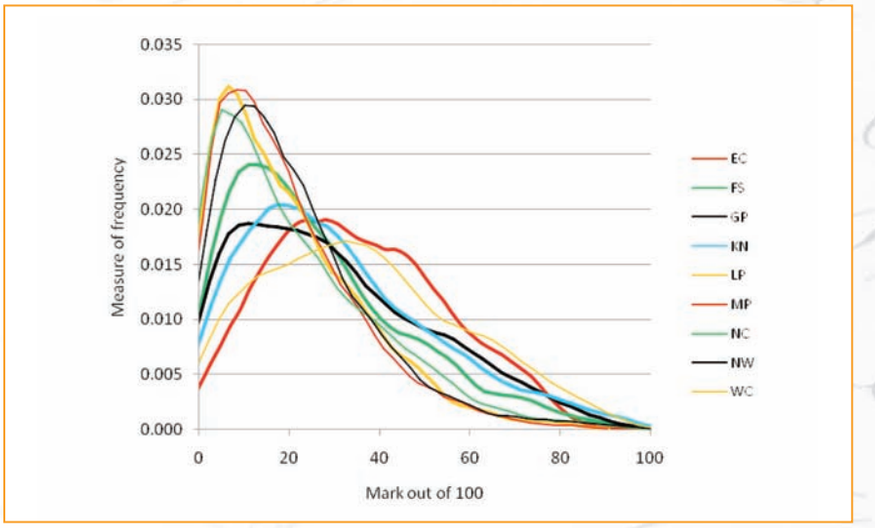


Figure 5: Distribution of Grade 6 languages scores

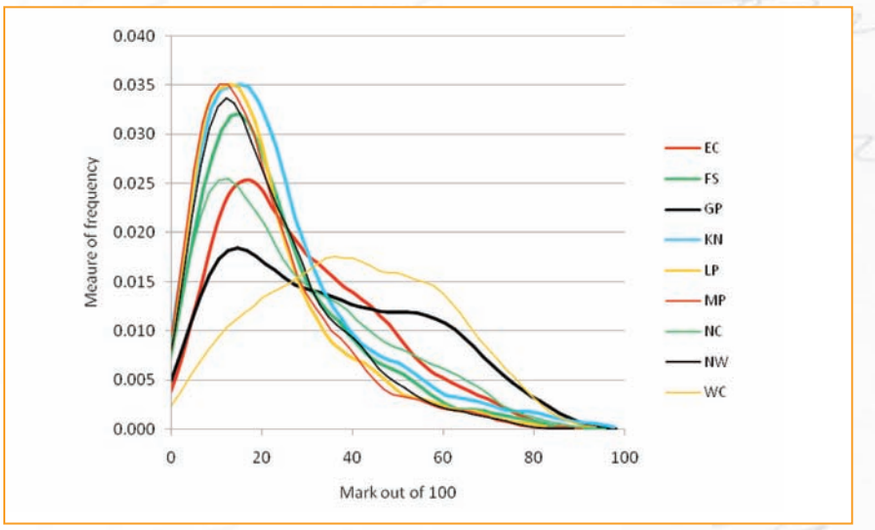
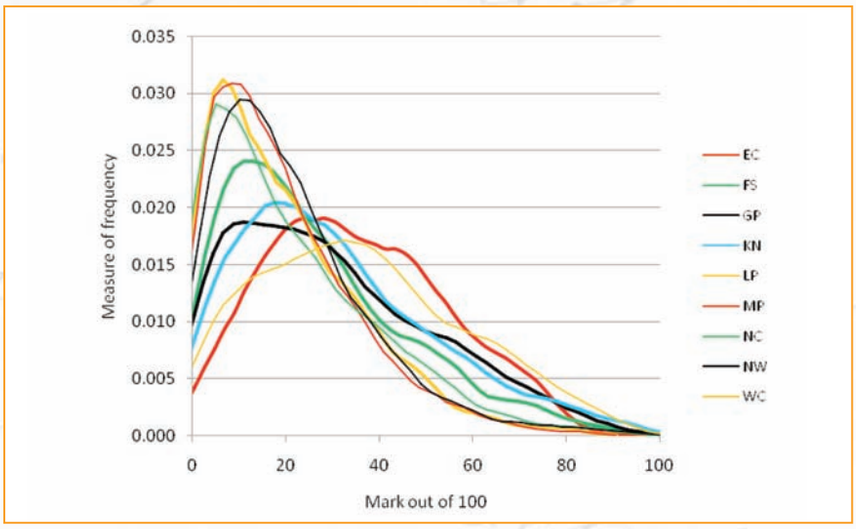


Figure 6: Distribution of Grade 6 mathematics scores



6.3 Differences between universal and verification ANA results

One of the more positive things to emerge from ANA 2011 is the finding that on average teachers are able to mark at the required standard and use the ANA marking memoranda correctly. This is especially true in the case of numeracy and mathematics. It is with the Grade 3 literacy tests that teachers had the greatest difficulty in marking accurately. Here teachers tended to mark at 5 percentage points *below* the correct level. In Grade 6 mathematics, it is those teachers who teach the worst performing learners who have the greatest difficulty in marking accurately. When these teachers mark incorrectly, it tends to be because they do not give learners credit when they should. This suggests that part of the problem in schools where learners struggle to learn is that teachers apply marking memoranda too rigidly, especially in the case of language scripts, and are not able to identify alternative but appropriate responses provided by learners. This has important implications for the emphasis within the professional development programmes.

A key aim of verification ANA was to examine the degree to which teachers marked accurately. The graphs that follow provide percentile plot curves for the mark out of 100 obtained by teachers themselves and obtained through the re-marking process. It is only in the case of Grade 3 literacy that the two sets of marks produced substantially different curves. In other words, in the remaining three tests the marks teachers obtained and the marks obtained through re-marking did not change the overall picture substantially and the accuracy of the marking of teachers is not as problematic as some had foreseen before ANA. To illustrate, using marks provided by teachers the percentage of learners achieving a score of less than 35% in Grade 6 mathematics is 71%. If we use the marks after re-marking, the figure becomes 69%. On average, teachers gave slightly lower marks than they should have in this test.

Figure 7: Grade 3 literacy distribution of scores

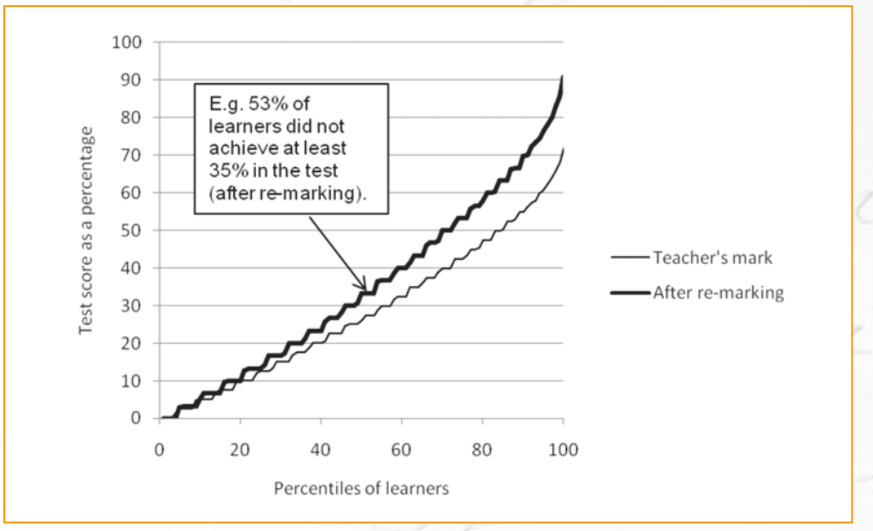


Figure 8: Grade 3 numeracy distribution of scores

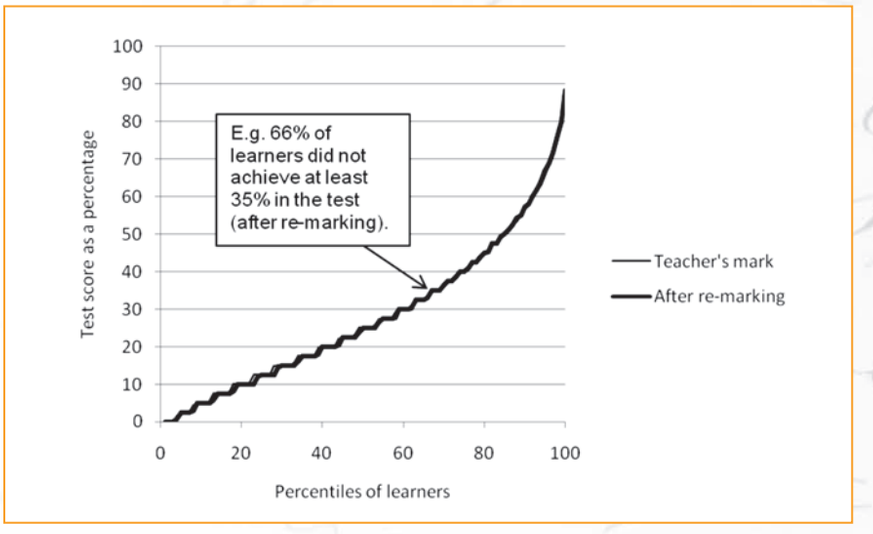


Figure 9: Grade 6 languages distribution of scores

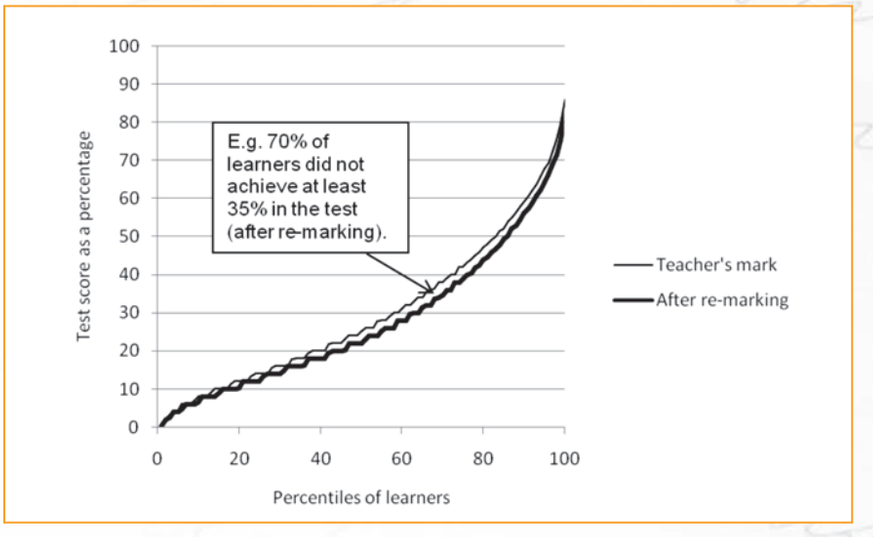
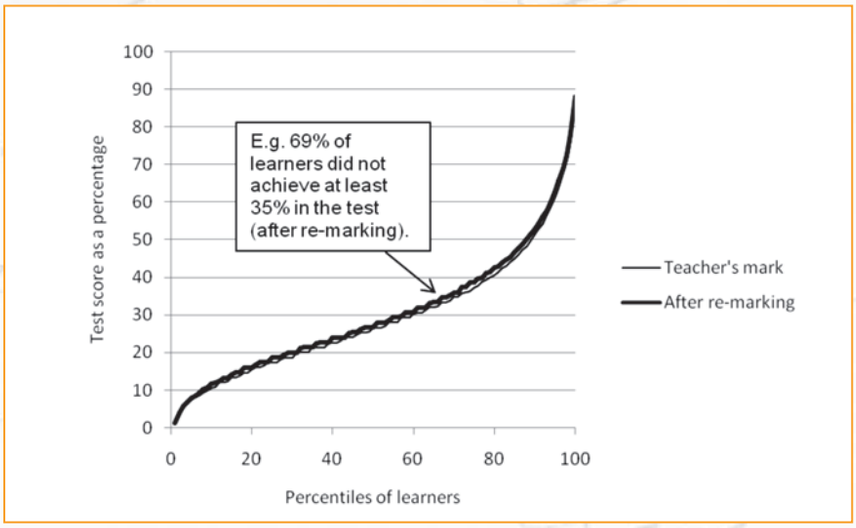


Figure 10: Grade 6 mathematics distribution of scores



A more detailed examination of the gap between the teacher's mark and the mark after the re-marking process reveals that though overall the teacher's mark provides a relatively accurate picture of learner performance, there is a substantial minority of teachers who require support in marking accurately. The next table indicates that the median gap between the teacher's mark and the mark after re-marking was 5 percentage points in the case of literacy. In other words, in the middle of the range the tendency was for teachers to mark in such a way that, for instance, a learner would be given 30% for the test instead of the 35% that he or she should get. At the two extremes of the range, the gap was much larger, however. For instance, for 5% of learners the gap was 21 or more than 21 percentage points, in other words a situation where the teacher was under-valuing the learner's responses to at least this degree. In Grade 3 numeracy, although the median teacher in every province was marking exactly as he or she should, the errors committed by some teachers were substantial. Specifically, 5% of learners were obtaining at least 10 percentage points *less* than they should and 5% of learners were obtaining at least 8 percentage points *more* than they should.

Table 4: Adjustments made to Grade 3 teacher marks (in percentages)

	Literacy			Numeracy		
	Lowest	Median	Highest	Lowest	Median	Highest
EC	-6	5	19	-10	0	10
FS	-3	6	20	-8	0	10
GP	-3	6	21	-8	0	8
KN	-3	8	24	-10	0	10
LP	-5	4	19	-8	0	10
MP	-4	3	18	-8	0	8
NC	-5	3	18	-5	0	8
NW	-4	4	18	-10	0	8
WC	-3	8	23	-5	0	10
SA	-4	5	21	-8	0	10

Note: In this table and the next one, 'Lowest' is the 5th percentile whilst 'Highest' is the 95th percentile.

The following table confirms that the small difference between the overall pattern of results if one uses the teacher's mark and if one uses the mark after re-marking is partly due to the fact that positive and negative errors cancel each other out. There is clearly a need for teachers in both the Foundation and Intermediate Phases to improve their marking skills, partly to ensure that learners are given positive feedback when this is due and are not discouraged in the learning process.

Table 5: Adjustments made to Grade 6 teacher marks (in percentages)

	Languages			Mathematics		
	Lowest	Median	Highest	Lowest	Median	Highest
EC	-12	-2	6	-5	1	9
FS	-12	-2	4	-5	1	8
GP	-12	-2	6	-7	0	8
KN	-12	-2	6	-5	1	8
LP	-14	-2	4	-5	1	8
MP	-12	-2	6	-4	1	8
NC	-10	0	6	-4	1	7
NW	-10	-2	6	-5	1	8
WC	-12	0	6	-4	1	8
SA	-12	-2	6	-5	1	8

What is clear in the following table is that the Grade 3 literacy marking problems are greater when it comes to better performing learners. On average, learners whose mark, according to the teacher, is 50% or above, should receive an additional 12 to 13 percentage points. This is a very large gap and suggests that teachers are often unable to appreciate responses from better performing learners who may provide more imaginative responses that go beyond the examples provided in the marking memorandum. This is a serious problem that could discourage learners and undermine learning. It is critical that this matter be strongly emphasised in the unfolding improvements to the teacher professional development system.

Table 6: Adjustments made to Grade 3 literacy scores by level

	Level 1	Level 2	Level 3	Level 4
	1-34	35-49	50-69	70-100
EC	3.2	7.8	8.7	7.9
FS	3.4	9.4	12.5	14.8
GP	3.3	10.3	14.1	11.0
KN	5.4	11.7	15.3	15.7
LP	3.2	9.1	11.3	14.8
MP	3.3	8.7	10.3	4.5
NC	2.5	7.4	11.0	14.6
NW	2.9	8.5	13.2	15.0
WC	5.3	11.9	15.0	17.1
SA	3.8	9.7	12.5	12.6

Note: Level refers to the level of the original teacher's mark. Each value represents the average percentage point increase applied to the original percentage score during the re-marking process (the same applies to the tables that follow).

The following table indicates a worrying trend in Grade 6 mathematics. Here it appears as if the teachers of the learners who struggle most are the teachers most likely to mark inaccurately, and when they do this they tend not to give learners the credit they deserve. The national figure of a 1,3 percentage point *under-valuation* by teachers with respect to level 1 learners means around one response per test was being marked incorrect when it should be marked correct. For learners with low scores and for whom passing the critical threshold of 35% is important, such errors can cause discouragement and confusion.

Table 7: Adjustments made to Grade 6 mathematics scores by level

	Level 1	Level 2	Level 3	Level 4
	1-34	35-49	50-69	70-100
EC	1.4	2.0	1.4	-0.4
FS	1.6	0.6	-0.9	-3.3
GP	1.1	0.4	-0.4	-1.2
KN	1.0	0.7	1.8	0.8
LP	1.5	0.8	-1.7	-8.2
MP	1.3	1.1	1.4	0.3
NC	1.2	0.7	-0.1	-0.9
NW	1.4	0.5	-0.4	-2.5
WC	1.9	2.0	1.3	0.5
SA	1.3	1.0	0.5	-0.8

6.4 Item-level patterns at the Grade 3 level

Patterns in the item- or question-specific test responses of learners in Grade 3 suggest that in most provinces ANA test administration procedures were relatively well implemented. However, in three provinces (Eastern Cape, KwaZulu-Natal and Limpopo), the degree of similarity between the responses to the same question by learners in the same school suggests that test administration procedures may not have been properly carried out. In this regard, clearer instructions and better monitoring in future runs of ANA seem necessary.

Verification ANA involved the data capturing of responses to individual test questions. This provides a rich source of information to inform specific issues in the design of future tests, the emphasis needed in professional development and what must become clearer in learning materials such as the national workbooks. Here these data are used for just one purpose, namely to examine possible problems in the test administration process.

There is a close relationship between the average test mark within a school and the degree to which learners obtain the same mark for specific questions in the test. In a school where all learners obtain 100% in the test, all learners will obviously have the same mark for each question in the test. However, the lower the average mark for the school, the lower will be the degree to which learners obtain the same mark in the same question. An analysis was undertaken to see to what degree schools in particular provinces deviated from the normal relationship between average test mark and degree of similarity in question-specific marks. This was done for

Grade 3 numeracy only. The results in the following table can be interpreted as follows. In the case of Eastern Cape, the percentage of learners with the same question-specific score tended to be around 3 percentage points higher than one would expect, given the average mark for the school. In other words, where one might expect 60% of learners to have the same question-specific mark, in Eastern Cape 63% of learners would have the same mark. What this suggests is that in Eastern Cape, learners would tend to provide the same *incorrect* response to a greater degree than one would expect. This could point to two things. It is possible that in some schools learners were guided during the test administration process, but in such a way that this guidance was not always correct. However, the pattern seen here could also reflect the incorrect teaching of certain types of mathematical operations within individual schools or classes. The problem, according to the table below, is greatest in Eastern Cape. However, it is also present in KwaZulu-Natal and Limpopo. In the case of the other provinces, the degree of similarity in responses to the same question was more or less what one would expect.

Table 8: Indicator of above normal similarity of learner responses[5]

EC	2.9
FS	-0.8
GP	-0.4
KN	1.6
LP	1.4
MP	0.0
NC	-0.3
NW	0.6
WC	-0.4

6.5 The percentage of learners performing at specific levels

In all four of the tests examined in this report, fewer than half of all learners in the country perform at a level that indicates that they have at least partially achieved the competencies specified in the curriculum. In Grade 6, the results indicate that only around 30% of learners fall into this category. At the top end, too few learners are able to achieve outstanding results. For instance, only 3% of learners in Grade 6 mathematics can be considered outstanding.

The ANA tests and marking memoranda are designed to produce the following correspondence between the percentage mark and descriptions of achievement.

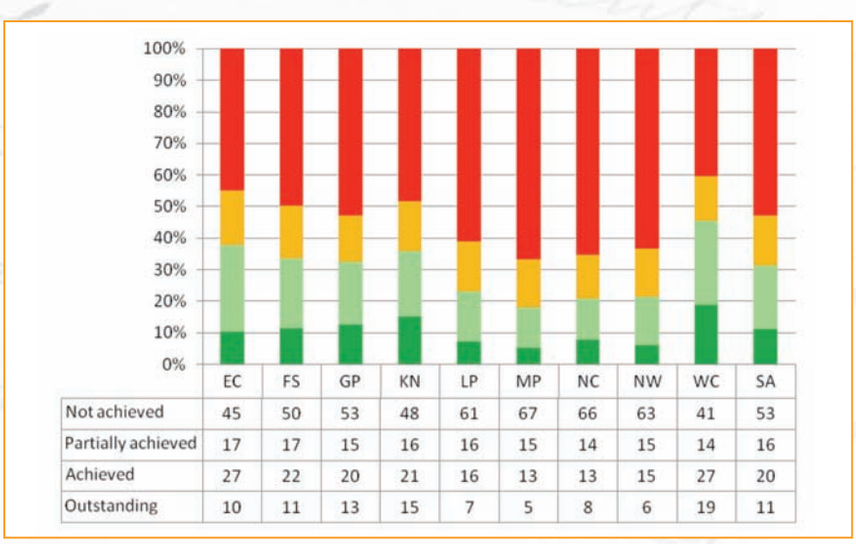
5 The figures in the table are slope coefficients for provincial dummies from a regression analysis, with MP serving as the reference province. Basically, an indicator of similarity of test responses was regressed on the average test score and provincial dummies. The regression occurred at the level of the school. The indicator of similarity of test responses for each school was calculated by first finding the percentage of learners having the most common mark for each question. Thereafter the average was found across all test items to obtain an average for the school.

Table 9: Levels of achievement

Level 1	Not achieved	Less than 35%
Level 2	Partially achieved	At least 35% but less than 50%
Level 3	Achieved	At least 50% but less than 70%
Level 4	Outstanding	At least 70%

The next four graphs illustrate the distribution of learners across the four achievement levels, with breakdowns by province. Given the fact that the data are from a sample, it is important to keep in mind the confidence intervals of the statistics. At the national level, each statistic, for instance that 11% of learners should have achieved an outstanding level of performance in Grade 3 literacy, is subject to a confidence interval of around 5 percentage points. In other words, we can be 95% certain that the true national statistic for outstanding performance for Grade 3 literacy lies between about 8.5% and 13.5%. The confidence intervals for the provincial statistics are around twice as large, in other words around 10 percentage points. A critical statistic is the percentage of learners achieving at levels 2, 3 or 4, meaning that they have achieved at least a reasonable part of the knowledge and skills they should have achieved by their grade. In Grade 3 literacy, this statistic was 47% at the national level. This is similar to the corresponding statistic found in the 2007 Grade 3 Systemic Evaluation, which stood at 48%. However, as has been emphasised above, comparisons between ANA 2011 statistics and previous assessment results need to be undertaken with much care, for a variety of reasons. The key overall finding is that in 2011, learner performance continued to be well below what it should be, especially for the children of the poorest and most disadvantaged South Africans.

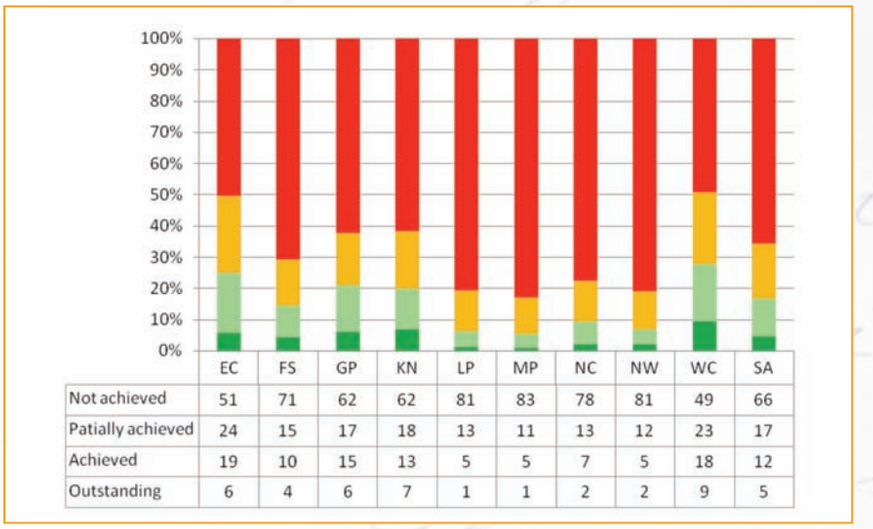
Figure 11: Performance in Grade 3 literacy by level



	EC	FS	GP	KN	LP	MP	NC	NW	WC	SA
Not achieved	45	50	53	48	61	67	66	63	41	53
Partially achieved	17	17	15	16	16	15	14	15	14	16
Achieved	27	22	20	21	16	13	13	15	27	20
Outstanding	10	11	13	15	7	5	8	6	19	11

In Grade 3 numeracy, 34% of learners attained a level of performance that represented at least partial achievement. The corresponding statistic from the 2007 Systemic Evaluation was 43%. The large discrepancy between the two figures highlights the need for a more in-depth item-level analysis to examine differences in performance with respect to similarly difficult questions, and the degree to which the overall test in ANA 2011 was overly difficult, or the 2007 Systemic Evaluation test was too easy.

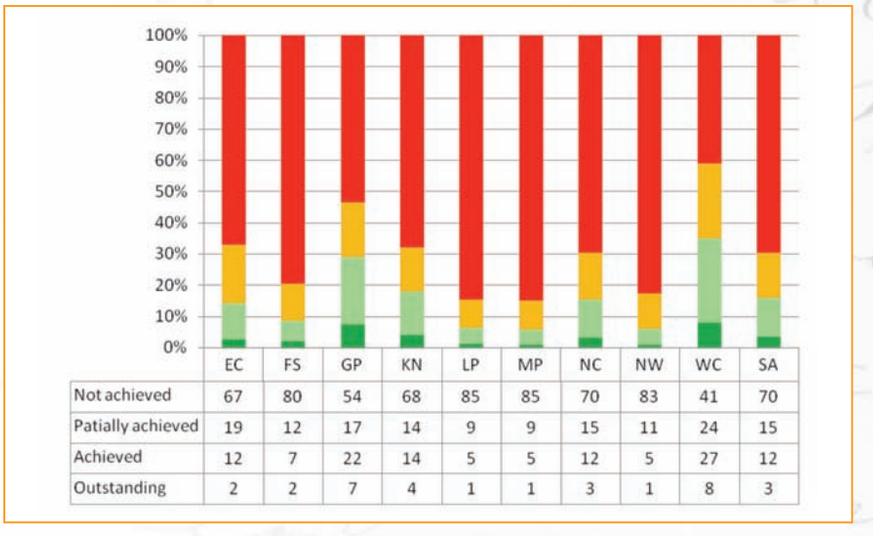
Figure 12: Performance in Grade 3 numeracy by level



	EC	FS	GP	KN	LP	MP	NC	NW	WC	SA
Not achieved	51	71	62	62	81	83	78	81	49	66
Patially achieved	24	15	17	18	13	11	13	12	23	17
Achieved	19	10	15	13	5	5	7	5	18	12
Outstanding	6	4	6	7	1	1	2	2	9	5

The Grade 6 languages results point to 30% of learners nationally reaching at least the partially achieved level. This compares to 37% in the 2004 Grade 6 Systemic Evaluation.

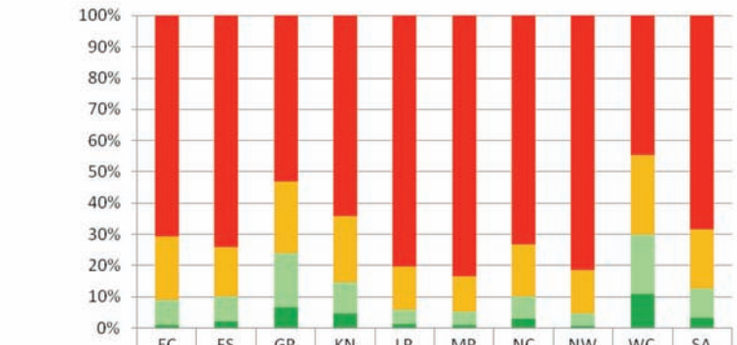
Figure 13: Performance in Grade 6 languages by level



	EC	FS	GP	KN	LP	MP	NC	NW	WC	SA
Not achieved	67	80	54	68	85	85	70	83	41	70
Patially achieved	19	12	17	14	9	9	15	11	24	15
Achieved	12	7	22	14	5	5	12	5	27	12
Outstanding	2	2	7	4	1	1	3	1	8	3

In Grade 6 mathematics, 31% of learners reached at least a partially achieved level of performance in ANA 2011. In the 2004 Systemic Evaluation, the corresponding statistic was 19%. Given the data limitations, it is not possible to conclude that this represents an unequivocal improvement in performance.

Figure 14: Performance in Grade 6 mathematics by level



	EC	FS	GP	KN	LP	MP	NC	NW	WC	SA
Not achieved	71	74	53	64	80	83	73	82	45	69
Patially achieved	20	16	23	21	14	12	17	14	26	19
Achieved	8	8	17	10	5	4	7	4	19	9
Outstanding	1	2	6	4	1	1	3	0	11	3

Preliminary results from universal ANA also point to the persistence of major challenges in improving learner performance. For example, the percentage of learners reaching at least the partially achieved level in Grade 6 mathematics, which stood at 26% and 27% in Free State and Northern Cape according to verification ANA, stood at just 28% and 29% according to universal ANA.

6.6 Challenges and opportunities at the level of the school

ANA provides opportunities that did not exist before to identify those schools within a district requiring the most urgent attention and those which can serve as examples of better practices to neighbouring schools. The data analysed here allow for a picture at the provincial level to be generated. This picture indicates that though many schools are clearly struggling, for instance in 45% of schools in the poorest quintile have almost all their learners performing at the 'not achieved' level in Grade 6 mathematics, there are also schools that do considerably better though they face the same socio-economic challenges. As an example of the latter, 13% of schools in the poorest quintile have at least half of their learners achieving at the top two levels (a score of at least 50%) in Grade 6 mathematics.

The following four tables indicate what percentage of schools fall into two categories, using marks obtained from the re-marking process. The first category is clearly struggling schools, defined here as schools where 95% or more of the learners performed at the 'not achieved' level (level 1). See the '!!' symbol in the tables. The second category is schools showing promise, defined here as schools where at least half of learners achieve at levels 3 or 4, in other words at least at the 'achieved' level. See the '✔ ✔' symbol in the tables. Most of the statistics in the next four tables point to the fact that the poorer the quintile, the greater the number of struggling schools and the smaller number of schools showing promise. This confirms the usefulness of quintiles in determining where the greatest support should be directed. At the same time, the fact that even in the better off quintiles there are struggling schools underlines the need to pay attention to under-performance in all socio-economic contexts. As an example of the latter, between 8% and 10% of schools in quintile 5 are clearly struggling with numeracy or mathematics.

Table 10: Percentage of schools clearly struggling and showing promise (Grade 3 literacy)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5		Overall	
	‼	✔✔	‼	✔✔	‼	✔✔	‼	✔✔	‼	✔✔	‼	✔✔
EC	16	53	5	68	0	80	0	88	20	60	8	70
FS	9	48	6	44	6	41	14	71	0	100	7	61
GP	30	30	45	0	8	54	4	48	0	100	17	46
KN	9	59	9	61	17	59	20	60	0	78	11	63
LP	22	31	9	49	20	40	0	33	0	0	13	38
MP	16	42	23	12	28	16	6	31	10	60	17	32
NC	35	14	17	31	18	24	0	44	9	82	16	39
NW	11	29	8	31	14	30	0	67	60	40	18	39
WC	14	71	14	43	0	72	0	70	0	96	6	70
SA	17	45	16	42	12	52	7	58	9	66	12	54

Table 11: Percentage of schools clearly struggling and showing promise (Grade 3 numeracy)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5		Overall	
	‼	✔✔	‼	✔✔	‼	✔✔	‼	✔✔	‼	✔✔	‼	✔✔
EC	28	38	16	21	13	33	13	50	0	40	14	36
FS	46	14	56	0	35	18	29	43	0	88	33	32
GP	30	0	50	0	27	42	36	18	0	83	29	29
KN	26	29	35	17	39	36	20	30	11	33	26	29
LP	66	6	29	3	44	4	67	33	0	0	51	12
MP	50	11	48	4	68	5	31	15	38	13	47	10
NC	50	3	57	7	18	12	11	22	11	44	29	18
NW	50	6	38	15	46	5	0	0	50	25	37	10
WC	21	36	17	17	6	22	4	33	0	68	10	35
SA	38	19	35	11	34	25	26	30	10	41	30	25

Table 12: Percentage of schools clearly struggling and showing promise (Grade 6 languages)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5		Overall	
	!!	✔✔	!!	✔✔	!!	✔✔	!!	✔✔	!!	✔✔	!!	✔✔
EC	38	17	63	16	33	11	0	20	0	33	27	19
FS	53	0	60	0	36	5	17	25	0	71	33	20
GP	15	31	43	0	14	17	12	60	0	100	17	42
KN	55	10	63	21	30	30	21	43	0	78	34	36
LP	74	2	59	3	50	4	25	25	0	67	42	20
MP	64	0	62	0	43	7	56	11	0	33	45	10
NC	85	0	20	12	8	23	13	25	7	57	27	23
NW	41	0	70	0	41	5	20	40	50	0	44	9
WC	0	50	0	20	17	33	0	53	0	75	3	46
SA	46	13	56	9	32	16	18	35	3	62	31	27

Table 13: Percentage of schools clearly struggling and showing promise (Grade 6 mathematics)

	Quintile 1		Quintile 2		Quintile 3		Quintile 4		Quintile 5		Overall	
	!!	✔✔	!!	✔✔	!!	✔✔	!!	✔✔	!!	✔✔	!!	✔✔
EC	45	14	47	5	25	6	0	10	22	33	28	14
FS	42	6	33	5	32	0	25	17	14	57	29	17
GP	23	38	14	29	20	9	8	35	0	86	13	39
KN	48	6	21	11	13	17	7	36	0	67	18	27
LP	54	0	56	7	25	10	33	33	0	100	34	30
MP	71	5	46	0	53	7	38	0	29	29	47	8
NC	61	4	38	8	31	8	38	25	7	43	35	17
NW	48	4	33	5	56	8	50	0	0	0	37	3
WC	8	38	9	18	6	17	0	34	0	77	4	37
SA	45	13	33	11	26	10	17	24	8	60	26	23

6.7 The ANA results in an international context

Whilst the percentage of learners performing at adequate levels is unacceptably low in South Africa, other developing countries, including relatively wealthy ones, face similar problems. This underlines the opportunities for learning from other countries and sharing South African experiences in dealing with the challenge.

The following table provides the percentage of learners in Grades 4 or 5 performing at a minimum level determined by the International Association for the Evaluation of Educational Achievement (IEA) within the 2006 Progress in International Reading Literacy Study (PIRLS) in 2006. Whilst South Africa's value was the lowest

amongst developing countries, similar problems are evident in countries such as Kuwait and Qatar, countries where inadequate education resourcing is generally not considered to be a problem. The challenge of ensuring that education resources translate into acceptable levels of learner performance is thus a worldwide one.

Table 14: Learners passing the low reading benchmark in 2006 PIRLS[6]

Participating country	% of learners
Indonesia	54
Iran	60
Kuwait	28
Morocco	26
Qatar	33
South Africa	22
United States	96

Though a rich country such as the United States does well according to the statistics used in the previous table, even in this country there is an understanding that learner performance is well below what it should be. In this regard, the figures reported within the United States No Child Left Behind programme are telling. Baseline values representing the percentage of school learners able to achieve a minimum level of performance in mathematics are around 10% in many states within the country.

7 THE WAY FORWARD FOR ANA BEYOND 2011

In the Delivery Agreement for basic education signed between the Minister of Basic Education and the President in 2010, the Minister commits herself to 'establish a world class system of standardised national assessments'. Experiences in other countries indicate that it takes several years for a programme such as the Annual National Assessments to reach the desired level of standardisation and impact. The 2011 wave of ANA, the first of its kind, was not expected to be perfect. Lessons have been learnt from the current wave and these lessons should inform the future implementation of this programme. Key lessons are the following:

- **Better logistics in the distribution of ANA materials to schools.** The late arrival of the required materials in the right languages to schools in the 2011 wave of ANA resulted in many schools administering the tests after the specified dates. This had a number of negative consequences, including the increased risk that different standards would be applied in different schools.
- **More rigorous quality assurance measures in verification ANA.** The need for measures that would ensure a higher degree of similarity in the administration of tests and sampling of test scripts within

6 See the PIRLS 2006 report at http://timss.bc.edu (p. 69).

schools in verification ANA has been discussed above. International experiences have shown that in universal assessment systems such as ANA it is important that a representative sample of schools produce highly reliable statistics that are able to inform policy and the reliability of the assessment system as a whole.

- **More standardisation within universal ANA.** In schools outside the sample, greater standardisation can be brought about through ensuring that tests materials are delivered to schools on time, but also through further training of school principals on their quality assurance role and training of teachers in how to mark ANA tests.

- **Better data collection procedures for universal ANA.** Important lessons have been learnt in 2011 on how to populate a national database of ANA results from all schools. Above all, it is important to ensure that human resources are available in all provinces to capture the data. It may moreover be preferable to focus on collecting data from all schools on key grades and subjects (which could be different in different years), as opposed to attempting to collect data from all six grades and both subjects in every year.

- **Improvements to the test design phase.** Inputs from teachers, their organisations and assessment experts in the country have indicated that there is room for improvement in the design of the tests and in the alignment of the tests with the curriculum. This matter is receiving close attention in preparation for the 2012 wave of ANA.

An important new feature in the 2012 wave of ANA will be the testing of all Grade 9 learners in languages and mathematics. Lessons from the 2011 Grade 9 pilot and the 2011 ANA process as a whole will inform this new feature of ANA 2012.

8 CONCLUSION

Though the challenges for the schooling system remain great, the 2011 wave of ANA provides a basis for optimism. Both the process of 2011 ANA and the information obtained from this process represent a basis for improvement that did not exist previously. Schools and the education departments have gained important experiences in better assessment and, through this, a better focus on what must improve. The unprecedented step of providing all Grades 1 to 6 learners with national workbooks in 2011 has, according to preliminary reports, shifted classroom practices in the right direction. The 2012 wave of ANA, to be conducted early in the 2012 school year, will serve as a critical instrument with which to monitor the degree to which national workbooks and other interventions, such as the streamlining of the national curriculum, have had an impact on learning. The 2011 wave of ANA has provided a wealth of experience in how to conduct a programme of this nature in a way that contributes to quality education.

Annual National Assessments