PROPOSALS ON THE RE-DESIGN OF THE ANNUAL NATIONAL ASSESSMENTS FOR 2016 AND BEYOND

(DRAFT VERSION – 27 JULY 2016)
### DEFINITIONS

<table>
<thead>
<tr>
<th>Term</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark</td>
<td>An actual measurement of group performance against an established standard at defined points along the path toward the standard. (LLCC, 2014: 1).</td>
</tr>
<tr>
<td>Accountability</td>
<td>The obligation placed on an education system by public officials, employers, and taxpayers for school officials to prove that money invested in education has led to measurable learning. An assessment system connected to accountability can help identify needs so that resources can be equitably distributed. (Gallaudet University, not dated, 2000: 1).</td>
</tr>
<tr>
<td>Assessment</td>
<td>&quot;The systematic collection, review, and use of information about educational programs undertaken for the purpose of improving learning and development.&quot; (Palomba &amp; Banta, 1999: 4).</td>
</tr>
<tr>
<td></td>
<td>&quot;Assessment is an ongoing process aimed at understanding and improving student learning. It involves making our expectations explicit and public; setting appropriate criteria and standards for learning quality; systematically gathering, analysing, and interpreting evidence to determine how well performance matches those expectations and standards, and using the resulting information to document, explain and improve performance.&quot; (Tom Angelo, 1995).</td>
</tr>
<tr>
<td></td>
<td>Assessment consists of two complementary processes: Measurement and Evaluation. A measurement is meaningless until it is evaluated.</td>
</tr>
<tr>
<td>Measurement</td>
<td>Collection of both qualitative and quantitative information / data about something (e.g. student learning, quality of instruction, support, curriculum, policies, etc.).</td>
</tr>
<tr>
<td>Evaluation</td>
<td>Making value judgments and informed decisions based on the measurements made.</td>
</tr>
<tr>
<td>National Assessment</td>
<td>An assessment &quot;designed to describe the achievement of students in a curriculum area aggregated to provide an estimate of the achievement level in the education system as a whole at a particular age or grade level.&quot; (Greaney &amp; Kellaghan, 2008: 7).</td>
</tr>
<tr>
<td></td>
<td>The gathering of “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.&quot; (Kanjee, 2007:13).</td>
</tr>
<tr>
<td>Diagnostic</td>
<td>Assessment carried out to identify strengths and weaknesses of an individual learner, a group of learners or a system as a whole.</td>
</tr>
<tr>
<td>Baseline Assessment</td>
<td>Measurement and evaluation of students' skills and knowledge upon entry to a learning programme aimed at providing a baseline against which to assess progress. (LLCC, 2014)</td>
</tr>
<tr>
<td></td>
<td>Assessment carried out, prior to instruction, to ascertain each student’s strengths, weaknesses, knowledge, and skills. (Dumit, 2012: 5)</td>
</tr>
</tbody>
</table>
| Systemic Assessment| Assessment carried out for the purpose of benchmarking performance and tracking “the progress made towards the achievement of the transformational goals of the
education system in respect to access, redress, equity and quality” (DoE, 2003: 2-3).

Assessment carried out “to assess the current performance and variability within a particular cohort of learners, according to some sort of external benchmark of desired proficiency, and to monitor progress, also according to some external standards for change and performance improvements over time.” (Dunne, et al., 2012: 2).

A national systemic assessment is “to identify systemic issues that are needed for success: curriculum design, development and enactment; teacher professional development; and creating and sustaining policy and management structures that support reform.” (Marx, et al., 2004).

Assessment carried out to provide “valuable data to planners in Government, the Basic Education Ministry, social partners and various institutional role players to improve the quality of basic education.” (DBE, 2014: 14).

<table>
<thead>
<tr>
<th><strong>Sample-Based Assessment</strong></th>
<th>Assessment carried out periodically involving selected representative groups of the national learner population for the targeted school grades or chronological grades as part of a national assessment to provide for “greater breadth of measurement, fuller coverage of the curriculum and avoid distortions deriving from ‘teaching to the test’. They can be carried out at comparatively low cost.” (OECD, 2011: 8).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full-Cohort (Universal) National Assessment</strong></td>
<td>The measurement and evaluation of every learner’s knowledge, understanding and skills in target school grades or chronological ages.</td>
</tr>
<tr>
<td><strong>Percentile</strong></td>
<td>A ranking scale ranging from a low of 1 to a high of 99 with 50 as the median score. A percentile rank indicates the percentage of a reference or norm group obtaining scores equal to or less than the test-taker’s score. A percentile score does not refer to the percentage of questions answered correctly, it indicates the test-taker's standing relative to the norm group standard.</td>
</tr>
<tr>
<td><strong>Norm</strong></td>
<td>A distribution of scores obtained from a norm group. The norm is the midpoint (or median) of scores or performance of the students in that group. Fifty percent will score above and fifty percent below the norm.</td>
</tr>
<tr>
<td><strong>Norm Group</strong></td>
<td>A random group of students selected by a test developer to take a test to provide a range of scores and establish the percentiles of performance for use in establishing scoring standards.</td>
</tr>
<tr>
<td><strong>Performance criteria</strong></td>
<td>The standard(s) by which student performance is evaluated.</td>
</tr>
<tr>
<td><strong>Continuous assessment</strong></td>
<td>Continuous assessment refers to making observations and collecting information periodically to find out what a student knows, understands and can do. Specific tasks are given to the learners based on what has been taught. Teachers observe the learners doing these tasks and make a judgment about how well they are doing. Continuous assessment is on-going and helps the teacher to find out what the learners have learned. Some other terms that are similar to continuous assessment are: classroom based assessment, running records, and teacher grading.</td>
</tr>
<tr>
<td><strong>Testing</strong></td>
<td>Testing is one way of assessing learners on a continuous basis. Tests usually come at the end of a topic or unit to find out what a student has learned. Testing can include a wide range of question types, but the most common are multiple choice, true and false, essays and matching.</td>
</tr>
<tr>
<td><strong>Exams</strong></td>
<td>Exams are usually carried out at the end of the year or cycle (for example, at the end of primary school). Apart from knowing what grade they got, students do not often get feedback on their performance on the exams. Exams are usually written in the same way that tests are written. Exams often have important consequences for students’ future.</td>
</tr>
<tr>
<td><strong>Assessment activities</strong></td>
<td>Assessment activities are activities given to learners to find out what they know and can do. An assessment activity is one in which the teacher is checking to see if learners have met the objectives of the syllabus, lesson or curriculum. Children often learn a lot from good assessment activities. Examples of assessment activities are writing a story or paragraph, making a model, solving problems and role playing.</td>
</tr>
<tr>
<td><strong>Quantitative Methods of Assessment</strong></td>
<td>Methods that rely on numerical scores or ratings. Examples: Surveys, Inventories, Institutional/departmental data, departmental/course-level exams (locally constructed, standardized, etc.).</td>
</tr>
<tr>
<td><strong>Qualitative Methods of Assessment</strong></td>
<td>Methods that rely on descriptions rather than numbers. Examples: Ethnographic field studies, logs, journals, participant observation, and open-ended questions on interviews and surveys.</td>
</tr>
<tr>
<td><strong>Curriculum Alignment</strong></td>
<td>The degree to which a curriculum’s scope and sequence matches a testing program’s evaluation measures. (LLCC, 2014: 2)</td>
</tr>
<tr>
<td><strong>High Stakes Assessment</strong></td>
<td>Assessment whose results have important consequences for students, teachers, schools, and/or districts. Such stakes may include promotion, certification, graduation, or denial/approval of services and opportunity. (LLCC, 2014: 2)</td>
</tr>
<tr>
<td><strong>Sampling</strong></td>
<td>A way to obtain information about a large group by examining a smaller, randomly chosen selection (the sample) of group members. If the sampling is conducted correctly, the results will be representative of the group as a whole. (LLCC, 2014: 4)</td>
</tr>
</tbody>
</table>
1. **INTRODUCTION**

The introduction & implementation of the Annual National Assessment (ANA) programme has been a major strategic intervention in education. The assessment has generated a wealth of valuable data that, in turn, has resulted in more focused & evidence-led public and professional engagement on issues of quality in education. Typical of any major intervention, the rollout of ANA began to raise critical issues and questions for both sustenance and enhancement of the initiative. In 2015, there was an impasse with teacher unions on the writing and impact of ANA.

The ANA programme is currently under review and there has been extensive consultation with the Teacher Unions to establish suitable models that are fit for purpose. The proposal is to conduct a two-tier model of administering a Universal National Assessment which is a test administered annually to all learners at Grades 3, 6 and 9 and a Systemic Assessment which is conducted bi-annually to a sample of learners at Grades 3, 6 and 9.

The Universal Assessment model comprises of:
(a) A diagnostic test which is Phase focused (based on test items from different grades in the phase) and,
(b) A summative (end of year) assessment which will be piloted in one grade (e.g. Grade 6) and later another Grade (e.g. Grade 9) will be added.

The Systemic Assessment requires an outsourced and suitably skilled service provider to report independently on system-wide monitoring of quality learning outcomes, as indicated in the sector plans of Government and the Ministry of Basic Education.

The aim of this document is to exemplify features of the Universal and Systemic assessments as possible models for the redesign of ANA in 2016 and beyond.

2. **BACKGROUND**

The historical success of ANA has been largely due to collaborative efforts and shared support between government, labour and parents to assess learners at regular intervals in gateway subjects of Mathematics and Languages. In support of improving the quality of basic education, policy interventions taken up at the 2009 African National Congress (ANC) Elective Conference resolved that conducting external tests, namely
the Annual National Assessments (ANA), should take place for all Grade 3 and Grade 6 learners every year, and the results should be provided to parents. This policy decision was further supported as a collective goal by all alliance partners at the 53rd National Conference in 2012 and has since formed part of the National Development Plan (NDP) and the basic education sector plan, Action Plan 2019: Towards Schooling 2030. Against this mandate of both the ruling party and Government, the Department of Basic Education (DBE) has ensured that Grade 3, 6 and 9 learners are measured every year.

The original plan for ANA launched within the auspices of the Foundations For Learning Campaign (in 2008) had been exceeded and all public schools and state-funded independent schools have since 2011, administered ANA according to the standardized time-table every year. With each round of the ANA, the programme continued to be improved and had become an important yardstick for the education sector to measure progress made against targeted interventions and programmes.

However, during the implementation of the 2015 ANA, teacher unions indicated their dissatisfaction with ANA and threatened not to participate in the writing. The following issues were raised by the Teacher Unions concerning the ANA:

(a) The tests are administered on an annual basis and hence the system is not given adequate time to remediate.
(b) There is a need for a more intensive programme of teacher development to address the shortcomings identified through ANA.
(c) ANA can only be written after it is remodelled.

In an attempt to resolve the impasse between the DBE and Teacher Unions, an agreement was reached between parties on the establishment of a Task Team to undertake the remodelling of ANA and an Inter-Ministerial Committee (IMC) to attend to the broader issues of dispute presented by the Unions. A concept document was developed by the Task Team and its proposals around a Universal and Systemic assessments have been welcomed as a foundational step in the re-design process. This document builds on proposals flagged by the Task Team and consolidates further inputs from academic experts and decision making structures within the Ministry such as HEDCOM and CEM.
However, before more narrow details of the proposals listed earlier are discussed, it is important to foreground and clarify how such models are viewed in assessment literature on large scale assessments.

3. CLARIFICATION OF CONCEPCTS

This section seeks to clarify some important terms and concepts pertinent to the discourse of national assessments. These are presented below to provide a theoretical basis for the modelling of a new national assessment models for South Africa.

3.1 Defining Assessment

Palomba and Banta (1999: 4) define assessment as "the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving learning and development." Coined differently, assessment may also be defined as a “systematic process or method of gathering information to better determine the knowledge possessed or achieved by a student.”

3.1.1 National Assessments

Kanjee (2007:13) sees national assessment as referring to the gathering of “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.” In South Africa, the annual national assessment was meant to provide "valuable data to planners in Government, the Basic Education Ministry, social partners and various institutional role players to improve the quality of basic education." (DBE, 2014: 14). Overall, results generated may be in the form of quantitative and/or qualitative formats.

3.1.2 Systemic Assessment

The DBE (2003: 3) uses the term 'systemic evaluation' to refer to the determination of the extent to which the education system achieves set social, economic and transformational goals through the measurement of learner performance as well as the context in which learners experience learning and teaching. This is in line with the Assessment Policy of the South African Education Department which stipulates that such systemic assessment be conducted in three grades of the education system, namely Grades 3, 6 and 9. Accordingly, the DoE states that “the main purpose of Systemic Evaluation is to benchmark performance and track the progress made towards the achievement of the transformational goals of the education system in
respect to access, redress, equity and quality” (DoE, 2003: 2-3). Thus, DoE (2003: 3) outlined the objectives of Systemic Evaluation as to:

(a) determine the context in which learning and teaching is taking place;

(b) obtain information on learner achievement;

(c) identify factors that affect learner achievement; and

(d) make conclusions about appropriate education interventions.

In its conceptualisation of systemic assessment, the DoE (2003: 3) envisions the collection of data at two levels:

The learner achievement component of Systemic Evaluation seeks to establish trends with respect to acquisition of key knowledge, skills, values and attitudes by learners at different points in the system. The contextual component is set to provide insight into the environment in which teaching and learning take place and to establish the performance of the education system with respect to the principles of access, redress, equity and quality.

Overall, it is envisaged that systemic assessment could draw the required data from the major components of the Education System, namely the teachers, education officials, provisioning officials (e.g. those providing teaching/learning support materials and allied services), as well as learners. According to Dunne, Long, Craig and Venter (2012: 2), whilst classroom-based assessment is generally fine-grained and topic specific, external systemic assessment is generally broadly banded, and attempts to cover the curriculum. Thus, the major purposes of systemic assessment are seen as “to assess the current performance and variability within a particular cohort of learners, according to some sort of external benchmark of desired proficiency, and to monitor progress, also according to some external standards for change and performance improvements over time” (Dunne, et al., 2012: 2). To Marx, et al (2004), the purpose of systemic assessment is “to identify systemic issues that are needed for success: curriculum design, development and enactment; teacher professional development; and creating and sustaining policy and management structures that support reform.” On their part, Greaney and Kellaghan (2008: 7) opine that national assessments are “designed to describe the achievement of students in a curriculum area aggregated to provide an estimate of the achievement level in the education system as a whole at a particular age or grade level.” The methodology typically makes use of sample data.
3.1.3 Universal Assessment

This model targets the measurement and evaluation of every learner's knowledge, understanding and skills, for the purpose of identifying and/or developing a suitable learning programme. It has the following characteristics:

(a) Conducted prior to instruction or intervention to establish a baseline from which individual student growth can be measured.

(b) Used to establish individual student's knowledge, understanding and skills level about the subject at the onset of a learning cycle; thus, helps the teacher to plan for the learning that follows more effectively and appropriately for different learning needs of students, individually and/or severally.

(c) May also be used as a ‘placement’ assessment – i.e. to place learners into different streams of learning, depending of identified learning needs.

(d) The marks obtained by individual students are not reported to parents or any other authority.

The Figure below illustrates processes of a diagnostic cycle.

Thus, diagnostic assessment is considered an essential device in a teacher’s “tool kit” that can be used to diagnose strengths and areas of need in all students. It involves the gathering and careful evaluation of detailed data using students’ knowledge and skills in a given learning area. The data assists teachers to plan for appropriate pedagogy and target learning to more effectively scaffold the learning needs of their students. Consequently, diagnostic assessment is used ‘for learning’ where taking action to adjust teaching plays a significant role in improving learning outcomes for all students.
4. REVIEW OF THE ANNUAL NATIONAL ASSESSMENTS (ANAS)

In July 2015, the Minister released the draft National Policy on the Conduct, Administration and Management of the Annual National Assessment (ANA) for public comment. As an initial step and departure point for, the re-design process considered a comprehensive review of the strengths and weaknesses of the ANAs inputs made by various members of the public. Through that process a number of weaknesses and limitations of the current model were identified, including the following:

- That the purpose of the ANA was unclear, and appeared to combine the purposes of both a systemic and universal (diagnostic, as it was referred to by participants);
- That the ‘confused’ purpose(s) of the ANA led to further confusion, or lack of clarity, regarding the utilisation of the results and the data emanating therefrom;
- That as a result of the above points, as well as the absence of clear protocols for the release and use of data, the results of the ANA were both abused and misused – including being used in a punitive manner, rather than developmental and transformational as some people had envisaged;
- That some of the abuse and misuse of the ANA results included comparing school performance which, in turn led to a number of undesirable and unintended consequences and practices, such as competition between schools and ‘teaching to the test’ – thereby compromising and adversely affecting the validity, reliability and credibility of the results;
- That, for a number of reasons the ANA reports had limited application as they were found by most teachers and schools not to be user-friendly;
- That the different data sets emanating from the ANAs – such as universal, verification and moderation data led to confusion.
- That the administration process of the ANA tests opened itself to abuse – thereby compromising the validity, reliability and credibility of the results;
- That the burden of accountability as a result of the ANA results appeared to fall on the shoulders of classroom teachers and schools, exempting other levels of the education system, such as the national, provincial and district;
- That the current process was capital intensive;
- That there was a lack of transparency at the various critical levels – including who was responsible for setting the ANAs;
- That the apparent link between the ANA results and ‘subsidy allocation’ remained unclear;
• That it was not clear how the ANA results were used to improve any aspects of the curriculum (planning, implementation and assessment);
• That the frequency of the ANAs (i.e. annual turn-around time) did not provide for sufficient time for the preparation of the necessary reports, plans and instructional interventions for the assessments to make any noticeable impact;
• That where there were interventions, these were not focused – and did not address the identified challenges;
• That there was no monitoring in respect of planned interventions, and that where efforts were made to do so, such efforts were uncoordinated and fragmented;
• The lack of anchor items (and a general absence of a well-defined questioning framework) made it difficult to make reliable and valid longitudinal comparisons of learner performance;
• That the verification of the ANA results was not related to contextual issues;
• That the ANAs took away teaching time, overburdened school resources and overloaded teachers with extra work – for which they received no compensation;
• That, as a result of school comparisons, the ANA had turned into high-stakes examination;
• That the timing of the ANAs made it impossible to know the actual levels of competency of the tested school grades in-so-far as the fourth term work was concerned.

5. INPUTS FROM ACADEMIC EXPERTS

To further clarify points of departure, academic experts in the field of national assessments were consulted to provide inputs on lessons drawn from international perspectives and best practices which provided an important frame for the South African context.

Inputs from academics on the ANA policy emphasised the point that purpose determines the design of the assessment, the types and extent of the data collected, as well as the sources of data and levels of analysis. In this regard, two possible purposes of national assessments fell under discussion. Both were diagnostic in nature; the difference was what was being diagnosed. In one case, the state of the system needed to be measured, to undertake a high-level assessment of the correctness and effectiveness of policies, structures and processes. This is similar to the type of assessment done in international studies like TIMMS, PIRLS and SACMEQ, and can be done at periodic intervals given the slow rate of change, especially in a large system.
like ours. This assessment would not make any demands on teachers, and most schools would not be included in the sample.

At the same time, all teachers need to know whether they are doing what they should for their learners, and if not, what needs to be attended to. Schools also need to know if their learners are progressing at the desired rate, and again what must be done to sustain or improve this. A sample-based, systemic assessment will not give this information to teachers, and it is therefore also necessary to have a more regular, universal assessment of all learners in a particular phase, which will provide invaluable developmental data to schools and teachers. This kind of assessment should be of a "low-stakes" nature, and used only to improve the quality of teaching and the performance of learners. Schools may also use these results to signal specific support needed from the District, the SGB and parents.

It is clear from the literature that data of a systemic or organisational level is required, as well as data relating to classroom practice. Both are important for the purpose of ensuring comprehensive and profound educational change (Kanjee, 2007; National Center for Education Statistics, 2010; Senghor, 2014; ACARA, 2015; Moloi, 2016).

An exclusive focus on systemic data will, by design and intent, place no emphasis on information needed to influence classroom practice. Similarly, it may also be said that an exclusive focus on the classroom through the collection of learner performance data will not yield the feedback needed to sharpen systemic processes and procedures. Although inferences can be drawn from universal learner-based performance data about systemic operations and efficiencies, and vice versa to a certain extent, there is no substitute for collecting purpose-driven data.

Looking at the respective purposes of systemic and universal assessments, it therefore makes sense to consider the option of a system of national assessment which seeks to achieve both purposes. Such a system would embrace both the above purposes, with appropriately designed instruments for each.

(a) Universal assessment
Experts strongly advised that in keeping with international practices, and given the above rationale and purpose, there should be an annual, universal assessment of all learners at the end of the Foundation, Intermediate and Senior Phases of the GET Band (Grades 3, 6 and 9).

The purpose of a universal assessment would be to identify strengths and weaknesses of individual or groups of learners in numeracy and literacy, at the endpoint of each Phase in the General Education and Training (GET) school band. These would be aggregated at different levels to provide an indication of achievement, in order to provide learners, parents, schools, the government and its social partners with important information that should be used to improve teaching and learning.

The assessment should be aligned to CAPS and designed to cover the work of the entire phase, and not just the Grade. The assessment should focus on the application of the content more than the recall of items, and questions of all orders should be included. The assessment shall not be taken into account for progression purposes, and copies of the tests, the model answers and the results should be freely available as development tools.

The standardised instruments will assess learner performance in Language and Mathematics, and all assessments shall be conducted in the Language of Learning and Teaching.

The primary level of reporting should be at school level to enable schools to diagnose learning gaps linked to CAPS. The annual cycle is appropriate in this case, since classroom and school-level interventions can be designed and implemented from year to year.

(b) Systemic Assessment

Experts also advised on the conduct of a sample based assessment. This assessment will collect information and data on the education system as a whole in order 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.” (Kanjeе, 2007:13).

The assessment will evaluate the functioning of the system, and test the validity of policies and practices of the Department. The data will identify “systemic issues that are needed for success: curriculum design,
development and enactment; teacher professional development; and creating and sustaining policy and management structures that support reform.” (Marx, et al., 2004).

A Systemic Assessment will also allow for international benchmarking and trend analysis across years, with confidential anchor items and questionnaires that are independently administered. It is recommended that the systemic assessment be conducted periodically, usually within a range of 2-4 year cycles. There are a number of reasons for this periodic sequencing:

- One is that sample data and systemic assessments provide systemic indicators, which are relatively stable over shorter periods of time and do not need frequent measurement;
- The second is that a three year cycle would lead to the same cohort being tested every three years, which is not scientifically valid, and may have the unintended consequence of favouring these cohorts in some way; and finally
- Four years provides a long enough period to allow for the impact of interventions to be measured and evaluated.

The sample based assessments should be administered in Grades 3, 6 and 9, as a representative sample of levels. The assessment should be done at the same time as the universal assessment, towards the end of the final year of the Phase. The sample based assessments should be independently conducted and quality assured, and should place no additional burdens on schools or teachers. Scientifically valid criteria should be used to ensure that the assessment takes account of all contextual factors that may be relevant to the analysis. This assessment should be benchmarked against regional and international assessments.

6. **THE PROPOSED FEATURES OF NATIONAL ASSESSMENT FOR 2016 AND BEYOND**

Taking into account the above arguments and context, there have also been increasing calls for external standardised assessments to be used for promotion of learners, partly to validate the reliability of school-based assessment but also to optimise the returns from the investment that the Department makes in the assessment. The 2016 Basic Education Sector Lekgotla, called for the introduction of a regulatory framework for school based assessment (SBA). This was considered a necessary measure to remedy weaknesses and strengthen the quality of SBA practices. Currently, the DBE is drafting guidelines on the conduct, administration and management of internal examinations at school level as the first step in providing assistance to schools on the conduct of SBA.
Within this context and the assessment literature, a two-tier model of conducting national assessments in South Africa, with a universal and systemic strand is proposed: This Universal Strand will comprise two legs with the following features:

(a) A universal diagnostic test drawn and exemplar items supplied to all schools for teachers to use at school level for focussed learner support.

(b) A universal summative test conducted annually at the end of the school year that will form part of the promotion of learners.

Systemic Evaluation strand is a bi-annual (sample-based) strand that could be used to monitor learner trends and report on the quality of learning outcomes.

Both strands will focus on Grades 3, 6 and 9. Grades 3, 6 and 9 are key focus grades of DBE and therefore assessment will be rigorously controlled in these three grades. These grades are regarded as key stages in the General Education and Training (GET) band since they contain the required competencies for exiting one curriculum phase and entry into another. It is therefore appropriate to assess learner competencies in these grades.

The content areas assessed will be Mathematics and Languages as these are considered gateway areas for establishing competency levels in other subjects offered in the curriculum. The design features of each leg of the proposed model are explained further below.

6.1 Universal Diagnostic

The DBE envisions providing sets of diagnostic tests for Grades 3, 6 and 9 in English Home Language, English First Additional Language and Mathematics to all public schools for teachers to use at school level for focussed learner support. Diagnostic tests will be developed by appointed panels of test developers with the purpose to inform instruction and better help learners learn. The 2016 test items will be sourced from a bank of items that were previously piloted for ANA but were not used. The items will be tweaked to serve a diagnostic purpose. The tests that will be released for 2017 and 2018 academic years will be drawn from new items developed by the test panels.
The diagnostic tests have a formative purpose, where teachers can assess learners while a topic is being taught (checking for understanding), or they may administer the assessment as a baseline at the beginning of the academic year or it may be divided into quarterly assessments based on the topics covered.

In 2016, the diagnostic test will comprise 50 MCQs that can be administered on an IT platform as well as open response items that teachers will be required to score. Thereafter, from 2017, diagnostic tests will be released incrementally, according to the topics as they appear on the CAPS work schedules.

The 2016 and 2017 tests will be released on an IT platform with a smart phone app and on CD. Schools may download the app and gain access to the items which may be administered either on an ICT platform or as a pen and paper exercise. The MCQs can also be scored on the ICT platform.

**Design features**

**Frequency:** Commencement in 2016 will be voluntary. Initial roll out will be annual and then be replaced with an item bank in 2018. In 2017, it may be used as a baseline assessment at the beginning of the year or divided into quarterly assessments based on topics covered.

**Test development:** Diagnostic tests, exemplar items and assessment guidelines.

**Test format:** The test focus will be diagnostic, low stakes and coverage will be for the phase (difficulty % spread: 5% basic; 25% slightly below grade level; 60% at the key grade; and 10% above grade level). For example a Grade 3 test will have the following spread of questions across grades:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Difficulty</th>
<th>% Spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>Basic</td>
<td>5</td>
</tr>
<tr>
<td>02</td>
<td>Slightly below grade level</td>
<td>25</td>
</tr>
<tr>
<td>03</td>
<td>Key Grade</td>
<td>60</td>
</tr>
<tr>
<td>04</td>
<td>Above grade level</td>
<td>10</td>
</tr>
</tbody>
</table>

**Test administration:** done by teachers; provided by PEDs, no time-table. IT platform to support writing and feedback on selected items (e.g. cell-phones). Diagnostics tests will be made available electronically and printing of tests will be the responsibility of PEDs where applicable.

**Marking:** at school level (moderation by SMT)
Capture: The capture and analysis of marks will be at school level (using school administration systems) or suitable alternative IT platforms (e.g. cell phone applications).

Reporting: A school report that shows areas of strength and support of learner may be provided to parents. Qualitative Diagnostic reports can be done at district and provincial levels. It will be the responsibility of the DBE to provide tools for qualitative reporting. No aggregation of learner scores beyond school level is expected.

6.2 Universal Summative

A Universal Summative test could be administered across all public schools and in state subsidised independent schools in Grade 6 and at a later point Grade 9 could be added. In the intermediate and senior phase, the Universal ANA tests will replace a formal school based assessment (SBA) task in the third term or fourth term of the academic school year.

The motivation for this leg is that the test will become integrated into the Curriculum and Assessment Policy (CAPS) progression and promotion requirements. The summative tests will be recorded as a formal assessment task and be included in the final SBA mark for progression (Grades 1-8) and promotion (Grades 9). A seamless integration of external standardised tests into the SBA tasks will also satisfy increasing calls for these tests to be used for promotion of learners, partly to validate the reliability of school-based assessment but also to optimise the returns from the investment that the Department makes in the assessment. In addition, it paves the way for an external set final examination in key Grades and subjects that takes place at the end of the academic year.

The implication for integrating the universal summative tests into the SBA tasks of CAPS will lead to reduced administrative work on the part of teachers to set a formal assessment task in Mathematics and Languages. Teachers will not have to mark the summative tests in addition to their own SBA tasks required in CAPS. The end of the year promotion schedules will only require slight modifications to identify the universal summative mark from the other SBA tasks.

Design features
Frequency: Annual, will replace the end of year examination in specific grades. It will initially commence with Grade 6 in 2018 and Grade 9 will be added in 2019. (Model will be similar to NSC).

Test format: Content and skills assessed will be curriculum based specific to a grade.

Test development: DBE will appoint examination panels for 2017 and 2018. Thereafter, tests will be drawn from item banks developed by a service provider.

Test administration: done by teachers; tests printed, packed and distributed by PEDs, national standardised timetable.

Marking: There should be national and provincial marking guideline discussions, and this will be cascaded to teachers who will do the marking done at school level. Marks should be moderation of by School Management Team (SMT) members / and sample marking should be done at district/cluster levels.

Capture: The capture of marks should be done at school level (using their own school administration systems (e.g. SA-SAMS)). The marks from schools will be fed into IT platforms that allows for aggregated scores to be established for system level reporting.

Reporting: Reporting is possible at all levels of the system (i.e. National/PED/District/School). Schools fill in promotion/progression schedules with the universal summative mark indicated.

6.3 Systemic Evaluation

The second strand of assessment is a bi-annual systemic evaluation that will be conducted on a sample of Grade 3, 6 and 9 learners. The purpose of this assessment is to monitor learner trends and report on the quality of learning outcomes. This type of assessment will allow for international benchmarking and trend analysis across years, with confidential items and questionnaires that will be independently administered.

The focus of the assessment will be on key competencies and on end of phase outcomes. The test development will be outsourced to a service provider. Systemic Evaluation will be conducted every two years by an independent agent in Grades 1 to 9 across a random sample of approximately 2000 public schools selected from all the nine (9) provinces. In a sub-sample of 50 schools per province, the independent agent will conduct an in-depth item analysis for each grade and subject in the sample.
The scope of the work shall include monitoring test administration, administering tests, marking tests, capturing learner scores, analysing results and reporting performance to the DBE. The DBE’s strategic document, “The Action Plan to 2019: Towards the realisation of Schooling 2030” provides a good overview of the need to separate the Universal and Systemic functions of national assessments. Systemic evaluation is better focused on producing performance information that is comparable over time and across provinces. It aligns with best practices on national assessments and researchers on assessment have indicated that in other countries (e.g. USA) have demonstrated that comparable statistics on language and mathematics performance over time need to be obtained through a sample-based testing system, as it is only within a limited exercise such as this that it is possible to achieve the required rigour and levels of reliability. The number of learners that will participate in this leg is reflected in Table 1 below.

Table 1: Maximum number of learners in the sample

<table>
<thead>
<tr>
<th>Grade</th>
<th>Schools</th>
<th>Learners (25 per school)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1,125</td>
<td>28,125</td>
</tr>
<tr>
<td>6</td>
<td>1,125</td>
<td>28,125</td>
</tr>
<tr>
<td>9</td>
<td>1,125</td>
<td>28,125</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>84,375</td>
</tr>
</tbody>
</table>

Design features:

Frequency: Once every two years, focusing on key competencies at Grades 3, 6 and 9. The first cycle should start in 2017 followed by a second cycle in 2019.

Test format: The focus will be on end of phase outcomes.

Test development: The design of tests will be outsourced to a service provider. The evaluation will include contextual questionnaires based on policy concerns relevant to the medium term strategic framework of the Department.

Test administration/Marking, Analysis and Capture: sample based (statistical models) will administered, collected, marked and captured by service provider.
**Reporting:** the reporting will be based on monitoring progress against quality learning outcome and the service provider will provide an independent National Statistical Report, 9 Provincial Statistical Reports and a qualitative Diagnostic Report on learner competencies.

7. **USE OF DATA AND INTERVENTIONS**

The ultimate purpose of assessment is to provide relevant, meaningful and user-friendly feedback to targeted users so that they can use the information to develop appropriate interventions and improve performance. Learners need the feedback so that they can know the content areas that they need to improve on. Similarly, teachers need to know the skills and content knowledge in which their learners require specific development and support. By the same token, subject advisors and curriculum development officials at district level need to provide teachers with support that addresses specific learning challenges as informed by the assessment report.

7.1.1 **Setting and reporting according to performance standards**

The Department collects massive amounts of data at a cost through national assessments. There is evidence that investment in national assessments has paid off in a number of education systems that have consistently implemented the assessments. However, inability to utilize the assessment results effectively has also been identified as one reason assessments do not lead to improved performance in many other systems that have selected this approach to improving educational outcomes. Some of the reasons given for under-utilization of information from national assessments include the nature and quality of feedback to the system.

An alternative approach that has great potential to improving national performance involves the report of learner achievement against performance standards. Performance standards are expected levels of learner competency in a given assessment area, as defined by the assessment standards. Performance standards are connected to assessment standards in that they articulate acceptable measurement forms and minimum learner performance levels necessary to achieve the set standards. Whilst it involves specialized expertise in psychometric measurement, by its nature standard setting is a subjective process. For this reason the participation of key stakeholders is critical in the process. Key stakeholders in this context will be a) policy and decision makers, b) curriculum specialists, c) teachers and d) learners. Besides necessary consensus

---

1 Uruguay, India and Liberia (Africa) are some of the systems that have used assessment effectively to improve education (Crouch, L. 2008).
on what the standards should be, direct involvement and responsibilities of the stakeholders varies at different stages of the standard setting process.

Most assessments use a dichotomous “pass/fail” classification to make decisions about learners or participants in a particular programme. Often high-stakes assessments use this classification. But for national assessments often three or more levels are identified. Each level must have a name that conveys the overall expectation for performing at the level. Popular performance level names include “Basic”, “Proficient” and “Advanced”. Once the names of the levels have been agreed upon by consensus, each level is described by short focused written statements called “performance level descriptors (PLDs)”.

Once developed, tested and agreed to be appropriate, PLDs will serve a number of purposes which will include:-

i) Providing coherent information (standardized) across the grades and subjects.

ii) Guiding teachers’ instructional efforts to ensure that learners reach acceptable performance levels because they will know the types and levels of complexity of items at the defined levels.

iii) Guiding test developers in the development of assessment items that are pitched at the required level of performance.

7.2.2 Interventions

When the diagnostic, summative and systemic results of the assessment are reported separately, each report will be comprehensive and specific to the target audience and purpose and will, therefore, meet fully and validly the needs of the target population. Diagnostic reports will provide detailed information on what each child in a grade can or cannot do at levels below, at and above the current grade. Teachers need this information for appropriate interventions.

In particular, the data from the universal assessments should be utilised as follows:

(a) To assist teachers, schools and districts to develop focused interventions to improve teaching and learning;

(b) To give an indication of any support (including additional learning materials) needed to improve classroom teaching, learning and assessment;

(c) To guide schools in setting standards

(d) To provide information to learners and parents
(e) To design and attract teachers to targeted development programmes

(f) To enable districts, parents and schools to have a common source of information.

The data from the sample-based Systemic Evaluation results should be used to:

(a) Improve operational systems at national, provincial and district levels with a view to enhancing the quality of teaching and learning

(b) Guide all levels of the Department in setting targets in relation to the national benchmarks for learner achievement

(c) Provide information for the system to benchmark performance and to provide a measure to track progress of learners over time

(d) Enable education stakeholders to interpret the information presented and use it to develop interventions to improve performance.

Follow-up interventions by the DBE should be multi-layered, focusing on those parts and levels of the system that are shown to be under-performing. The following responsibilities should inform the intervention strategies:

**National level**

(a) Develop a national intervention plan and strategy arising from the data

(b) Construct a management plan for the interventions based on the results

(c) Monitor national progress on learner achievement in mathematics and languages against set targets.

(d) Use the National Assessment results as a systemic tool for the development and review of policy.

(e) Write and amend guidelines on the utilisation of results as necessitated by circumstances.

**Provincial and District levels**

(a) Analyse the quantitative and qualitative data per District

(b) Develop improvement plans to address systemic challenges

(c) Support Districts in preparing development programmes

(d) Monitor schools and Districts to evaluate the effectiveness of intervention strategies

**School level**

(a) Schools, under the leadership of the Principal, and with the support of the District, to analyse their own results in detail
8. IMPLEMENTATION PLAN

The DBE is planning to convene panels of test developers to develop the exemplar (diagnostic) test for Grades 3, 6 and 9 in June 2016. The exemplar tests will be developed and implemented on an ICT platform. The administration will be voluntary and will be administered at the schools convenience. The test development panels will also develop summative assessment tests to be piloted in October/November 2016 and administered in 2017.

Parallel to this process will be the appointment of a service provider to develop an Item Banking system as well as a service provider that will develop, administer, mark and report on the systemic tests in 2018. Table 2 below provides an indication of the implementation plan for the 2016/2017 financial year.

9. FINANCIAL IMPLICATIONS

The key cost drivers for the re-design process includes:

a) Universal
   - Development of materials (exemplar tests and guidelines; summative tests).
   - Appointment of examination panels.
   - Editing and quality assurance.
   - Development of Item bank/outourcing to service provider.
   - IT platforms for administration and feedback (e.g. Vodacom).

b) Systemic Evaluation
   - Appointment of service provider.

The cost for the implementation of the universal diagnostic and summative tests is estimated at R 23 196 539. The cost for the implementation of the systemic assessment is estimated at R 40 200 000.

An estimated breakdown of test development costs are indicated in Table 2 below.
<table>
<thead>
<tr>
<th>No.</th>
<th>ACTIVITY</th>
<th>COST IN “R”</th>
<th>TIMEFRAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Recruitment and appointment of Grades 3, 6 and 9 test panels, training of test panels, versioners, editors and adaptors</td>
<td>1 503 552</td>
<td>May-June 2016</td>
</tr>
<tr>
<td>2.</td>
<td>Finalisation of the item specifications for item banking, TOR for item writing and appointment of a service provider: Part 1 in 2016/7</td>
<td>11 158 432</td>
<td>June –July</td>
</tr>
<tr>
<td>3.</td>
<td>Development of test items for the 2016 test exemplars in Grades 3, 6 and 9 Language and Mathematics, versioning, moderation and editing (diagnostic)</td>
<td></td>
<td>Ongoing from June</td>
</tr>
<tr>
<td>4.</td>
<td>Development of test items for the 2017 assessment instruments in Grades 6 and 9 Language and Mathematics, versioning, moderation, adaptation and editing of test items (summative)</td>
<td>5 195 423</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Compilation, formatting and in house printing of test booklets for piloting (Grades 3, 6, 9)</td>
<td></td>
<td>July-September</td>
</tr>
<tr>
<td>6.</td>
<td>Recruitment and appointment of markers for the Pilot Study (2017 test instruments)</td>
<td>1 939 132</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Training of PEDs on the Pilot Study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Pilot Study in Grades 3, 6 and 9 for the 2017 assessment instruments, marking, capture and analysis of data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Quality assurance of test exemplars</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Refinement of items for the 2017 instruments and selection of items and compilation of final test drafts</td>
<td>2 800 000</td>
<td>October -November</td>
</tr>
<tr>
<td>11.</td>
<td>Quality assurance of the 2017 assessment instruments and signing off of tests</td>
<td></td>
<td>March 2017</td>
</tr>
<tr>
<td>12.</td>
<td>Pilot Study of the exemplar test on an IT platform in 50 schools</td>
<td>600 000</td>
<td>October -November</td>
</tr>
</tbody>
</table>

Estimated total budget: R23 196 539

Part 2: Systemic Evaluation estimated costs

<table>
<thead>
<tr>
<th>No.</th>
<th>ACTIVITY</th>
<th>COST IN “R”</th>
<th>TIMEFRAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Development of the data collection manuals</td>
<td>1 800 000</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Training of fieldworkers</td>
<td>1 200 000</td>
<td>2017/2018</td>
</tr>
<tr>
<td>3.</td>
<td>Development of Language and Mathematics items for Grades 6 and 9 and field testing of items</td>
<td>10 000 000</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Test administration and monitoring</td>
<td>6 000 000</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Collection of scripts from schools, marking and moderation of scripts</td>
<td>10 200 000</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Capturing of data, analysis and report writing</td>
<td>2 500 000</td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Printing of manuals, test booklets and reports</td>
<td>3 500 000</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>Sundries</td>
<td>3 000 000</td>
<td></td>
</tr>
</tbody>
</table>
Estimated total budget | R40 200 000

Added costs will include the use of IT platforms and dissemination of tests to PEDs which is estimated at R500 000 for the 2016/2017 financial year.

10. **PROTOCOL ON PROVINCIAL ASSESSMENTS**

A protocol on conducting National Assessments must be developed to regulate all assessment processes and avoid “assessment overload” in schools. This should take account school based assessment programmes, as well as Provincial or District level assessments and common examination processes.

11. **CONCLUSION**

Going forward, all parties must respect and protect the integrity of the agreed upon proposals on national assessments. To ensure the credibility of the system, there should be stringent monitoring of potential risks and the implementation of the follow-up interventions at each level, and regular impact assessments should also be conducted.

The proposals listed in this document provide a broad outline for the new design and format of the National Assessment. In the next few months, further work will be done on explicating:

(a) The purpose, format and structure of the Universal Assessment.
(b) The purpose, format and structure of the Systemic assessment.
(c) Policy amendments.
(d) Implementation plan for the new model at different levels of the system.
(e) Provincial assessments and the need to streamline programmes to avoid assessment overload.
(f) Role and use of teachers in centralised marking (summative).
(g) Use of the data and results to inform teaching and learning interventions.

The process of developing this document has been guided by local and international literature on national assessments, and by the views of specialists in this area, which have been overlaid on the deliberations of the Task Team.