

MAINTAINING STANDARDS IN THE NATIONAL SENIOR CERTIFICATE

Vijayen D Naidoo

Umalusi – South Africa

Acknowledgements: The presenter would like to place on record that this paper draws substantially from the Umalusi Maintaining Standards Project which is a project that cuts across all the operational units of Umalusi. To this end, recognition is given to all those who contributed to the project.

Introduction

Most large-scale examination systems include measures to ensure consistency over periods of time. Umalusi (Council for Quality Assurance in General and Further Education and Training) currently manages a statistical moderation process whereby results obtained in final exit examinations are adjusted to maintain reasonably consistent standards over time. Prior to 2008, this moderation included several measures including adjusting raw scores on the basis of historical averages (*norms*) calculated from learner performance over five-year periods; *pairs analysis* in which the average results for a particular subject in each instance are compared to the average results of all other subjects, in turn, for the same group of learners; and comments (expert judgement) made by internal and external moderators.

In 2008 Umalusi urgently needed to review its systems in this area – the main reason being the introduction of the new qualification—the *National Senior Certificate* (NSC) based on the National Curriculum Statements. The first national exams at Grade 12 level took place at the end of 2008. What had to be addressed immediately was that there were no historical norms for the associated examination results. To ensure the integrity of these results, Umalusi had to have a valid understanding of the quality and levels of cognitive demand of the new curricula relative to those just superseded- the Senior Certificate (SC) based on the NATED 550 curriculum. Furthermore, unlike the SC, the NSC does not comprise of higher and standard grades, but single sets of papers which encompass levels previously separated into higher and standard grades.

Umalusi’s Quality Assurance of Assessment (QAA) and Statistical Information and Research (SIR) units, together with the Statistics and Assessment Committee of the Umalusi Council, put in place a range of different strategies with regard to strengthening Umalusi’s quality assurance of assessment in 2008. The overall strategy included the creation of new “guideline norms” and in-depth research into the levels of difficulty of key curricula and their associated exams.

The intention was that the findings of the research involving in-depth curriculum evaluation and exam paper analysis be used to further the just use of *pairs analysis* and new norms in 2008. The research was specifically designed to provide Umalusi with succinct information on the

comparability of the old SC and new NSC curricula, and on the comparative difficulty of the exams associated with each to adjudicate on the standard of the new NSC exams in 2008, in relation to the standard of the previous SC exams.

The main (high enrolment) ‘gateway’ subjects used to assess suitability for entrance to tertiary institutions were selected viz. English First Additional Language; Geography; Life Sciences (previously Biology); Mathematics; Mathematical Literacy; and Physical Science.

STRATEGY FOR MAINTAINING STANDARDS IN THE TRANSITION FROM THE NATED 550 TO THE NEW NATIONAL CURRICULUM: THE CREATION OF “GUIDELINE” NORMS

Given Umalusi’s view that one of the key determinants of the fairness of learners’ results is contextualisation of these results in relation to historical norms – situating the results within the context of the previous performance of *several* similar cohorts of learners in similar subjects – it was essential to have a valid and fair set of norms within which to standardise the 2008 results. Given also that many of the NCS subjects were thought to be similar to their SC predecessors, or to combinations of SC counterparts, it was decided to base the new norms on previous ones.

The Statistical Model

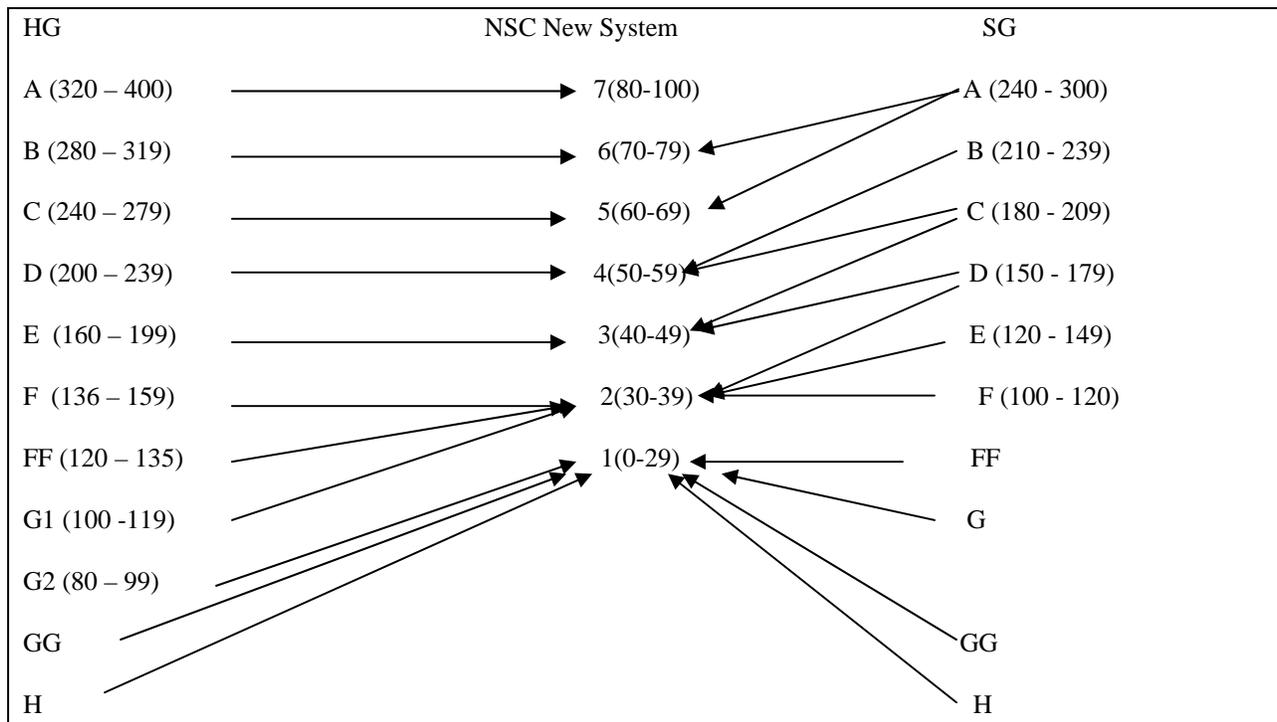
Where subjects had predecessors in terms of similar subjects, and even if these subjects had been offered and taken at differing levels in the old National Education (NATED 550) system, the norms of the preceding subjects were used. In cases where these subjects had been offered on Higher and Standard Grade levels, *combinations* of previous Higher and Standard Grade norms were used in 2008. In the case of Mathematics, the previous Higher Grade norm was used in 2008; for Mathematical Literacy, the previous Standard Grade norm was utilised.

For new subjects such as Civil-, Mechanical-, and Electrical Technology; Consumer and Hospitality Studies; and Information and Computer Applications Technology, hybrid norms were constructed. These hybrids incorporated the norms of old NATED 550 subjects thought to have been close in nature to the new *National Curriculum Statement* (NCS) subjects. In some cases up to eight or ten previous norms were combined in these hybrids.

Table 1 indicates the statistical model used for constructing the guideline norms. Grade 12 learner data (raw marks) for the period 2002 to 2007 was used to determine the new “guideline” norms. The model suggests that the top achievers in Higher Grade will still remain top achievers in the new system and that learners in 2008 will not be unnecessarily advantaged or disadvantaged in comparison with learners from the previous years. It does not suggest grade inflation.

While such use of the existing norms is logical, for the integrity of the 2008 standardisation process it was necessary to test the validity of the approach. The in-depth research into the six NSC subjects was part of this testing exercise: the study was used to triangulate as it were, the validity of the 2008 results in relation to the new norms. The research also assisted the positioning of the non-researched subjects through comparison facilitated by pairs analysis.

Table 1: Statistical model for constructing “Guideline Norm”



Note: The previous HG and SG had a grading scale of A to H (The mark intervals are indicated in brackets). The new NSC has a grading scale of 1 to 7 (The percentiles are indicated in brackets).

STRATEGY FOR MAINTAINING STANDARDS IN THE TRANSITION FROM THE NATED 550 TO THE NEW NATIONAL CURRICULUM: IN-DEPTH CURRICULUM AND EXAM PAPER EVALUATIONS

Four critical assumptions underlay this aspect of the research, first, that a comparison of the three curricula (NATED HG and SG, and the *National Curriculum Statement*) would provide an indication of whether the demands made by each are comparable. It was assumed that the old and new curricula did require comparable levels of knowledge and skill in order to pass.

It was also thought that a comparison of the expressed requirements for the setting of final exit examinations would provide an indication of whether learners are required to perform at similar

levels in the old and new examination systems. The underlying assumption here was that the requirements for the new subject examinations were basically comparable with a 'blend' of the old Higher and Standard Grade exam requirements.

A third assumption was that a comparison of the analyses of the 2008 NSC *exemplar* exams and old SC exams would give an active indication of the level of difficulty of the 2008 NSC exemplars in relation to the old Higher and Standard Grade SC versions of those exams. Similarly, it was expected that a comparison of the analyses of the *final* 2008 NSC examinations with the analyses of the 2008 NSC exemplars and old SC exam papers would indicate the level of difficulty of the final 2008 NSC exams in relation to both the exemplars, and exams set prior to 2008. The underlying thought here was that the 2008 NSC exams would be poised at such a level that they would enable learners achieving at the levels of past learners just passing at Standard Grade level to pass. In addition, the new exams would also contain sufficient difficult items so that learners achieving at the highest levels would be earning results equivalent to the 'A-grades' achieved by previous learners at Higher Grade levels. Finally, the new papers would contain items that distinguish accurately between learners with a range of academic proficiencies. In other words, the assumption was that the examinations would be set in such a way that their levels of cognitive difficulty could be used to identify learners achieving at the top and bottom ends of the spectrum, as well as to discriminate a range of levels between these extremes, capturing the range of achievement displayed in the old HG and SG exams.

The fourth assumption was that if all of the related research findings were presented in a sufficiently clear and systematic manner, they would be able to support the Umalusi Statistics and Assessment committee in making decisions associated with the fairness, reliability, and levels of the new 2008 examinations.

Methodology

Teams of four researchers evaluated the NATED 550 Higher and Standard Grade, and *National Curriculum Statement* curricula for each subject. They also analysed all Higher and Standard Grade exam papers from 2005 to 2007, as well as the 2008 exemplar and final papers for their subjects. In each case, the evaluators had to make a myriad of judgments prior to commenting on the respective levels of difficulty of the curricula and exams. Their final judgments are based on a series of prior judgments, for which they were required to provide trails of evidence in each instance. Research instruments used ensured consistency of reporting across individuals.

Curriculum Analysis

The specific research questions (or concluding task) for the curriculum evaluation was:

Is the assumption that the NATED 550 curricula and the National Curriculum Statement require similar levels of knowledge and skill in order to pass a justifiable assumption? Regarding the levels of cognitive difficulty comprised by the three curricula, in an overall sense, how do the National Curriculum Statements rank against the NATED 550 Higher Grade curricula and the NATED 550 Standard Grade curricula respectively? Are the NCS curricula comparable to the Higher Grade or Standard Grade curricula, or to mixtures of the two previous curricula? If the level of difficulty of the NCS curriculum for the subject in question is somewhere between that of the earlier Higher Grade and Standard Grade curricula, in what proportions are the respective percentages of the levels of each of the earlier curricula? How should it be rated – for example, would it be, say, 60:40 HG to SG – based on actual counts of ratings recorded for all preceding sets of questions?

Main curriculum evaluation findings

Regarding determining the precise levels of difficulty of the respective curricula, in the process of making judgments on the relative levels of difficulty of the NATED 550 and NCS curricula, the subject teams drew on various aspects of their fine-grained analyses. All the teams drew, for example, on their findings relating to the specification, weighting, and foci of content and skill topics. The Physical Sciences, Mathematics, and Geography teams found that information on amounts and levels of difficulty of content and skill topics yielded solid evidence of the respective overall levels of difficulty of the curricula. The Mathematical Literacy team focused on cognitive types and levels of skills in order to make their decisions. The English FAL team compared degrees of specification of content and progressive increase in complexity of skills in their comparisons.

The Life Sciences team drew on a wide range of aspects including specifications for external assessment, when making their judgments. The Geography team included the nature of the organising principles, finding that the *transmission* of disciplinary knowledge and skills required in the NATED 550 curriculum was easier than the *application* of this knowledge and these skills in the NCS system.

In all, three teams (Physical Science, Life Sciences, and Mathematics) found their NCS curricula to be midway between the NATED 550 Higher and Standard Grade equivalents, in 50:50 proportions. The Geography team found the NCS Geography curriculum between the NATED 550 Higher and Standard Grade levels, but closer to that of the Higher Grade, in a 60:40 Higher: Standard Grade relation. The English FAL team found the NCS curriculum for their subject more difficult than both the NATED 550 Higher and Standard Grade courses.

Specific Curriculum Findings: Physical Science

The most important points arising from this comparison for Physical Science were the following:

- With regard to breadth of content it **cannot not** be concluded that the new curriculum falls somewhere between the old HG and SG curricula. The breadth of content in the NCS far exceeds that for either of the previous curricula (requiring roughly 30% more class time). Hence it can be concluded that the NCS is **far more difficult** in terms of breadth of content than the NATED curriculum.
- With regard to difficulty of content, it **can** be concluded from the estimates that the new curriculum falls somewhere between the old HG and SG curricula.

Table 2: Results of curriculum analysis for Physical Science

Curriculum name	Estimated class time (45 minute periods)		% class time (Full curricula)			% class time (Examined curricula)		
	Full curriculum	Examinable curriculum	Difficult	Medium	Easy	Difficult	Medium	Easy
NATED SG	241	119	34%	57%	9%	37%	53%	10%
NATED HG	254	144	37%	55%	9%	47%	44%	9%
NCS	368	162	38%	49%	13%	43%	48%	9%

Question Paper Analysis

The specific questions for the analyses of the exam papers were:

- Do the 2008 exemplar and final papers allow for learners who would have achieved A-grades in the old Higher Grade papers to achieve A-grades in the new NSC exams, *where the new A-grades are comparable to the old Higher Grade A's?*
- Do the 2008 exemplar and final papers allow for the average learner passing at the level of the old Standard Grade papers to pass the new NSC exams?
- What distinguishes most significantly, the Higher from the Standard Grade exams?
- From the analysis of the Higher and Standard Grade papers, how do the 2008 NSC exemplars compare? And how do the 2008 NSC final papers compare with the exemplars on one hand, and the Higher and Standard Grade papers on the other? Are there any points of comparison not already covered?

Main exam paper analysis findings: general comment on the difficulty levels of the 2008 final NSC papers

The subject teams commented accurately on the overall cognitive character and difficulty levels of the final 2008 *National Senior Certificate* exam papers in relation to their Higher and Standard Grade counterparts in the years 2005–2007, and 2008 exemplars, based on total counts

of items at specified cognitive type (use of taxonomies) and difficulty levels (easy, moderate and difficult)

Three teams (Physical Sciences; Life Sciences; and English FAL) gave differing fine-grained results for the respective final 2008 papers for their subjects, but on the *whole*, showed that the papers were closer to the old NATED 550 *Higher Grade* than the Standard Grade papers for the subjects. A fourth team (Geography) found that their 2008 final papers contained *more comprehension* and *problem-solving* questions than the previous Higher Grade papers for this subject – these (2008) questions being of a cognitively demanding type, and in addition, set at *difficult* levels.

The two remaining teams found their papers, on the whole, easier. Since Mathematical Literacy is a new subject and they had no previous papers to consider, the Mathematical Literacy team evaluated the 2008 final papers in relation to requirements in the *Subject Assessment Guidelines* for their subject. They found that while the spread of items in Paper 1 roughly matched those in the *Subject Assessment Guidelines*, the percentage of questions at the lower cognitive levels in Paper 2 was higher than that recommended. The Mathematics team found the final 2008 papers 1 and 2 (based on the core curriculum) closer to those of the old NATED 550 *Standard* than the Higher Grade papers. The NSC makes provision for an optional Mathematics Paper 3 which contains some of the more cognitively demanding aspects of the curriculum.

Comparability of A-grades in the NATED 550 Higher Grade and 2008 NSC papers

The subject teams commented, again based on accurate counts of the types and difficulty levels of items in the exam papers, on whether the 2008 exemplar and final papers allowed for learners who would have achieved A-grades in the old Higher Grade papers to achieve A-grades in the new NSC exams where the A-grades *were comparable to the old Higher Grade A's*.

Four teams (English FAL; Geography; Life Sciences; and Physical Science) found that because the spread of types and levels of questions in the respective papers was similar, A's in the 2008 NSC papers would be equivalent to A's in the NATED 550 Higher Grade papers.

The Mathematics team found that learners typically achieving at the level of high Cs, Bs and As in the NATED 550 Mathematics Higher Grade exams would be able to score As in the final 2008 NSC Mathematics papers-as a result of only the core curriculum being assessed.

The Mathematical Literacy team, again not having previous Higher and Standard Grade papers, considered the value of potential A-grades achieved in relation to the proportions of items at stipulated difficulty levels in the *Subject Assessment Guidelines* for the subject. It was found that the final 2008 NSC papers would *not* discriminate between top-end achievers in the subject, as the papers included on average only 22% rather than the 40% of higher cognitive-level questions recommended in the *Subject Assessment Guidelines*.

Whether the 2008 NSC papers allow for learners just passing at Standard Grade-type level to pass

Notwithstanding the overall difficulty levels of the papers, two Umalusi teams (English FAL and Mathematics) found that percentages of the lower cognitive order *basic conceptual* items were similar to those in the old Standard Grade papers for the subjects, and would therefore allow learners achieving at just-passing-Standard-Grade levels to pass. The Umalusi Mathematical Literacy team noted that while there were more than enough *easy* items to enable these learners to pass, the ambiguity of many questions would lower the pass rate from that expected from the levels of the questions.

Three teams (Geography; Life Sciences, Physical Sciences) found the proportions of *easy* items in the 2008 NSC final papers *lower* than those in the average Standard Grade papers for the subjects. The Geography team noted, for example, that the amounts of *basic conceptual* questions in the NSC papers were closer to percentages in the old Higher Grade than in the Standard Grade papers. The Life Sciences team pointed out that the number of *easy* questions in the NSC papers was very close to that needed to pass, leaving very small margins for error at that level. The Physical Science group found that it would be much harder for a learner achieving at this level to pass the 2008 NSC exams than it would have been to pass the Standard Grade exams: the 2008 final exams contained an average of 23% of *easy* items, while the average for the Standard Grade papers between 2005 and 2007 was 39%. The papers for these subjects would clearly have been very difficult for learners at the lower end of the achievement spectrum, and in the case of Physical Science, especially so.

Specific Question Paper Analysis Findings: Physical Science

A summary of the examination Paper analysis can be seen in Table 3 below:

Table 3: Results of analysis of examination papers for Physical Science

Paper description	Type of cognitive demand			Level of Difficulty		
	Factual	Conceptual	Problem solving	Level 1 (Easy)	Level 2 (Moderate)	Level 3 (Difficult)
Average all SG exams (2005 to 2007)	21%	29%	50%	28%	67%	5%
Average all HG exams (2005 to 2007)	12%	30%	58%	16%	65%	19%
Average all HG&SG (2005 to 2007)	16%	30%	54%	22%	66%	12%
2008 Exemplar Paper 1&2	25%	28%	47%	21%	65%	14%
2008 Final DoE Paper 1&2	12%	37%	50%	16%	63%	20%

Distinguishing highest level achievers (final papers)

The percentage of marks allocated at Level 3 difficulty for the final DoE examination is 20%. This shows that there is a comparable differentiation of A-grade learners with the old HG exam, and hence that Umalusi's recommended allocation of the highest grade (Level 7 on the new system) is accurate.

Determining average achievers (final papers)

The total percentage of Level 1 with the additional percentage of factual questions gives the total percentage achievable by the average SG learner. This is shown in Table 4:

Table 4: Total achievable percentage by average SG learner in Physical Science exam papers

Paper description	Level 1 questions	Additional factual questions (beyond Level 1 difficulty)	Total percentage achievable by average learner
Average all HG (2005 to 2007)	16%	5%	21%
Average all SG (2005 to 2007)	28%	10%	39%
Average all HG&SG (2005 to 2007)	22%	8%	30%
2008 Exemplar Paper 1&2	21%	14%	34%
2008 DoE Final Paper 1&2	16%	7%	23%

For the final DoE examination the total achievable marks for the average SG learner is much lower (23%) than in the old SG exams (39%). The pass mark of 30% on the new system does not adequately address this issue. Average low achieving learners are likely to score between 20% and 25%. (It was also argued that taking into account the breadth and unfamiliarity of the NCS content that this could be lower)

Comparison of overall standard of final DoE papers with the combination of HG and SG and the exemplar paper

The overall standard of the DoE final papers does not compare favourably with the combination of the Higher and Standard Grade papers. The final exam contains fewer Level 1 (easy) questions, and more Level 3 (difficult) questions than the combination of the Higher and Standard Grade papers. The final exam does compare favourably with the previous HG. There are similar percentages of questions at all levels of difficulty between these exams.

Overall Impact of strategies

Regarding the new norms, use of the constructed and hybrid norms was deemed successful by Umalusi's Statistics and Assessment committees in light of the patterns emerging from the

triangulation of findings from all sources considered during the 2008 standardisation. Actual learner performance was in most cases not far from the constructed norms. In just under 50% of the subjects moderated, raw marks were accepted. For these subjects, the average difference in the means of the constructed norms and the means of actual learner performance was 3, 44. This pattern suggests a match between desired and actual levels of learner achievement, and the suitability of the system within which the new norms were operating. It must be remembered that the *National Senior Certificate* results incorporate those marks of learners who would previously have achieved at levels typically found separately in Standard or Higher Grade exam papers. The fact that the Umalusi Committee found it possible to use raw marks signals that the 2008 papers for these subjects were seen to differentiate successfully between learners performing at, on one hand, the old NATED 550 Higher Grade levels, and on the other, those achieving at levels previously associated with Standard Grade papers.

The findings for Physical Science are significant as is borne out by the actual learner performance in the 2008 final examinations.

- The number of candidates achieving a grade 7 was approximately 800 (0.4% of 211382), which compared well with the 1295 (1.8% of 71924) candidates who obtained distinctions in the 2007 higher grade examinations.
- Approximately 40% more of the 2008 cohort performed below the 30% percent pass mark as compared to the 2007 cohort. This corresponds with the judgements made by the team the average low achieving learners are likely to score between 20% and 25%
- Approximately 12 680 of the 2008 candidates passed at the 40% level as compared to the approximately 13 665 who passed at the 40% level in the 2007 higher grade exams. This indicates that the question paper was indeed pitched closer to the previous higher grade

Pairs analysis made possible the comparison of average learner performance in any particular subject against the average performance of those same learners in all other subjects taken by them. It was possible to look at how *groups* of learners fared in similar subjects such as Physical Sciences, Mathematics, and Information Technology; or Accounting, Economics, and Business Studies. Where there were research findings for analysis of the difficulty levels of the exam papers concerned, these findings could be compared with the *pairs analysis* results. Further, it was possible to extend the reach of the research findings, through comparison of learners' average results for these subjects, with the average grades for similar subjects not yet forming part of the research sample, via the *pairs analysis* results.

Conclusion

Where the findings from the qualitative curriculum and exam paper analyses were available for a subject, these systematically arrived-at findings were considered in relation to the comments of

moderators, results of *pairs analysis*, and new norms. Looking at all of this information (results of *pairs analysis*, moderators' comments, and research findings) *in relation to* the new norms gave a relatively three-dimensional view of learner performance in relation to desired standards for particular subjects. Where research had not been conducted for any particular subject, there was heavier reliance on the new norms, results of *pairs analysis*, and moderator comments.

In all, given the triangulation of the different types of findings, the 2008 standardisation is deemed to have been very rigorous, systematic, and fair. The results of the curriculum and exam paper analyses were particularly useful in the 2008 standardisation process, as they provided more robust information than other items in the background information set. The norms had been constructed *a priori* – before the exams had been written – and the extent to which these norms would be valid was not known before the availability of the 2008 learners' results. *Pairs analysis*, at best, provides *a relative measure* of performance (performance in one subject can be viewed against performance in another subject). Moderators' comments are the comments of individuals only. The judgments emanating from the research, in contrast, comprised the combined work of *four* individuals in each instance, and constituted the integrated overall findings of a wide array of a myriad of smaller judgments made in systematic and consistent ways. The Umalusi Statistics and Assessment Committees have requested that Umalusi continues with this research, including additional subjects in the group of subjects being studied, and continues to analyse subsequent exam papers for subjects already covered.