May 2009

Signalling performance

An analysis of continuous assessment and matriculation examination marks in South African schools

UMALUSI

Council for Quality Assurance in General and Further Education and Training

Signalling performance: An analysis of continuous assessment and matriculation examination marks in South African schools[1]

Servaas van der Berg and Debra Shepherd
Department of Economics
University of Stellenbosch
South Africa
(email: svdb@sun.ac.za)

Final report to Umalusi
May 2008

1 Earlier draft versions of this report were presented to Umalusi in October 2007 and to the 5th Conference of the Association of Commonwealth Examinations and Accreditation Bodies on "Improving the Quality of Public Education in the Commonwealth", Pretoria, 9-14 March 2008. The authors wish to thank Tim Dunne and Emmanuel Sibanda for advice on the quantitative methodologies used, Derek Yu for assistance with analysis of the data, and various members of Umalusi's Research Committee for useful comments.

PUBLISHED BY

UMALUSI
Council for Quality Assurance in General and Further Education and Training

37 General Van Ryneveld Street, Persequor Technopark, Pretoria
Telephone: 27 12 3491510 • Fax: 27 12 3491511
Email: Info@umalusi.org.za • Web: www.umalusi.org.za

Contents

Abstract:

Continuous assessment (CASS) is formally becoming an increasingly important part of the evaluation of South African students at matriculation level, with the weight attached to CASS in the final matriculation mark having been increased to 25%. More importantly, perhaps, CASS is the form in which students get feedback on their performance during the matriculation year, and such feedback is likely to have an important impact on their preparation and effort for the matriculation examination. Weak assessment in the school system could thus mean that students are getting the wrong signals. This may have important consequences for the way they approach the final examination. Moreover, if such wrong signals have also been common earlier in their school careers, it may also have had an important influence on their subject choice and career planning.

To evaluate the quality of the school-based CASS compared to the externally moderated matriculation examination which, for present purposes is regarded as the "correct" assessment of student performance, this study analysis data on CASS and matriculation examinations for three years for a number of subjects. The results are presented in a format where the extent of "inaccurate" CASS can be evaluated and analysed. There are two dimensions to inaccurate assessments in terms of signalling: On the one hand, an inflated CASS can give students a false sense of security that they are likely to do well in the matriculation examinations, thereby leading to unrealistic expectations and diminished effort. Secondly, a weak correlation between the mark obtained in CASS and the examination means that the signalling to students is weak in another dimension: Relatively capable students may get relatively (compared to classmates) low CASS marks. This is referred to in this paper as poor reliability of assessment, on the premise that the examinations and continuous assessment both should be testing the same underlying understanding of a school subject as enunciated in the national curriculum. We analyse the extent of each of these two dimensions of weak signalling and attempt to present it in a format that allows an overview of the magnitude of these problems amongst South African schools, by subject area, subject, province, and socioeconomic background of schools, and whether schools are public or independent schools. The analysis draws conclusions with some disturbing policy implications for a relatively large part of the school system.

*To be technically sound, assessments must be both **valid** and reliable. An assessment is valid when it is used for the purposes for which it is designed, allowing appropriate interpretations of the results. A **reliable** assessment provides test scores that consistently measure a student's knowledge of what is being tested. Assessments used in standards-based systems should meet a third criteria (sic), **alignment**, or the degree to which the assessment adequately reflects the standards on which it is based. (Pearson Education, no date)*

Part 1:

Introduction

Continuous assessment (CASS) is formally becoming an increasingly important part of the evaluation of South African students at matriculation level, with its weight in the final matriculation mark having been increased to 25%. Matriculation results determine options for university entry, bursaries, career choice, and labour market prospects. CASS gives feedback to students on their performance during the matriculation year, and is therefore likely to influence their preparation and effort for the matriculation examination.

From the student's point of view, the objective is to get through the assessment tasks successfully (Nicol & Macfarlane-Dick, 2006). Feedback from continuous assessment ought to provide students with signals to aid them in understanding how well they have mastered the curriculum material. Unintended consequences arise when the quality and nature of the continuous assessment do not correspond with the desired learning outcomes. In the South African context, for instance, poor quality CASS, which is set by the teacher at school level, may not support the learning outcomes set by the national curriculum standards for matriculation. Poor quality assessment in schools would give students wrong signals that could influence their learning strategies, their effort for the final examination and their planning for the future. Moreover, if poor quality assessment is common in the CASS mark in matriculation, it is likely to be equally common in lower grades. Thus it may also already have exerted an influence on subject choice, career planning and even the decision to persevere to matriculation rather than pursuing alternative options, such as electing to go into vocational training in Further Education and Training Colleges (FET Colleges).

This study evaluates the quality of the school-based CASS compared to the externally moderated matriculation examination (which will for present purposes be regarded as the "correct" assessment of student performance), using Umalusi data on CASS and matriculation examinations for a number of subjects for 2005 as well as the two previous years. Alternative measures of assessment accuracy in CASS are used and applied to the data, to determine to what extent CASS marks gave poor signals to students as to their likely matriculation performance.

An analysis of assessment data compared to examination data can help to illuminate two important questions. The first relates to the subject knowledge of teachers. Teachers with poor subject knowledge are likely to give inaccurate assessment marks. Thus, such an analysis could help to identify teachers who do not teach to the curriculum standard, whether owing to poor subject knowledge or other reasons. It is also possible to determine where such teachers are located (e.g. in what province, district or school), and in which subjects this problem is particularly severe. Secondly, assessment marks act as information to matriculation candidates about how to prepare for the examination: It informs them how well they are prepared and thus what confidence they can have about succeeding in different subjects, as well as where their weaknesses are, thus allowing them to prepare better for the matriculation examination. Thus, a low signal to noise ratio in assessment implies that students have less information on how to prepare, contributing to low examination results.[2]

The paper proceeds as follows: Section 2 shows how assessment acts as signalling and how signalling can be measured; Section 3 details the data and especially the methodology followed; Section 4 provides empirical results in terms of a comparison of all marks nationally: Section 5 deals with assessment accuracy at the level of the school; and Section 6 concludes the paper and offers some policy suggestions.

2 In unpublished work, Leibbrandt and Lam (2006) have suggested another interesting hypothesis. In earlier grades, poor assessment of examinations makes it more worthwhile for weak candidates to continue in the school system, whereas better candidates have a higher probability of failing than they should have, which would make them less likely to persevere. Thus inadequate subject knowledge in the school system may have devastating results even before matriculation. Inadequate subject knowledge reflected in assessment at matriculation level would be even worse at earlier levels, as matriculation teachers are usually more experienced and often better qualified than other teachers

Part 2:

Assessment as signalling: how to measure it

CASS marks are determined at the school level, based on tasks that are not standardised across schools but vary by in terms of number, level of difficulty, and marking accuracy. Thus they are likely to be less accurate than the externally set, marked and moderated matriculation examinations. It is, therefore, a fair assumption that the examination mark is the standard against which to judge CASS marks.

This study distinguishes two types of assessment inaccuracies, in terms of their statistical qualities, with different signalling dimensions:

- **Assessment leniency (where CASS marks are much higher than examination marks):** An inflated CASS mark, evident in a large gap between CASS marks and examination marks, can give students a false sense of security that they are better prepared for the examinations in that subject than they in fact are. This could elicit inappropriate studying behaviour (e.g. diminished effort in that subject) and thereby further lower examination results. For present purposes, the examination mark will be taken to have been an accurate measure of performance in each subject.[3] Some teachers adopt a strategy of purposely limiting CASS marks, to encourage candidates to work harder for the examination. But the data showed that only in about 12.2% of cases in the subjects selected for this analysis did examination marks exceed CASS marks, an indication that this was not a very widespread practice. The rising weight attached to CASS marks in the final mark also militated against such a strategy. Thus teachers should ideally aim to minimise the difference between the CASS mark and the examination mark. But even if this difference is small or even zero for an individual class or school, the information content of the CASS mark could still be inadequate, if the second measure (correlation) is poor.

- **Low assessment reliability (where performance measured by CASS and the examination marks are only weakly correlated):** A poor correlation between CASS and the examination mark indicates that the latter is also an unreliable indicator of the individual's relative ability compared to classmates in a particular subject. In such a case, a student who scores low in the CASS mark may score unexpectedly well in an examination, compared to others who obtained better CASS scores, or the converse could also occur: that relatively high CASS marks would lead to disappointment when the examination marks become available. In such cases of a weak correlation in a particular school and subject between the CASS and the examination mark, signalling to students is weak in another dimension: CASS marks do not act as a good predictor of examination marks even in relative terms. But as even a high correlation can still not provide full information to candidates of what to expect and thus how to prepare for the examination; the first measure (the gap) is also important in this regard.

Poor performance in CASS in either or both of these two dimensions of assessment (gaps or correlations) will be referred to as "inadequate" or "inaccurate" assessment. Ideally, there should be small gaps and high correlations between the two types of assessment. The extent of each of these two dimensions of inaccurate signalling will be analysed by subject, province, socioeconomic background of schools (school quintiles), and independent versus public schools.

3 This ignores the possibility of inconsistent examinations marks or that the examination marks may be endogenous, determined in part by learning behaviour responding to the assessment mark. Even if there were full correspondence between CASS and examinations in levels of difficulty, one would expect a less than perfect correlation between marks. Firstly, it is likely that students put in enhanced effort in the final matriculation examination, particularly if they performed weakly in CASS. This should serve to reduce or even reverse the gap actually observed, i.e. higher CASS than examination marks. Secondly, there is some variation in how well candidates perform even when the test difficulty and their preparation remain unchanged.

Given the standards and content set by the national curriculum, a fair degree of consistency in the assessments for each subject should be possible. In well-functioning schools, one would expect, in the terms of the opening quote of this paper, validity (tests being well designed to test curriculum knowledge), reliability (consistently measuring the student's knowledge) and alignment with the national standards.

Moskal and Leydens (2002) saw reliability as follows:

> Reliability refers to the consistency of assessment scores. For example, on a reliable test, a student would expect to attain the same score regardless of when the student completed the assessment, when the response was scored, and who scored the response. On an unreliable examination, a student's score may vary based on factors that are not related to the purpose of the assessment.

Elsewhere they also stated that "*When the cause of variation in performance and the resulting scores is unrelated to the purpose of the assessment, the scores are unreliable*." (Moskal & Leydens, 2002). The North Central Regional Technology in Education Consortium (2002) glossary of education terms declared that reliability meant that "*The same person is likely to get approximately the same score across multiple test administrations.*" This, however, does not distinguish adequately for present purposes between two sources of difference between scores on different assessments, viz. differences in the means (levels) and differences in the relative scores (correlation). The former of these two was earlier referred to as the alignment, "*the degree to which the assessment accurately reflects the standard being measured*" (Burger, no date), and should thus ideally be distinguished from assessment reliability. It is useful, therefore, to consider another discussion of this concept:

> Equivalent forms reliability examines the extent to which scores acquired from the same population on two different versions of an assessment are comparable. If different items (or test – the authors) truly measure the same concept, then it would be expected that the results of individual responses across these items would be highly correlated. (Moskal, Leydens & Pavelich, 2002)

Alignment in the standards of the examination and CASS should ensure minimal gaps between these marks. However, in addition, the strong correlation referred to above is also sought, implying that the two types of assessment "*truly measure the same concept*", knowledge of the curriculum.

Some examples taken from the 2005 Biology Higher Grade (HG) of the use of the different measures may be pertinent (for ease of presentation, only cases with relatively small classes are shown). Each of Figures 1 to 5 reflects the raw CASS mark and the raw examination marks of individual students in a particular school in Biology HG, to illustrate how CASS marks can send signals of varying clarity to students.

Figure 1 shows a school where the correlation between the CASS and the examination mark was only 0.40. The 18 candidates from this school all performed better in the CASS than in the examination, reflected in all observations being located to the right of the diagonal. If CASS marks were the same as examination marks, all the candidates would have been on the diagonal. But CASS marks exceeded examination marks considerably: The arrow shows that, for Candidate b, the CASS mark should have been 24 marks (percentage points) lower to have corresponded with his/her examination mark.

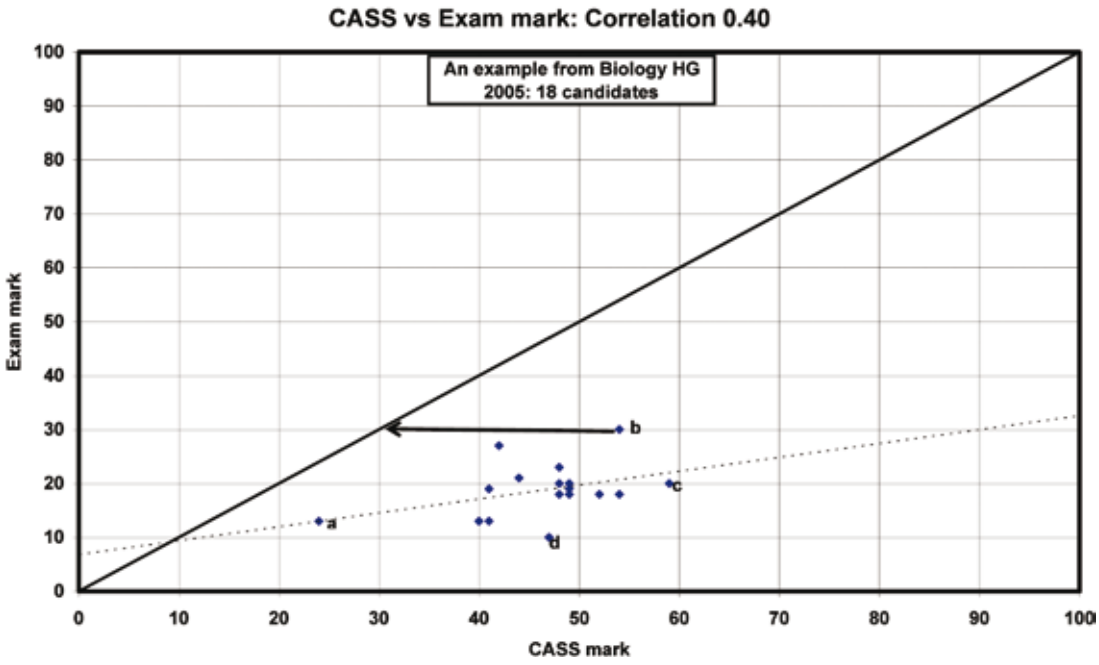
In this figure, there was much deviation around the dotted regression line: If all observations were on this line, or indeed on any other upward sloping line in this figure, it would have implied a perfect correlation (i.e. the correlation coefficient r would have been 1.00)[4]. On such a line, there would

4 If all observations were on a downward sloping line, the correlation would have been –1.0. a perfectly linear negative relationship. Essentially, the correlation measures how well data can be mapped from one variable onto another.

have been a perfect relationship between the CASS and examination marks, so that, ex post, one would have been able to predict examination marks perfectly from the CASS marks. But because candidates did not know how the CASS mark was likely to relate to the examination mark, ex ante, they would have had just the magnitude of the CASS mark as a measure of their performance. Candidate a would indeed have had the correct information, in terms of knowing that his/her prospects for passing Biology HG were very poor. But though all candidates except Candidate a achieved between 40% and 60% in CASS, the best performer in the examination was Candidate b, who achieved only 30% despite a CASS mark of 54%. Candidate c, the best performer in CASS with almost 60%, would have been completely misled, since his/her examination marks was only 20%. Candidate d would also have been misled about relative performance, i.e. what mark to expect in the examination compared to his/her classmates.

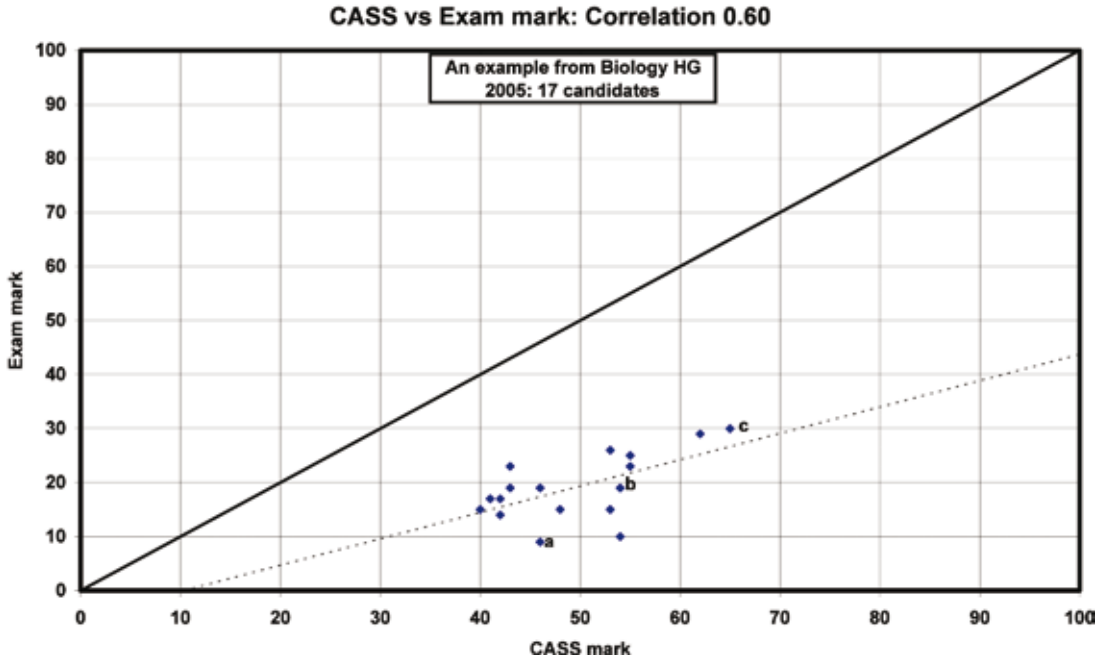
Therefore, the closeness of the observations around the regression line gives an indication of the correlation between CASS and examination marks, i.e. how well the relative ordering of CASS marks matches that of examination marks. In contrast, the deviation of observations from the diagonal indicates how inflated CASS marks are (the lower the observations lie below the diagonal, or the more to the right, the more inflated the CASS marks relative to examination marks).

Figure 1:



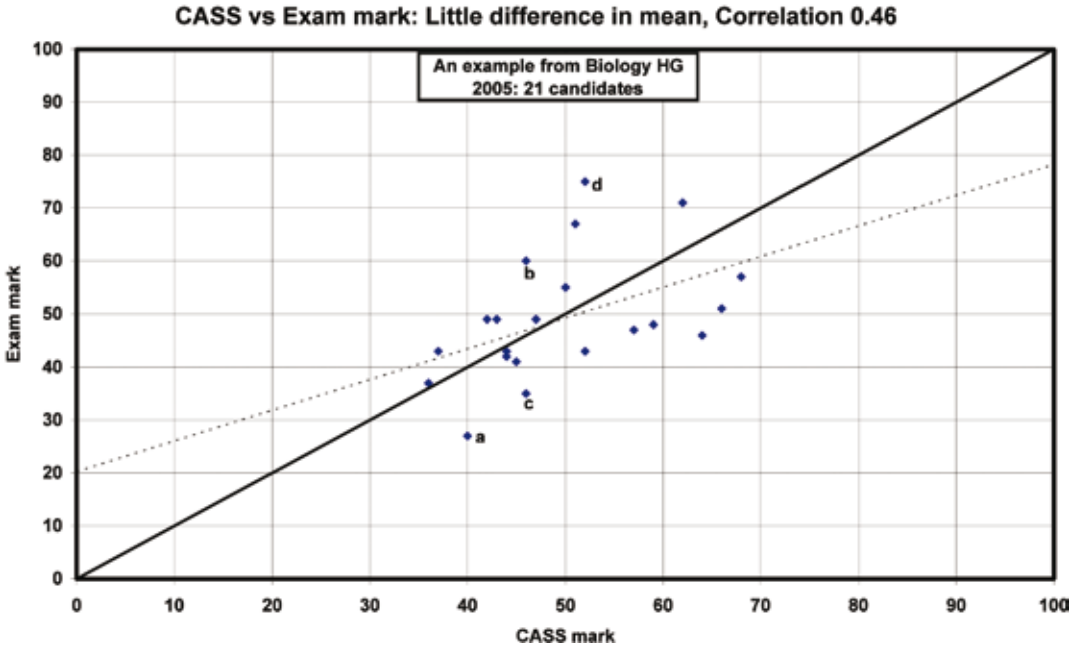
An improvement in the correlation between CASS and examination marks indicates that more reliable information is available to students about relative performance. In **Figure 2** below, the correlation coefficient was somewhat better (r=0.60), as reflected in the observations being more tightly arranged around the regression line, with the result that these 17 candidates had a little more information about what to expect in the examination, in terms of their relative performance within the class. But here, as in the previous figure, the average gap between the CASS and examination marks was again a massive 28 percentage points. With such a large gap, it was possible, as here, that the greater accuracy of relative performance levels still offered little useful information: All candidates failed the examination.

Figure 2:



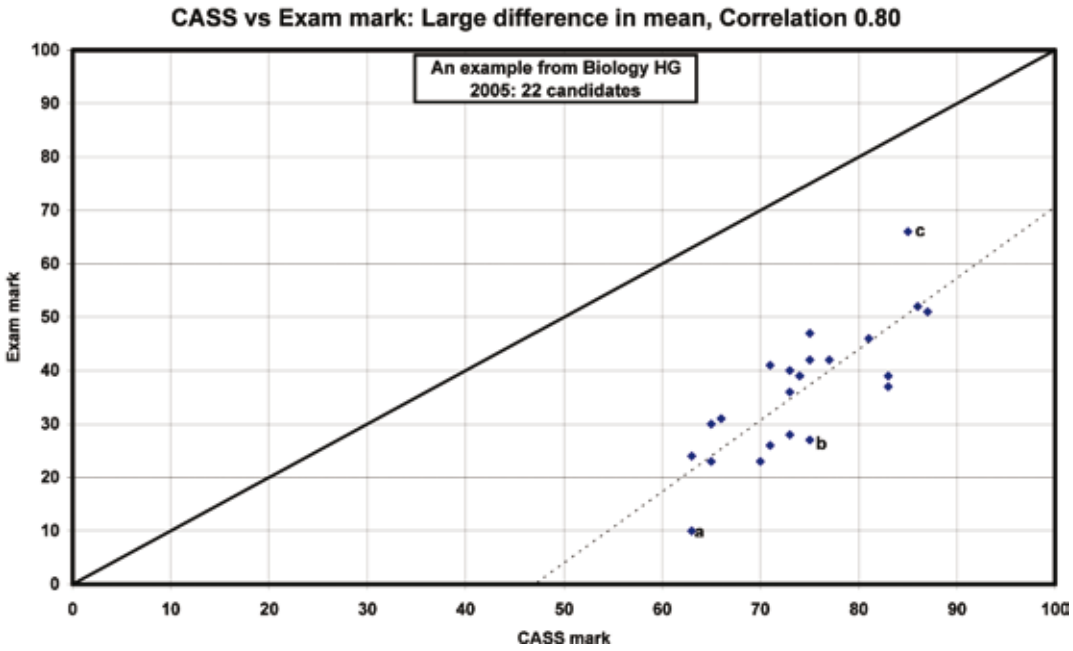
In another school (**Figure 3**), the mean gap between the CASS and examination marks was negligible (less than 1 percentage point). Yet here the relatively weak correlation coefficient (r=0.46) gave little information to students as to their relative performance. Candidate a may perhaps have had an indication that the examination would be a challenge, as this candidate had a low CASS mark and a relative performance (near the bottom of the class) that accorded with this. Indeed the location to the right of the diagonal showed that this candidate had an even worse examination mark. In contrast, there were some observations in this school to the left of the diagonal, candidates who did better in the examination than in CASS. Candidate b was in this position, although his/her CASS mark was not all that different from that of Candidate c. The examination brought different surprises for them, though: While Candidate b obtained a much higher mark, Candidate c obtained a lower mark. Candidate d, meanwhile, rose to the top position and if he/she was better informed, might well have put in enough effort to achieve a distinction (80%): But with a CASS mark of 52%, a 75% examination mark must have come as quite a surprise.

Figure 3:



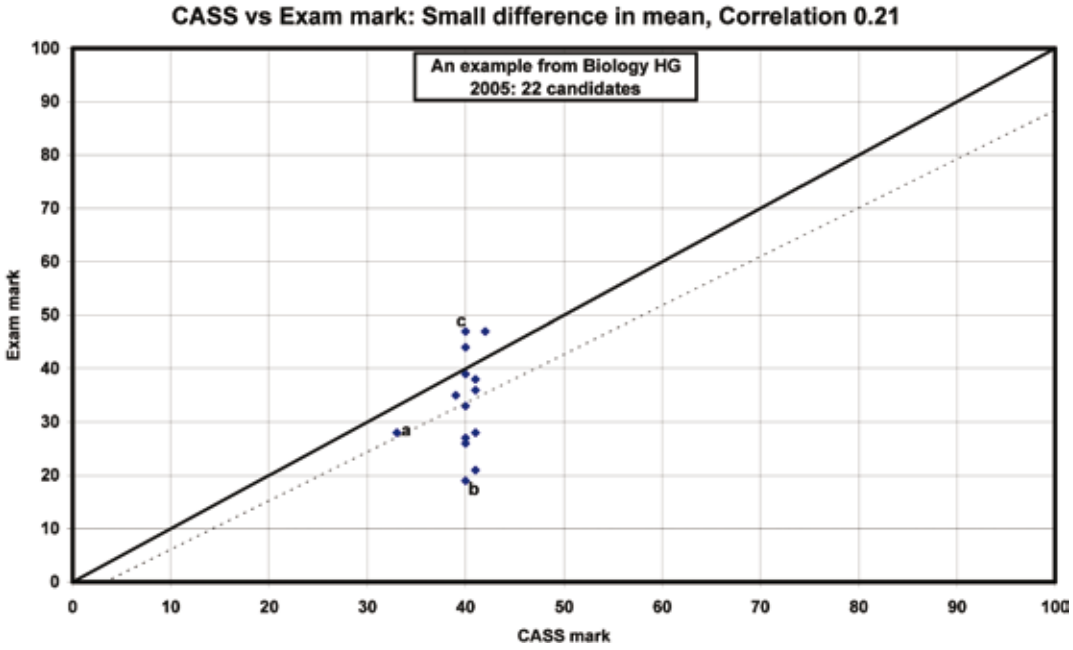
In some cases, relative positions could tell a great deal. Suppose that, in the case of the school in **Figure 4**, candidates in 2005 had received feedback from results of examinations of previous years that indicated that the bottom half of candidates in Biology HG were in danger of failing, then Candidate a and Candidate b should have got a clear message that they were in danger of failing, despite their high CASS marks. In contrast, the high mark of Candidate c, though exaggerated (this candidate also lay considerably to the right of the diagonal), should have signalled to the candidate that he/she was quite safe, though it would have given the false impression that a distinction was within reach.

Figure 4:



In contrast to the above, **Figure 5** below shows a school with little differentiation between candidates in CASS. In this school, 21 out of the 22 candidates had marks between 39% and 42% for CASS. Candidate a, the lone exception with a worse CASS mark of 33%, almost managed to achieve the same mark in the examination. For Candidate b and Candidate c, both with CASS marks of 40%, the examination brought considerably different rewards: 19% and 49%, respectively. Even though the mean result was quite similar for CASS and the examination, the CASS mark was largely bare of information to assist these candidates in preparing for the examination. Such a pattern may indicate that the teacher had little confidence in distinguishing between candidates in CASS marks, perhaps owing to inadequate subject knowledge or a poor understanding of the demands of the curriculum.

Figure 5:

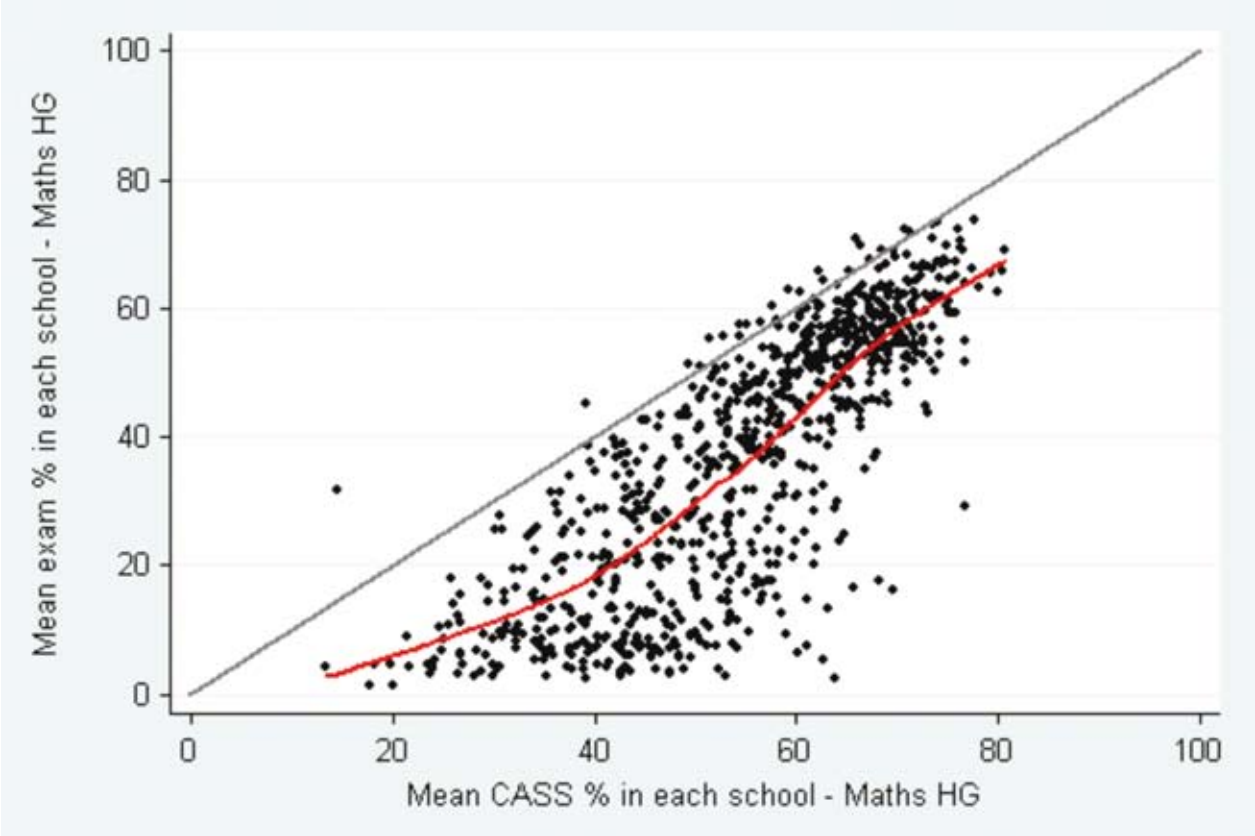


The above examples, all taken from actual situations in Biology HG in 2005, illustrate how important information is, and also that many teachers give weak signals to students. They also show that both the relative mark (as reflected in a good correlation) and a small gap between the CASS and examination marks are important measures of the quality of information available to candidates on their examination prospects. However, with the large weight now given to CASS marks in the examination, teachers may be playing safe by giving exaggerated CASS mark rather than taking the risk of prejudicing the final matriculation results of their students by giving too low CASS marks.[5]

5 Note, however, that where a school's mean assessment mark for a subject exceeds its examination mark by more than 10%, the assessment marks are adjusted downwards so as not to exceed this limit.

Figure 6 shows a similar picture to the foregoing figures, but in this case each dot represents a school with more than 15 candidates who entered the Mathematics HG examination. The majority of schools had CASS marks considerably in excess of their examination marks (i.e. they were located to the right of the diagonal in the figure). The trend line, a Lowess (locally weighted regression) curve indeed also lay considerably to the right of the diagonal. Note, however, that schools in which performance in this subject was higher appeared to have less lenient CASS marks compared to the examination marks: The line moved closer to the diagonal.

Figure 6: School level CASS and examination marks for Mathematics HG 2005, and Lowess (locally weighted) regression trend line



Part 3:

Data and methodology

1 THE DATA

The total dataset obtained from Umalusi consisted of all data for matriculation students of South African high schools from the nine provinces for the years 2003, 2004 and 2005. Student level information provided was gender, race and the raw scores (before adjustment) on school continuous assessment and the matriculation examinations of students for each subject offered by them. At the school level, information on the province, quintile and sector (public or independent schools) was available. Owing to some inconsistencies in coding that made it difficult to assign individual students to their respective schools for 2003 and 2004, North-West province was excluded from all cross-time analysis. Further coding problems meant that adequate assignment of schools to provinces, quintiles and sectors was only possible for 2005.[6] Students for whom no information was available on either or both sets of marks (assessment and matriculation examination) were also excluded, and school-level correlations based on fewer than 15 pairs of observations were dropped. The final data set (excluding North-West) consisted of 5 162, 5 549 and 5 547 schools for 2003, 2004 and 2005 respectively, whilst the 2005 data set including North-West consisted of 5 968 schools. The mean matriculation class size per school was 79, 82 and 87 for 2003, 2004 and 2005 respectively.

2 WHAT IS AN ADEQUATE CORRELATION BETWEEN CASS AND EXAMINATION MARKS?

The correlation coefficient of two variables measures both the strength and direction of the linear relationship between them.[7] It can take a value ranging between +1 (an increasing linear relationship) and –1 (a decreasing linear relationship). Thus a large and positive value correlation between the CASS and examination marks of matriculants in a school could indicate a close positive relationship between the two. Note that correlation in no way implies causation: if a linear relationship exists between X and Y, it cannot be said that X causes Y or vice versa. It is only possible to say that a relationship exists.

For purposes of this study, "school correlations" between assessment and matriculation examination marks were calculated for each of seven subjects[8], distinguishing where appropriate also Higher Grade and Standard Grade subjects, if there were at least 15 candidates in that school and subject. This part of the analysis necessarily did not involve all schools, as many schools had fewer than 15 candidates in some subjects. A weak or even negative school correlation suggested poor reliability of student assessment; in that the school's continuous assessment was poorly matched to the outcomes measured by the matriculation examination. This would be disquieting if it applied to a considerable number of schools.

6 Even then, there was conflicting information in a small number of cases on the quintiles and provinces to which some schools belonged.

7 The most popular method to calculate correlations is the Pearson product-moment correlation coefficient $\rho_{x,y}$, calculated by dividing the covariance of the two variables by the product of their standard deviations. An unbiased estimate of $\rho_{x,y}$ can be calculated using the sample correlation coefficient r as follows:

$$r = \frac{S_{xy}}{\sqrt{(SxxSyy)}}$$

8 These were English (first and second language), Mathematics, History, Biology, Geography and Physical Science (Chemistry and Physics).

Several authors have offered guidelines for interpreting the "strength" of a correlation coefficient. Cohen (1988), for example, suggested the interpretations for correlations in psychological research as shown in **Table 1**. However, such criteria are somewhat arbitrary, and no single correlation value can be used as identifying a sharp cut-off between accurate and inaccurate assessment. Moreover, for the case considered here, negative correlations would unquestionably indicate extremely weak reliability assessment.[9]

Table 1: Strength of correlations

Correlation	Negative	Positive
Small	-0.29 to -0.10	0.10 to 0.29
Medium	-0.49 to -0.30	0.30 to 0.49
Large	-1.00 to -0.50	0.50 to 1.00

Source: Cohen (1988)

As no prior published research had been performed in this area (to the knowledge of the authors), the literature offered no guidance as to what correlation value constituted a strong, positive relationship between the assessment and matriculation examination marks of a particular school. One way of determining an appropriate correlation threshold was to turn to significance testing. For the minimum number of observations used for comparisons, viz. 15, a correlation of +0.513 is sufficient to state with 99% confidence that there is a significant relationship between the two sets of marks, i.e. there is no more than a 1% probability that this correlation would occur by chance if CASS marks were generated randomly.[10]

9 This is similar to using a poverty line in poverty analysis, in that a correlation threshold, like a poverty line, provides a threshold level with which to determine whether schools are assessing poorly or not. This comparison with poverty analysis will be drawn on in using some of the tools of poverty analysis.

10 This follows from the t-value calculated for the correlations, with $t = (r \times n^{½})/[(1-r^2)^{½}]$, with n the sample size and r the correlation. For a one-sided test of significance at $p^2 \geq 0.99$. it turns out that t needs to be 2.650 for n=15 and n-2= 13 degrees of freedom, thus r should exceed 0.513.

As another alternative to determining an appropriate correlation value to test school assessment accuracy, the calculated correlation coefficients were also used to determine a "synthetic" threshold value, in effect allowing the data to provide its own threshold value that was unique to it (i.e. using the actual distribution of class sizes). The t-statistics and accompanying probability values were calculated for all school correlations. These were subsequently categorised as either significant or insignificant (if the probability that such a correlation would be found by chance was set this time as being below 0.1%). This was then plotted in overlaid bar charts such as that shown for Biology in **Figure 7** to offer a visual means of ascertaining where a threshold value of assessment accuracy may lie. Thus one could determine where "accurate" assessment began by observing where the "area of insignificance" ended and the "area of significance" began. These bar charts indicated that the region of "insignificance" ended at a correlation value lying somewhere between 0.4 and 0.6. The latter value will be used in some of the analysis.

Figure 7:

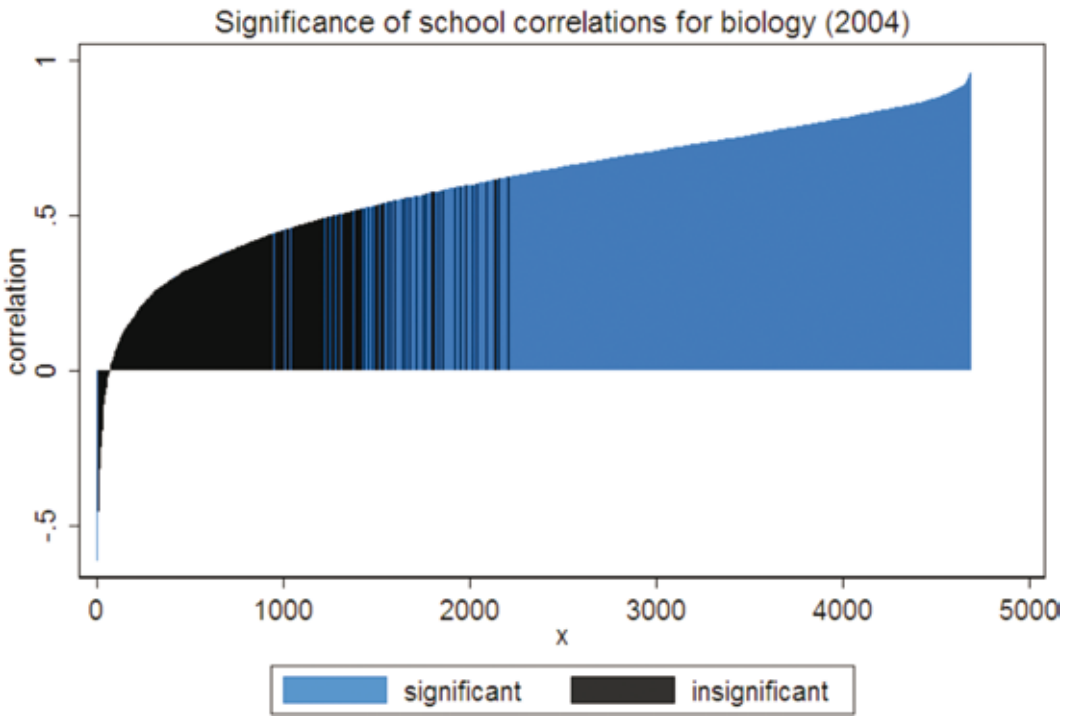


Table 2 offers another clue as to what an adequate correlation between CASS and examination marks is. It is sometimes argued that CASS and examination marks should not be closely correlated, as they are intended to test different things. But underlying both continuous assessment and the examination mark are the same general ability, and the same subject curriculum knowledge that should be reflected in the tests. One would thus, for CASS and examination marks in the same subject, expect a higher correlation than between unrelated subjects. **Table 2** shows pairwise correlations for the same students combining different subjects. Even between such seemingly unrelated subjects as English Second Language and Mathematics SG (two of the subjects most commonly encountered), the correlation was 0.505 in 2005. If such a correlation was found between such unrelated subjects, based only on underlying general ability and motivation of students, one would expect a far higher correlation between the CASS and the examination mark for the same subject, where the same underlying subject knowledge derived from the curriculum was being tested.

Table 2: Correlation coefficients between individual performance in subjects in examinations, 2005 (Number of pupils in parentheses)

	Biology HG	Biology SG	English 1st	English 2nd	Geogr HG	Geogr SG	History HG	History SG	Maths HG	Maths SG
English 1st language	0.728 (26990)	0.542 (13270)								
English 2nd language	0.805 (84001)	0.620 (190290)								
Geography HG	0.853 (47608)	0.754 (48884)	0.745 (17044)	0.616 (6033)						
Geography SG	0.745 (8067)	0.666 (90487)	0.715 (93156)	0.580 (95425)						
History HG	0.830 (11786)	0.662 (19194)	0.735 (8620)	0.494 (3453)	0.794 (16677)	0.657 (7504)				
History SG	0.689 (4165)	0.560 (54766)	0.615 (28693)	0.488 (58597)	0.625 (9905)	0.608 (39415)				
Mathematics HG	0.852 (23963)	0.706 (5596)	0.573 (15206)	0.518 (31906)	0.768 (11856)	0.553 (1718)	0.596 (1986)	0.595 (87)		
Mathematics SG	0.745 (62490)	0.613 (94204)	0.728 (28205)	0.505 (213739)	0.622 (46807)	0.432 (37209)	0.647 (5937)	0.481 (3348)		
Physical Science HG	0.873 (35928)	0.740 (13695)	0.624 (16890)	0.525 (12271)	0.766 (19843)	0.607 (2628)	0.600 (1695)	0.650 (57)	0.915 (30073)	0.813 (31888)
Physical Science SG	0.794 (34779)	0.694 (66899)	0.730 (45092)	0.551 (100927)	0.648 (22657)	0.523 (26723)	0.577 (1178)	0.551 (726)	0.835 (5696)	0.785 (110494)

As indicated, when the scores for the two assessments are totally unrelated to one another, the correlation will be approximately zero (r=0). This would imply that the CASS score would be useless for predicting the examination score. Knowing the CASS score for a particular student would give no information about the possible values of the examination score. At another extreme, if the CASS score (X) was to be perfectly related to the examination score (Y), by a simple linear formula of the type Y= a + b.X, the correlation coefficient would be (approximately) one (i.e. r=1). Here the same value of a and b would apply for every candidate, and b would be a positive number. This would imply that once the information is known for the CASS mark (X), the examination mark (Y) would be perfectly predictable and would not provide further insight into the performance of the candidates.

In general, when dealing with examination data, having some additional information of a similar type so as to more properly nuance the final allocation of marks, seems appropriate and desirable.

While the correlation coefficient is a measure of the similarity of the paired data values, a related quantity (the square root of $(1-r^2)$) is a measure of the natural variation in scores remaining after using one of the variables to make a best guess rule of the type Y= a + b.X, to estimate the Y-value for a specific X-value. Specifically, the question is what variation one should expect for the examination marks between students who had similar marks for CASS. The specific question is what sort of values of the correlation coefficient might be useful in an education setting. **Table 3** presents the levels of correlation coefficient required to attain a nominated percentage of the original variability.

Table 3: Shrinkage factors and the correlation coefficients required to obtain such shrinkage relative to a random relationship between CASS and examination marks

Standard deviation of examination mark set at 1	Shrinkage factor (Std deviation of examination mark, given CASS mark)	Shrinkage percentage	r^2	Correlation coefficient
1	1	100%	0	0
1	0.9	90%	0.19	0.436
1	0.8	80%	0.36	0.600
1	0.7	70%	0.51	0.714
1	0.6	60%	0.64	0.800
1	0.5	50%	0.75	0.866
1	0.4	40%	0.84	0.917
1	0.3	30%	0.91	0.954
1	0.2	20%	0.96	0.980
1	0.1	10%	0.99	0.995
1	0.0	0%	1.00	1.000

Thus to shrink the variability in the examination mark for a given CASS mark to 50% of its natural spread (where there is no correlation, i.e. the variability is completely random), we need a value of r = 0.866, but to shrink it to 20% of its original extent, the necessary value is r=0.980. As an informal rule of thumb, correlations below 0.75 offer only flimsy evidence of any meaningful relationship.

This can be illustrated by plotting data sets in which the pairs of examination marks and CASS marks pairs have been artificially constructed to have correlation coefficients ranging from 0 to 1. If CASS marks were completely uncorrelated with the examination mark, the situation may have looked as in **Figure 8a**. The 1 000 observations for this figure were generated to reflect a situation where the mean gap between examination and CASS marks is zero, i.e. on average, there is no inflation of CASS marks, and with a mean mark of 50 and a standard generation of 16 for each of the assessment methods. A slightly better relationship is observed in **Figure 8b**, where the coefficient of determination (R-squared) is 0.2, and the correlation 0.447. In this case, the variation in the examination marks for given levels of the CASS mark is reduced to 89.4% of the levels in **Figure 8b**, i.e. a "shrinkage factor" of 0.894 is applied to the variance in examination marks, given CASS marks. As can be seen, this still reflects a very small improvement. If the variability is shrunk further, to 77.5% of its original value, the R-squared would improve to 0.4, as in **Figure 8c**, i.e. 40% of the variance in the examination marks would be explained by the CASS mark. This already requires a correlation coefficient r of 0.632, though, as can be seen in the figure, it still implies much randomness of the examination mark around the CASS mark (i.e. around the diagonal). Further shrinkage of the variability of examination marks to 63.2% of its original value (as in **Figure 8d**) yields an R-squared of 0.6 and therefore requires a correlation of 0.775, whilst reducing the variability even further to 44.7% of its original value, to obtain an R-squared of 0-.8, requires a correlation of 0.894. Even this leaves a fair amount of variability of the examination mark around the CASS mark, and is still a far cry from the perfect match between the assessments that would result if both the R-squared and the correlation coefficient were 1, i.e. where variability is shrunk to 0% of its original value, and all observations lie on the diagonal.

Figure 8a:

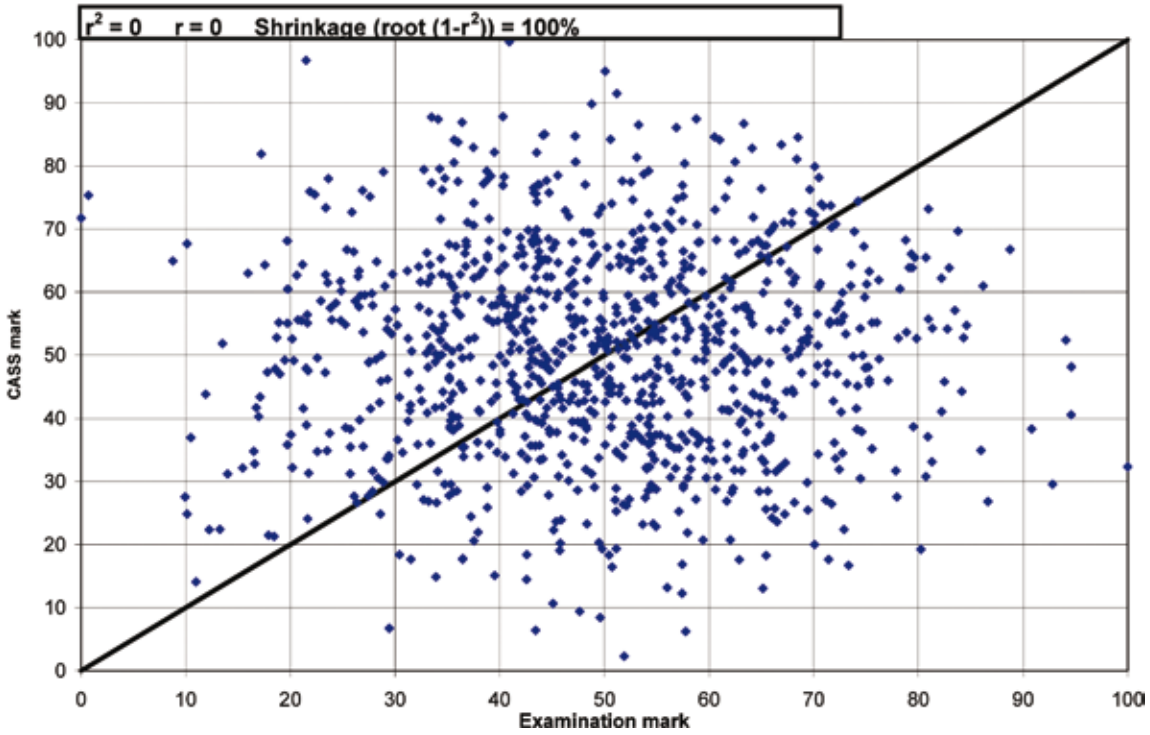


Figure 8b:

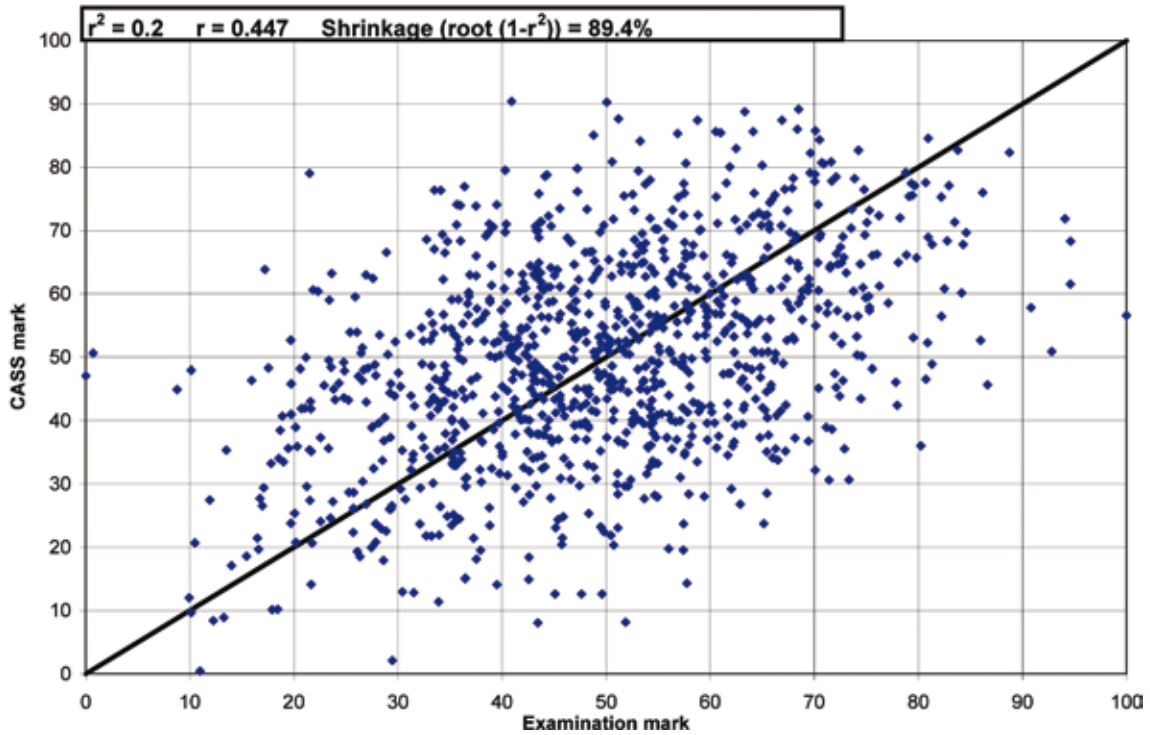


Figure 8c:

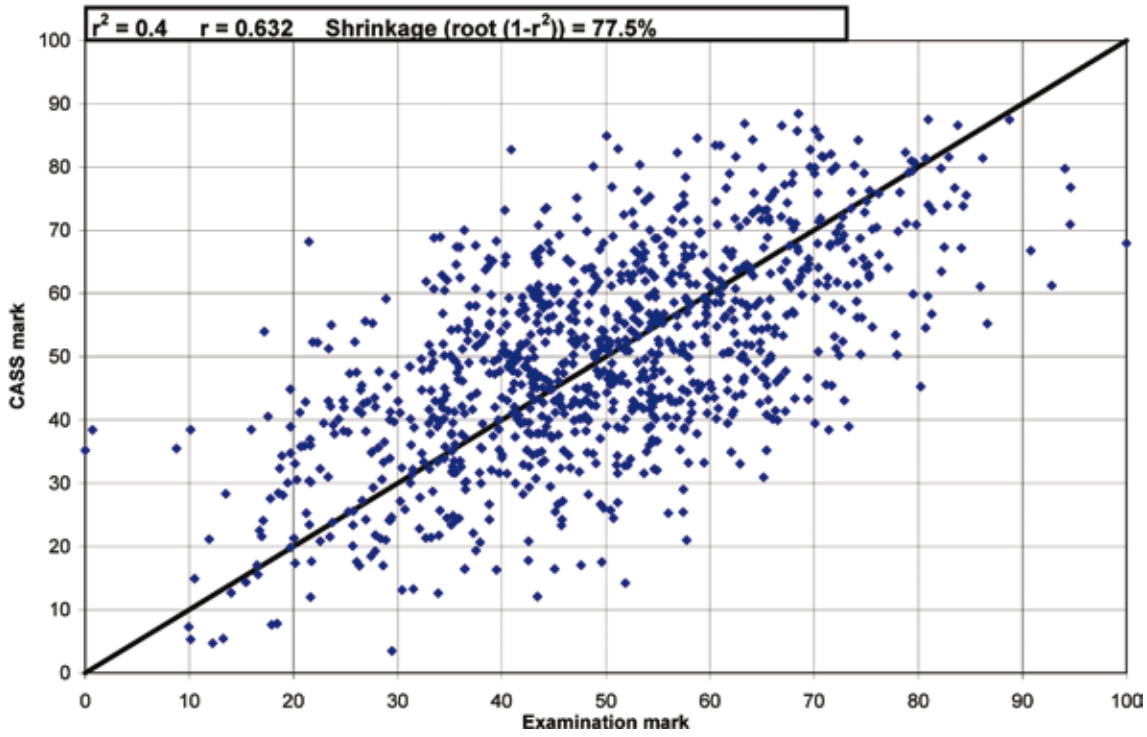


Figure 8d:

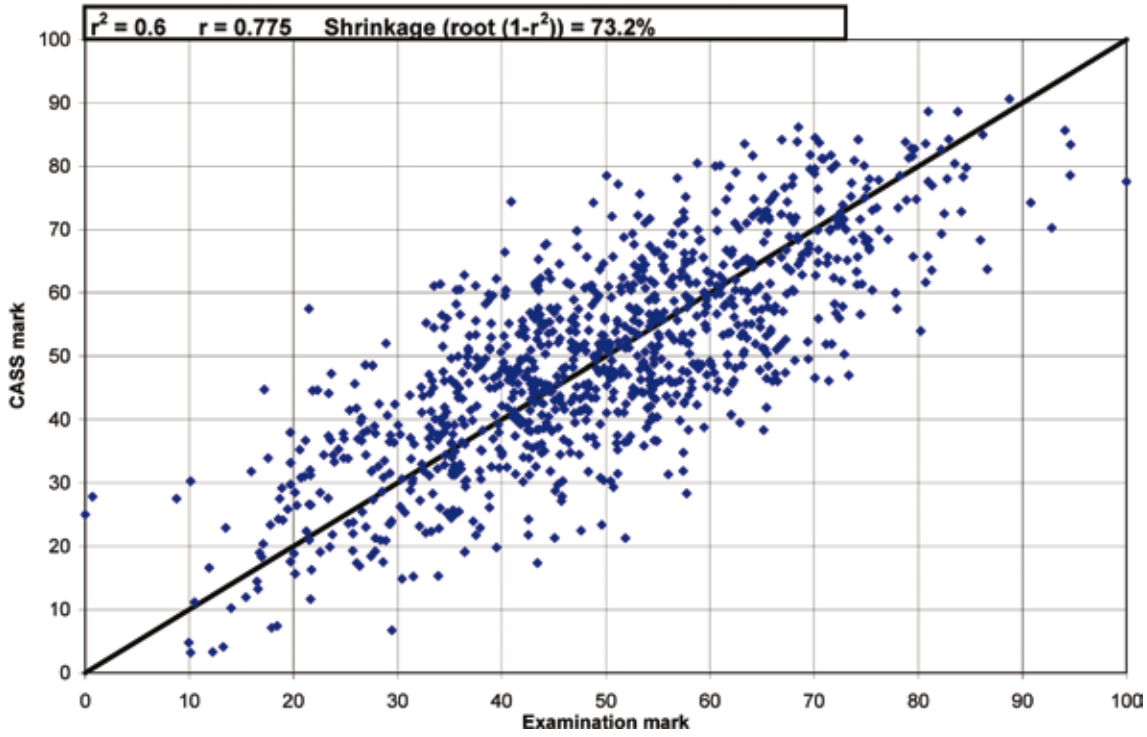
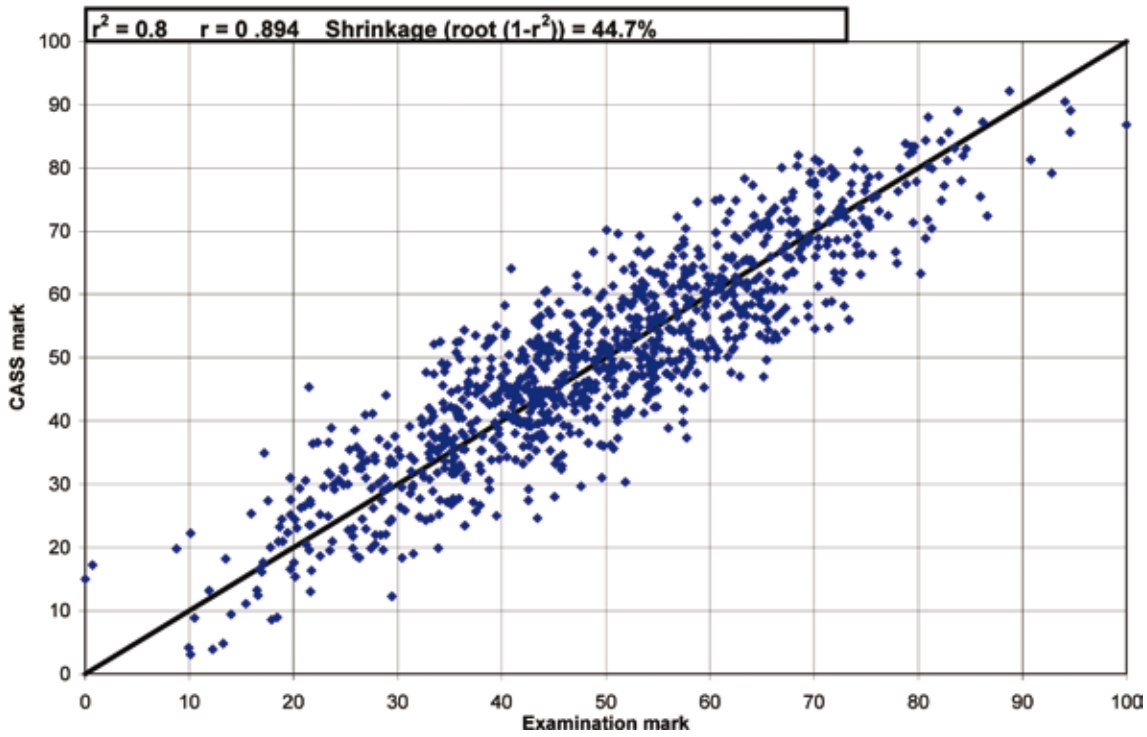


Figure 8e:



It would appear reasonable to consider any shrinkage less than 50% as inadequate. Where the shrinkage factor is 50%, the variability in the examination marks around the CASS marks is still half as large as it would have been if examination marks were purely randomly assigned to candidates with given CASS marks, i.e. the CASS mark still does not provide a very clear signal and contains much "noise" (randomness). But as **Table 3** shows, a relatively large correlation (0.866) is required to achieve this level of shrinkage of the variability in CASS marks.

In the light of the above, using correlation coefficients as low as 0.6 to separate "strong" from "weaker" correlations may thus be too lax a yardstick for assessment accuracy. Far higher correlations should hold in a system that assesses well. As will be shown later (e.g. in **Figure 9**), in better performing schools and in most subjects, correlations of above 0.8 are quite standard.

Part 4:

Individual CASS and examination marks

1 INDIVIDUAL CASS AND EXAMINATION MARKS BY SUBJECT AND PROVINCE

Table 4 (**attached as an Appendix**) summarises the means and standard deviations of the assessment and examination marks of all individuals in South African high schools for 2005, by province and subject. (Note that these results include North-West Province.) For each individual student, the actual mark in the examination was compared to the mark obtained for the same subject in the continuous assessment. For every subject and in all provinces, the mean CASS mark was consistently and substantially higher than the examination mark. The fact that the standard deviations of the examination marks were in most cases slightly larger than those of the assessment marks suggested some uncertainty amongst assessors. Perhaps many were "playing it safe", i.e. giving similar marks to high and low performances, a strategy that may be indicative of uncertainty resulting from inadequate subject knowledge.

A few results are worth mentioning. For English First Language there was a surprisingly small gap between the assessment and examination marks for KwaZulu-Natal of only 1 percentage point, but only a relatively smaller share of students took this subject. Most gaps were larger than 10 percentage points, sometimes much larger. For example, in Mpumalanga many subject gaps were in excess of 20 percentage points. Considerable gaps were observed for most provinces in Biology and Physical Science, whereas gaps for Geography and English were relatively smaller. Cumulative density functions[11] (not shown) indicated that no matter what threshold was chosen for the gap value, all other subjects first-order dominated Science and Biology in 2005, i.e. no matter what magnitude of the gap one looked at, large gaps were more frequent for these two subjects.

2 TRENDS AT INDIVIDUAL LEVEL IN CASS AND EXAMINATION MARKS AND THE ASSESSMENT GAP

One would expect a gradual closing gap between the CASS and examination marks, as teachers adjust their continuous assessments based on feedback from previous examinations, to bring it more closely in line with the curriculum standards as indicated by the examination mark. This did not occur. On the contrary, the widening gaps in **Table 5** rather provided evidence of declining examination marks in the majority of cases, with a relatively stable assessment mark. Small reductions of 1 to 2 percentage points in some of the CASS marks were more than made up for by relatively large falls in the examination marks, which may have resulted from examinations becoming more rigorous or from the declining of the average quality and preparation of matriculation candidates. Declining examination marks were indicative of an increasing disparity between what was being taught and assessed within the schools, and what was being tested in the examination (in line with the national curriculum). The sharp rise in the mean gap for History HG (from 1.7 to 15.7) and Mathematics HG (8.6 to 18.8) resulted from a very large drop in the examination marks for these subjects. Only for Geography SG did the gap close significantly, and that because of a rise in the examination mark.

11 The probability distribution can be represented by a cumulative distribution function (CDF) that can be used to ascertain whether one distribution of correlations first-order stochastically dominates another (Madden & Smith, 2000) or, in this case, whether one distribution of correlations is statistically worse than another, irrespective of the criterion value set. If the CDF of Mathematics mark correlations within schools lies below that for Biology for all possible correlation values, then relative assessment reliability in Mathematics between schools is always better than in Biology irrespective of what threshold correlation value has been chosen; that is, the proportion of schools who are assessing their students unreliably will always be smaller in Mathematics than for Biology, irrespective of the criterion threshold chosen. However, if two CDFs intersect, ranking becomes ambiguous and it is unclear whether or not one subject is more reliably assessed than another. The ranking of the assessment accuracy for these two subjects, for instance, will then depend on what criterion one sets for the correlations – technically, first-order dominance then no longer holds (Madden & Smith, 2000). The threshold correlation value is similar to the poverty line employed in poverty analysis, e.g. using the class of poverty measures devised by Foster, Greer and Thorbecke (1984).

Inspection of kernel density distributions[12] of these individual gaps for 2003, 2004 and 2005 showed distributions that were indeed steeper than a normal distribution (i.e. with a higher kurtosis), with a positive mean, indicating that most students achieved a much higher assessment mark than the matriculation examination mark. The distributions also appeared to have been shifting rightwards from 2003 to 2005, implying that the divergence between examination and school assessment worsened for the average matriculant.

The standard deviations of the gap were in most cases relatively large. For the more exact subjects one would have expected a smaller gap between the CASS and examination marks, and smaller standard deviations of this gap. However, surprisingly the gap for English (Second Language and especially First Language), traditionally regarded as a "less exact discipline", was smaller and showed even less variance than for the more exact discipline, Mathematics. This may have reflected a high level of consensus amongst English teachers about the standards at which they expected students to perform, or that the curriculum was more highly specified, or both.

Table 5: Means and standard deviations of individual assessment and examination marks by subject and grade, 2003-2005 (standard deviations in parentheses)

		CASS Marks			Examination marks			Gap		
		2003	**2004**	**2005**	**2003**	**2004**	**2005**	**2003**	**2004**	**2005**
Biology	**HG**	57.7	57.3	55.9	39.8	35.6	35.1	17.9	21.7	20.8
		(15.9)	(15.3)	(15.2)	(20.7)	(19.0)	(18.3)	(15.5)	(15.2)	(15.3)
Biology	**SG**	45.5	45.8	45.6	30.5	27.2	25.6	15.0	18.5)	20.0
		(13.6)	(13.6)	(13.4)	(13.1)	(13.1)	(12.8)	(13.3)	(14.1)	(14.3)
English 1st language	**HG**	59.0	59.0	58.4	54.0	52.3	52.0	5.0	6.7	6.4
		(12.5)	(12.7)	(12.5)	(14.2)	(14.1)	(14.1)	(9.0)	(8.9)	(8.9)
English 2nd language	**HG**	49.5	49.3	48.9	41.4	36.3	36.2	8.1	13.0	12.7
	HG	(13.3)	(12.9)	(12.9)	(14.3)	(12.6)	(12.8)	(10.5)	(9.9)	(10.2)
Geography	**HG**	46.7	46.2	45.0	36.6	36.4	34.7	10.0	9.8	10.3
	HG	(15.2)	(14.7)	(14.0)	(17.9)	(18.7)	(17.7)	(11.8)	(12.3)	(11.8)
Geography	**SG**	41.6	41.2	41.1	34.8	35.7	36.9	6.8	5.5	4.2
		(12.4)	(12.0)	(11.9)	(14.2)	(13.8)	(14.4)	(12.6)	(12.0)	(12.7)
History	**HG**	50.5	52.3	51.4	48.8	39.6	35.9	1.7	12.7	15.5
		(16.8)	(15.7)	(15.6)	(20.7)	(20.0)	(18.3)	(14.7)	(15.2)	(14.5)
History	**SG**	43.1	43.8	42.8	39.6	35.5	30.9	3.5	8.4	11.9
		(13.1)	(12.8)	(12.8)	(15.9)	(15.4)	(14.4)	(15.4)	(14.6)	(14.0)
Mathematics	**HG**	55.5	56.6	54.5	46.9	42.0	35.9	8.6	14.6	18.6
		(19.9)	(17.5)	(17.8)	(24.6)	(23.5)	(24.3)	(12.4)	(13.8)	(14.5)
Mathematics	**SG**	38.9	40.6	40.1	27.1	27.8	25.8	11.7	12.8	14.4
		(17.7)	(16.1)	(15.9)	(20.5)	(20.2)	(19.7)	(12.6)	(13.2)	(13.5)
Science	**HG**	55.7	54.1	52.8	38.4	33.8	30.8	17.3	20.3	22.0
		(15.2)	(14.9)	(14.7)	(20.1)	(18.6)	(17.6)	(13.0)	(12.1)	(12.2)
Science	**SG**	45.1	44.1	43.8	31.6	29.7	27.6	13.5	14.4	16.2
		(12.8)	(12.4)	(12.0)	(12.8)	(11.6)	(11.5)	(11.8)	(11.2)	(11.3)

12 Kernel density methods are often used to estimate the probability density function of a random variable. The kernel density estimator is sensitive to the choice of bandwidth, as too large a bandwidth can lead to over smoothing. (Davidson & MacKinnon, 2004). The kernel densities for different distributions of correlations were compared for years, subjects, provinces and quintiles.

The large standard deviations for most other subjects were a cause for concern, pointing to an inconsistent understanding amongst teachers of the level of performance required of students in these subjects. These difficulties may either have stemmed from the learning/teaching process or from the evaluative judgments made when assessing learners, or both.

Though the gap between assessment and examination marks was large throughout the spectrum, it was particularly large for those with very high and very low assessment marks. Also interesting is that in almost a quarter of cases where candidates had achieved 50% or above for one of these subjects in CASS, their examination marks lay below 30%.

Umalusi has noted the large gaps between CASS and examination marks and, in its moderation, imposed a limit of a 10-percentage-point deviation between these two. Where the mean CASS mark for a subject in a school deviated by more than 10 marks from the examination mark, all CASS marks were adjusted to reduce the mean gap to 10 marks. **Table 6** shows that the share of matriculants who achieved an examination mark of more than 10% lower than their school assessment marks (equivalent to a difference of one letter symbol) had increased for all subjects except Geography. It is also evident that increases were larger at the Higher Grade level. The proportion of History HG students achieving more than 10 percentage points lower in the matriculation examination than in CASS more than doubled over this short period, from 26.7% to 62.7%, and Mathematics HG too experienced an increase of almost 30 percentage points in this proportion. It is a cause for concern that students were being assessed at lower levels in the school tests and examinations, particularly at the Higher Grade level, as matriculants who took Higher Grade subjects were often academically more ambitious, aiming at matriculation endorsement (a requirement for university studies). If they were too leniently assessed in school, they were being given false signals of their performance potential in the final matriculation examinations.

Table 6: Share of matriculants with a gap of more than 10%, by subject and in aggregate (% of all candidates)

		2003	2004	2005
Biology	**SG**	62.6	71.3	74.2
	HG	64.3	75.6	72.7
English First Language	**All**	24.6	29.9	29.2
English Second Language	**All**	41.4	60.1	58.5
Geography	**SG**	38.1	33.2	31.1
	HG	50.0	48.5	50.1
History	**SG**	32.8	43.8	53.8
	HG	26.7	54.0	62.7
Mathematics	**SG**	53.9	56.1	60.7
	HG	39.8	56.8	68.3
Science	**SG**	59.0	63.8	69.6
	HG	67.6	78.7	82.3
Aggregate	**All**	43.0	57.0	59.9

3 INDIVIDUAL LEVEL GAPS AND CORRELATIONS BY SUBJECT, 2005

Table 7 again displays the mean gaps, alongside the mean correlations between CASS and examination marks of all individual candidates by subject.[13] The table shows that there were few subjects where there were both high correlations between CASS and examinations marks and a small gap between the two marks. One important exception was English First Language, with a gap of only 6.4 percentage points and a correlation reaching almost 80% (r=0.78). This could perhaps have been attributed to the fact that this subject was elected by only a relatively small group of students, mainly concentrated in historically more privileged schools. But English Second Language, taken by a large group of students from heterogeneous but often poor backgrounds, also had both high correlations (0.69) and a small mean gap (12.7 percentage points). It was less surprising that correlations for an exact discipline like Mathematics were very high in both Higher Grade and Standard Grade – but then, the gaps between the CASS and examination marks were unexpectedly wide. Judged on these figures, the worst assessments were those in Biology (particularly SG) and Physical Science SG, although there were also serious problems of assessment in History SG. The picture in Geography SG was mixed, with a weak correlation but a smallish gap.

Table 7: Mean gap and correlation between CASS and examination marks of individual candidates, 2005

		Mean gap	Correlation
Biology	**HG**	20.8	0.596
	SG	20.0	0.409
English First Language	**HG**	6.4	0.782
English Second Language	**HG**	12.7	0.688
Geography	**HG**	10.3	0.745
	SG	4.2	0.547
History	**HG**	15.5	0.643
	SG	11.9	0.477
Mathematics	**HG**	18.6	0.808
	SG	14.4	0.730
Physical Science	**HG**	22.0	0.729
	SG	16.2	0.541

13 In subjects with a large variance in the assessment gaps between different schools, one would expect a lower correlation between individual CASS and examination marks. In such cases, some schools would gain much in their CASS relative to their examination marks, but other schools not, dampening the overall CASS/examination marks relationship.

Part 5:

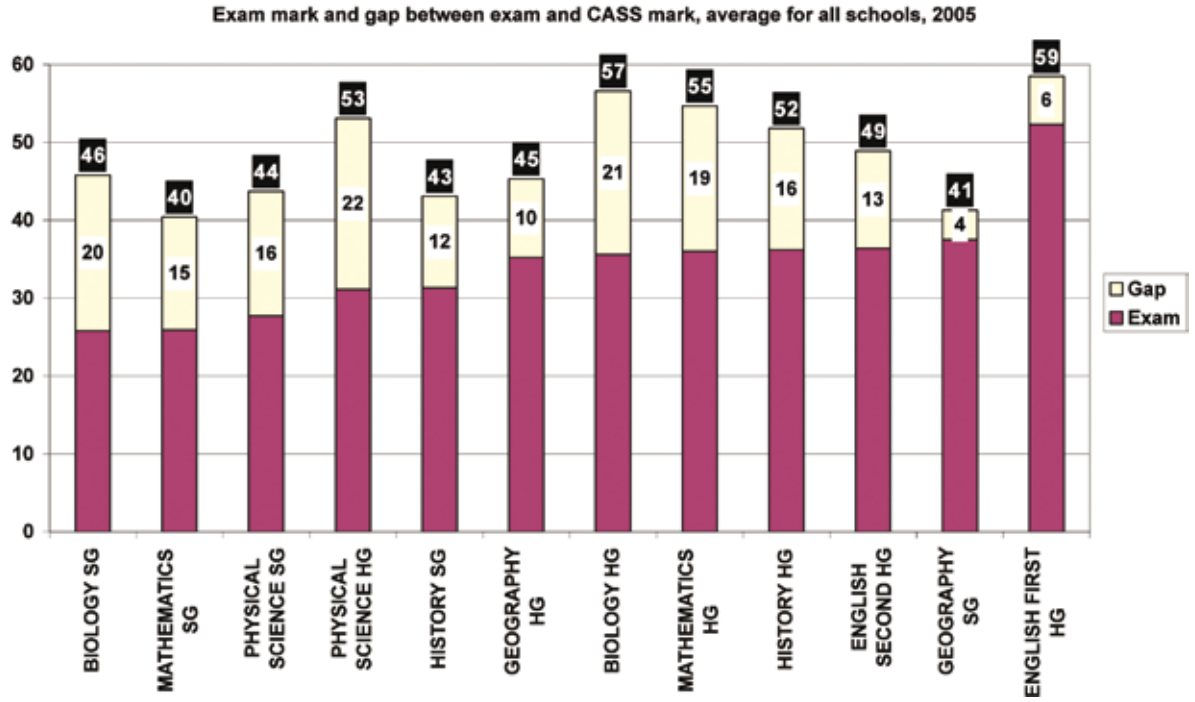
School level assessment

Thus far, analysis was confined to data at the individual level. To determine how well assessment takes place in individual schools and classrooms, it is necessary to aggregate within schools.[14] Unless specified differently, the analysis from this point onwards is *at the level of the individual school, not weighted by the number of candidates in each school*. Essentially, the picture obtained is intended to ascertain how accurately teachers assess.[15] [16]

1 GAPS AT THE SCHOOL LEVEL BY SUBJECT, 2005

Figure 9 shows that the smallest gaps between CASS and examination marks within schools related to Geography SG followed by English First Language, which was also the subject with the highest mean examination mark across schools. Large gaps occurred for some Higher Grade subjects (Science 22, Biology 21, Mathematics 19) and for Biology SG.

Figure 9:



14 Strictly speaking, this analysis is at the level of examination centres, which may also contain some private candidates. However, in practice only a small percentage of all candidates are private candidates doing the matriculation examination..
15 In some cases, more than one teacher may have been involved in assessment in the same school, but assessment across classes within the same school was likely to be relatively consistent as teachers would have tended to share assessment across classes.
16 In order to derive stable data, schools were dropped from the analysis for individual subjects where fewer than 15 students entered for examination in that subject.

2 INTRA-SCHOOL CORRELATIONS BETWEEN CASS AND EXAMINATION MARKS AT THE SCHOOL LEVEL BY SUBJECT

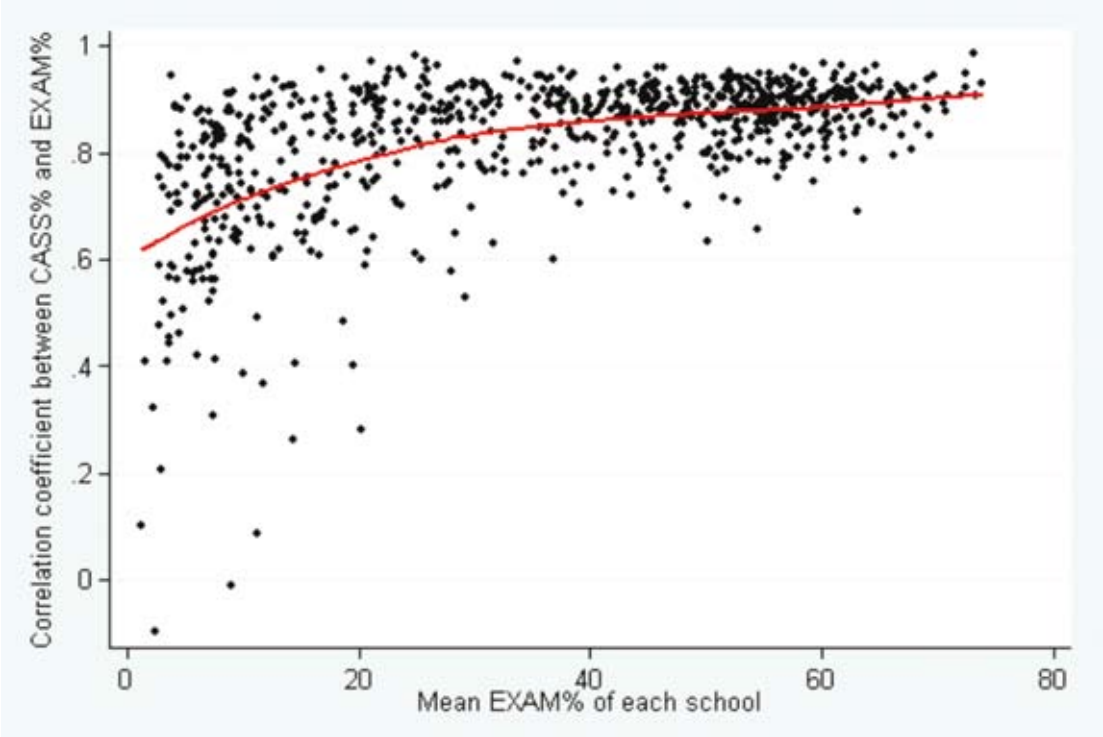
Table 8 shows mean subject correlations for South African schools. Subject correlations in some cases had large standard deviations, perhaps largely owing to small numbers of candidates because of subject choice (only some candidates in a school took a particular subject; where there were fewer than 15 candidates in a subject, such a school was not considered in the analysis). There appeared to be stability in the means and standard deviations over the period, with slight improvements in the reliability of assessment within schools of Geography SG and History SG and slight deterioration in that for Biology HG. Mathematics, especially at Higher Grade level, had the highest average correlation within schools (as was the case for all individuals), as well as the smallest standard deviation and thus a larger frequency of more reliable assessment. However, the share of schools where CASS marks showed low reliability also increased, as evident from the increase in the standard deviation of the school correlations for Mathematics. Schools appeared to assess the two English subjects reliably, with distributions of correlations within schools almost similar to that in Mathematics. History SG was the least reliably assessed subject, with a mean correlation over schools of around 0.5. It would, therefore, appear that a substantial number of schools do not assess this subject reliably, although there has been some improvement. This, combined with the fast widening gap between mean CASS and examination marks for History discussed earlier, suggested that assessment accuracy in some schools was even further declining.

Table 8: Means and standard deviations of within-school correlations between CASS and examination mark by subject and grade, 2003-2005

Subject	Grade	Mean correlations			Standard deviation of correlations		
		2003	2004	2005	2003	2004	2005
Biology	**HG**	0.70	0.68	0.67	0.19	0.20	0.21
	SG	0.59	0.58	0.59	0.19	0.20	0.21
English First Language		0.75	0.74	0.75	0.16	0.16	0.16
English Second Language		0.71	0.69	0.69	0.16	0.17	0.18
Geography	**HG**	0.70	0.71	0.72	0.16	0.16	0.17
	SG	0.57	0.59	0.61	0.20	0.19	0.20
History	**HG**	0.64	0.61	0.63	0.20	0.20	0.20
	SG	0.49	0.49	0.53	0.21	0.21	0.20
Mathematics	**HG**	0.86	0.84	0.83	0.10	0.13	0.13
	SG	0.82	0.80	0.79	0.13	0.14	0.15
Science	**HG**	0.74	0.76	0.74	0.19	0.17	0.18
	SG	0.64	0.65	0.65	0.19	0.18	0.18

Figure 10 shows a Lowess regression that indicated, for Mathematics HG, the relationship between a school's mean examination mark and the correlation between the CASS and examination marks for that subject, a measure of assessment reliability. As can be observed, schools performing better in the examination also tended to have more reliable assessment – more specifically, unreliable assessment was only common amongst some schools with low performance in the examination. This figure appears to support a working hypothesis that unreliability of CASS marks may have been associated with weak examination preparation.

Figure 10: The relationship between schools' examination marks and the correlation between CASS and examination marks for Mathematics HG, 2005



3 RELIABILITY AND LENIENCY OF ASSESSMENT WITHIN SCHOOLS BY SUBJECT AND SCHOOL QUINTILE OR TYPE

Given clear divergences in assessment accuracy of schools by province and by subject, it is interesting to know whether thus accuracy (measuring both reliability and leniency) also differed by school poverty level and by school type (public versus independent schools). **Table 9** shows mean school-level correlations and gaps by subject and grade for 2005 by province, **Table 10** by school quintile, and **Table 11** by school type. Poor assessment values by either criterion (mean gap above 20, or school level correlation below 0.60) are highlighted in the tables.

There was a major problem in assessment in Mpumalanga but, perhaps more surprisingly, the Northern Cape also showed relatively poor assessment in terms of these two sets of criteria separately. This analysis also confirmed that there were widespread problems in schools in assessing accurately in Biology SG and, to an even greater extent, in History SG across provinces: The very weak mean correlations encountered here across schools in all provinces, implying poor reliability between CASS and examination marks, raised the suspicion that teacher subject knowledge might have been deficient in schools offering these two subjects, that the curriculum might have been poorly specified or that teachers might not have been setting assessment tasks that could act as good preparation for the examination. Geography HG also bore further investigation.

Table 9: Mean school level correlations and gaps between CASS and examination marks by province, 2005

	Western Cape	Northern Cape	Free State	Eastern Cape	KwaZulu-Natal	Northwest	Mpumalanga	Gauteng	Limpopo	SA
Mean school level correlation between CASS and examination marks										
Biology HG	0.75	0.72	0.82	0.76	0.69	0.64	0.56	0.76	0.69	0.67
Biology SG	0.61	0.54	0.73	0.68	0.56	0.52	0.49	0.65	0.59	0.59
English 1st	0.77	0.79	0.76	0.72	0.84	0.77	0.71	0.74	0.76	0.77
English 2nd	0.77	0.80	0.74	0.65	0.66	0.70	0.71	0.73	0.66	0.69
Geography HG	0.54	0.51	0.66	0.63	0.63	0.58	0.59	0.66	0.53	0.61
Geography SG	0.73	0.70	0.80	0.72	0.76	0.72	0.69	0.76	0.67	0.72
History HG	0.73	0.73	0.59	0.61	0.66	0.59	0.56	0.67	0.63	0.63
History SG	0.54	0.50	0.48	0.52	0.54	0.56	0.49	0.52	0.54	0.53
Mathematics HG	0.86	0.89	0.86	0.85	0.85	0.80	0.76	0.85	0.86	0.83
Mathematics SG	0.82	0.82	0.85	0.77	0.79	0.77	0.75	0.82	0.81	0.79
Science HG	0.84	0.75	0.84	0.77	0.74	0.69	0.68	0.77	0.75	0.74
Science SG	0.71	0.70	0.70	0.65	0.64	0.60	0.63	0.65	0.63	0.65
Mean gap between CASS and examination marks										
Biology HG	8.0	18.8	10.5	10.6	20.9	18.4	32.6	16.8	17.1	21.3
Biology SG	11.5	25.2	11.5	12.8	23.0	19.3	29.4	19.3	18.1	20.1
English 1st	8.1	4.0	10.4	13.0	2.9	14.2	19.3	6.6	12.0	7.9
English 2nd	7.9	7.3	8.8	12.4	16.6	13.6	15.0	8.0	11.6	13.2
Geography HG	1.8	8.2	3.2	7.8	-0.5	9.7	8.6	-1.6	9.5	3.8
Geography SG	7.7	8.0	5.9	8.3	4.9	11.9	12.9	6.7	20.3	10.1
History HG	10.8	20.5	17.7	13.6	9.2	11.8	22.9	12.7	16.8	15.1
History SG	7.4	14.5	17.0	13.0	8.9	12.7	20.3	11.6	9.7	11.6
Mathematics HG	12.0	15.0	12.9	12.7	16.8	16.7	29.6	11.6	14.2	18.0
Mathematics SG	9.4	12.9	7.3	12.1	16.8	12.1	20.9	10.9	11.1	14.4
Science HG	16.6	24.6	14.3	18.3	19.3	22.0	29.3	19.8	18.3	22.4
Science SG	15.1	20.7	10.3	13.8	17.5	17.9	21.7	16.3	13.6	16.1

Highlighted values: Gap = CASS – examination mark > 20 marks, or correlation <0.60

The pattern across quintiles of the school SES distribution (national quintiles) was repeated across almost all subjects: Assessment accuracy differed little by either criterion in the bottom three quintiles for all subjects, but it improved by either criterion (reliability and leniency) in the fourth and especially fifth quintiles (**Table 10A**). Quintile 5 contained more accurately assessing schools. **Table 10B in the Appendix** shows greater detail, but with the marks and gaps shown at the individual level, i.e. unlike in **Table 10**, where each school rather than each individual carries the same weight.

Table 10A: Mean school level correlations and gaps between CASS and examination marks by school poverty quintile, 2005 (Quintile 1 contains the poorest and Quintile 5 the richest schools)

	Mean school level correlation between CASS and examination marks					Mean Gap between CASS and examination marks				
	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Biology HG	0.64	0.61	0.64	0.66	0.74	25.5	25.8	24.5	21.5	14.8
Biology SG	0.61	0.59	0.57	0.58	0.61	21.7	21.8	21.7	19.2	14.6
English First Language	0.74	0.74	0.74	0.75	0.81	11.7	10.7	9.9	7.5	5.6
English Second Language	0.66	0.68	0.69	0.71	0.73	14.5	14.5	14.0	12.6	10.1
Geography HG	0.62	0.62	0.61	0.59	0.61	7.6	4.4	4.0	2.3	0.0
Geography SG	0.71	0.69	0.70	0.71	0.77	12.6	12.7	11.2	9.6	4.9
History HG	0.55	0.57	0.59	0.62	0.72	18.9	19.2	17.3	15.5	8.6
History SG	0.54	0.52	0.52	0.52	0.53	14.3	11.8	11.5	9.7	8.2
Mathematics HG	0.79	0.78	0.77	0.80	0.86	20.9	26.3	25.2	21.4	14.3
Mathematics SG	0.78	0.77	0.78	0.78	0.82	17.0	16.3	15.2	13.4	10.3
Science HG	0.67	0.67	0.68	0.71	0.82	26.8	26.8	25.1	22.3	18.4
Science SG	0.64	0.63	0.64	0.64	0.67	17.9	17.8	16.3	14.6	13.7

Highlighted values: Gap = CASS – examination mark > 20 marks, or correlation <0.60

Differences by school type were difficult to interpret given the small number of independent schools and ambiguity in what it meant to be an independent school: Both very rich and relatively poor schools were included in this group, explaining the larger standard deviation for independent schools compared to public schools. There were no substantial differences between these two categories, except that the mean gap was usually smaller in independent schools (i.e. CASS marks are less inflated), and for all the Higher Grade subjects (but not English), correlations were considerably better than in public schools.

Table 11: Mean school level correlations and gaps between CASS and examination marks by school type, 2005

	Mean school level correlation between CASS and examination marks		Mean Gap between CASS and examination marks	
	Independent schools	Public schools	Independent schools	Public schools
Biology HG	0.72	0.66	17	22
Biology SG	0.59	0.59	16	20
English First Language	0.74	0.78	18	7
English Second Language	0.69	0.68	12	14
Geography HG	0.56	0.61	2	4
Geography SG	0.73	0.72	9	10
History HG	0.68	0.62	9	15
History SG	0.45	0.53	8	11
Mathematics HG	0.87	0.82	15	19
Mathematics SG	0.77	0.79	12	15
Science HG	0.75	0.73	20	23
Science SG	0.64	0.64	16	16

Empirical testing for first-order stochastic dominance in terms of the reliability of CASS revealed the following patterns:

- There was no unequivocal trend over time.
- Western Cape reliability of assessment first-order dominated all eight other provinces. This implied that Western Cape schools were assessing more accurately than schools from the other provinces regardless of the threshold correlation value for assessment reliability chosen. There was no stochastic dominance between other provinces, thus no further general conclusion can be drawn – the answer depended on the correlation threshold selected.
- Quintile 5 first-order dominated over all the other quintiles over the entire range of correlations, whilst Quintile 4 appeared to dominate the others over most of the relevant range.

Taken across the mean of all subjects (an admittedly weak measure), of the 5 968 schools included in the data for 2005, 1 107, just shy of a fifth of all South African schools, were assessing less reliably, measured against a correlation of 0.6. The share of schools within each province that possessed a school correlation of less than 0.6 is displayed in **Table 12**. Over a quarter of all KwaZulu-Natal and Limpopo schools assessed matriculants unreliably, in comparison to the fewer than 10% in the Western Cape and the Free State. The Western Cape, Northern Cape and Free State combined contained only approximately 5% of schools assessing unreliably, whereas Mpumalanga and KwaZulu-Natal made up almost two-thirds of such schools in South Africa.

The measure used here is, however, a little problematic, as it does not distinguish between subjects and may thus be affected by factors such as differences in subject weights across schools. An alternative is to investigate provincial differentials in some individual subjects. English Second Language and Mathematics SG, the most assessed subjects, were investigated and the results presented in Table 13. Again, in terms of reliability of assessment in English Second Language, Limpopo and KwaZulu-Natal schools fared the worst. Although the Northern Cape had a reasonable share of schools faring poorly in terms of overall assessment reliability, a mere 4% were assessing unreliably in English Second Language. Once again, two provinces were responsible for a substantial share of poor assessment reliability. KwaZulu-Natal and Mpumalanga accounted for more than 50% of the unreliably assessing schools in English Second Language. In the case of Mathematics SG, the picture was similar, in that KwaZulu-Natal and Mpumalanga again accounted for more than half of the unreliably assessment schools (though unreliably assessment was less common in this subject), with the Eastern Cape also comprising a substantial share at 18%.

Returning to **Table 12**, more than a fifth of the schools in each of the bottom three quintiles under-performed in terms of student assessment reliability, much more than the top two quintiles. Of all schools assessing unreliably, more than 80% came from the bottom three quintiles. The clear divergences between assessment reliability of poor and rich quintiles was in line with patterns of educational outcomes in South Africa.

Table 12: Share of schools with correlation between all CASS and examination marks across all the selected subjects below a threshold of 0.6, by province and quintile, 2005

Province	Share of schools assessing inaccurately (%) within province	Provincial share in sample of schools assessing inaccurately
Western Cape	6.0	1.9
Northern Cape	15.1	1.4
Free State	7.4	2.0
Eastern Cape	12.6	10.0
KwaZulu-Natal	28.2	38.2
North-West	16.9	6.4
Mpumalanga	17.8	21.8
Gauteng	15.0	8.4
Limpopo	27.3	9.8
South Africa	18.5	100.0

Province	Share of schools assessing inaccurately (%) within province	Provincial share in sample of schools assessing inaccurately
Quintile 1 (poorest)	23.3	31.3
Quintile 2	24.7	25.3
Quintile 3	20.1	24.8
Quintile 4	12.9	11.1
Quintile 5 (richest)	8.9	7.5
South Africa	18.9	100.0

Note: Discrepancies in the share for South Africa result from the fact that some schools were not ordered by quintiles.

Table 13: Share of schools with correlation below threshold of 0.6 for English Second Language and Mathematics SG, by province (2005)

	English Second Language		Mathematics SG	
Province	**% of schools assessing poorly**	**Provincial share of poorly assessing schools**	**% of schools assessing poorly**	**Provincial share of poorly assessing schools**
Western Cape	9.0	2.2%	3.1	2.3%
Northern Cape	3.8	0.3%	5.1	0.7%
Free State	15.0	3.7%	3.6	1.8%
Eastern Cape	18.5	13.0%	10.3	17.8%
KwaZulu-Natal	30.6	34.3%	8.9	27.7%
North-West	23.3	7.8%	11.5	9.5%
Mpumalanga	19.7	22.6%	14.7	28.4%
Gauteng	17.6	6.3%	5.8	7.4%
Limpopo	30.3	9.7%	5.7	4.4%
Total		100.0%		100.0%

4 ASSESSMENT QUALITY AT THE SCHOOL LEVEL: CONSIDERING BOTH RELIABILITY AND LENIENCY

A fuller reflection of the quality of assessment should consider both the gap between mean CASS and examination marks and the correlation between these two. In **Table 14**, criteria for "very accurate" and "extremely inaccurate" assessment standards in a school were set as follows:

- If the difference between the mean assessment mark and mean examination marks in a school was less than 10 percentage points and the correlation between these two sets of marks is 0.6 or more, a school was taken to be assessing very accurately.
- If the difference between the mean assessment mark and mean examination marks in a school was greater than 20, and the correlation between these two sets of marks was 0.3 or less, a school was taken to be assessing extremely inaccurately.

By the criteria used here, there was a high frequency of schools assessing accurately. Good assessments by these criteria were especially found in History HG (65% good assessment) and English First Language (61%), followed by Geography (HG & SG) and History SG. Schools assessing very accurately were uncommon in Physical Science (both HG and SG) and Biology (especially SG).

In contrast, extremely inaccurate assessment was less common, judged by these rather lax criteria (examination marks more than 20 percentage points below assessment marks, and the correlation between the two marks less than 0.4). Biology (SG 19%, HG 16%), and Science HG (11%) stood out as the subjects where inaccurate assessment occurred relatively frequently.

Table 14: Proportion of schools assessing accurately (correlation>=0.6, mean gap <=10) and inaccurately (correlation<0.4, mean gap>20) using combined criteria

	Very accurate assessments	Extremely inaccurate assessments
Biology HG	20%	16%
Biology SG	15%	19%
English First Language HG	61%	0%
English Second Language HG	28%	3%
Geography HG	47%	1%
Geography SG	44%	2%
History HG	65%	0%
History SG	40%	0%
Mathematics HG	30%	2%
Mathematics SG	34%	2%
Physical Science HG	9%	11%
Physical Science SG	13%	6%

However, the assessment performance using a single criterion rather than these combined criteria appeared to be far less satisfactory. This related especially to the mean assessment marks, which was extraordinary high in some schools, even where assessment marks did correlate with examination marks. For example, for Biology SG, more than 200 schools out of 5 300 had assessment marks more than 40 percentage points above the examination mark. Yet even in these schools, more than a third had a correlation of 0.6 or higher between these two marks. Similar figures applied to other subjects. So it appeared that the tendency to give high assessment marks was not necessarily always closely related to poor correlation with the examination mark. It appeared as if teachers could rank student performance relatively well, but gave leniently high marks. The question arises whether this may have been related to a culture of setting marks too high in earlier grades in order to reduce failures or deflect parent protests.

Table 15: Share of schools with correlation below threshold of 0.6 by subject and grade, 2005

		% of schools assessing subject unreliably	Number of schools assessing subject unreliably	Subject share of schools assessing unreliably
Biology	**HG**	0.30	2 362	8%
	SG	0.45	4 358	14%
English First Language	**HG**	0.13	961	3%
English Second Language	**SG**	0.22	5 330	18%
Geography	**HG**	0.18	2 375	8%
	SG	0.40	2 517	8%
History	**HG**	0.35	843	3%
	SG	0.60	1 794	6%
Mathematics	**HG**	0.06	820	3%
	SG	0.09	4 774	16%
Physical Science	**HG**	0.17	1 339	4%
	SG	0.34	2 969	10%
Total for these subjects			30 442	100%

Note: "Unreliable" assessment is here taken to be a CASS mark correlated below 0.6 with the examination mark in that subject, for subjects entered by at least 15 candidates in a school.

Part 6:

Conclusion

This paper started from the premise that assessment provides important signals to students that should assist them to prepare for examinations and make informed choices about career options, further studies and subject choice, both at school and beyond. Analysis indeed showed that schools where there was less reliable continuous assessment tended to perform worse in examinations.

Analysis of CASS took place against the context of examination marks, taken to be the more accurate reflection of student performance. Two measures of accuracy of continuous assessment were used and applied across a number of subjects: the leniency with which CASS marks were awarded (compared to examination marks), measured as the gap between these marks, and the reliability of CASS marks in terms of their correlation with examination marks.

The analysis was conducted at two levels. First, all individual marks in different subjects were analysed using the above two measures of assessment accuracy, and patterns were investigated. Secondly, a similar analysis was conducted at school level rather than of individual marks, to determine in which schools assessment was less accurate, and to attempt to identify patterns in this regard.

The broad conclusions of this study are that continuous assessment accuracy was weakest in terms of the great leniency of assessment in many schools (inflated CASS marks), although unreliability of assessment also was a cause for concern in some cases. This required targeted interventions. There was also evidence of a clear hierarchy in terms of assessment accuracy. The bulk of inaccurately assessing schools by both measures combined were in Mpumalanga and Kwazulu-Natal, with the Eastern Cape also a large contributor. The Western Cape, and schools in the top and even the second quintiles of the SES distribution, assessed much better. In terms of subjects, Mathematics, both at Higher and Standard Grade levels, was the best assessed subject, with English First and Second Language falling close behind. This contrasted to History, where there was a larger share of poorly assessing schools than for any other subject.

Added to this was the disturbing finding that these gaps had in most cases been increasing, in some cases substantially so. KwaZulu-Natal and Mpumalanga witnessed large increases in their aggregate gaps, whilst the small gap in History had increased four-fold. The gaps were widening largely as a result of falling examination marks, in part the results of tightening up of the national examinations, but perhaps also because of a larger number of under-prepared students entering the examinations.

Apart from the fact that inaccurate continuous assessment in matriculation and, presumably, also assessment in earlier grades, were sending wrong signals to students and parents, resulting in inappropriate subject choices, career planning and preparation for the matriculation examination, there was a further issue that Umalusi and the education authorities should take note of. With the 25% weighting given to CASS marks in matriculation, and the limit of a mean deviation of 10% either way between examination and CASS marks that Umalusi imposed, differences in strategic behaviour between teachers or schools could have important consequences. Schools setting high standards in CASS in order to induce more intensive learning in preparation for the examination may place their candidates at a considerable disadvantage in the final matriculation mark (of up to 5 percentage points) relative to schools who persist with exceedingly lenient assessment.

There is a wider consideration also, though. It is extremely worrying that differentials between CASS and examination marks appeared not to result in feedback to the following year's CASS marks. Teachers did not appear to be seriously re-evaluating their own assessment standards on the basis of the examination marks, thus the already weak link between CASS marks and curriculum standards remained weak. This cries out for further investigation and action by the education authorities.

Bibliography

Cohen, J. (1988). Statistical power analysis for the Behavioral Sciences (2nd ed.) Hillsdale, NJ: Lawrence Erlbaum Associates.

Davidson, R., & MacKinnon J.G. (2004). Econometric Theory and Methods. Oxford: Oxford University Press.

Burger, Don. (no date.) Using Standards-Led Policy to Align Assessment and Accountability Systems. Pacific Resources for Education and Learning. . [WWW document].

URL http://www.prel.org/products/re_/standards-led.htm. 12 Feb. 2008.

Madden, D., & Smith, F. (2000). Poverty in Ireland, 1987–1994: A Stochastic Dominance Approach. The Economic and Social Review. 31,(3), 187–214.

Moskal, B.M., & Leydens, J.A. (2000). Scoring rubric development: validity and reliability. Practical Assessment, Research & Evaluation 7(10). [WWW document].

URL http://PAREonline.net/getvn.asp?v=7&n=10. 11 February 2008.

Moskal, B.M., Leydens, J.A., & Pavelich, M.J. (2002). Validity, reliability and the assessment of engineering education. Journal of Engineering Education, Jul 2002. [WWW document].

URL http://findarticles.com/p/articles/mi_qa3886/is_200207/ai_n9121811/pg_1. 12 February 2008.

Nicol, D., & Macfarlane-Dick, D. (2006) 'Formative assessment and self-regulated learning: a model and seven principles of good feedback practice', Studies in Higher Education 31(2).

North Central Regional Technology in Education Consortium (2002). Glossary of Education Terms and Acronyms. [WWW document].

URL http://www.ncrel.org/sdrs/areas/misc/glossary.htm. 12 February 2008.

Pearson Education. (no date). Assessment: Reliability/Validity. [WWW document].

URL http://www.ecs.org/html/issue.asp?issueid=12&subIssueID=123. 12 February 2008.

Umalusi data set (2003-2005)

Appendix

Table 4: Mean CASS marks, examination marks and gaps for all candidates, selected subjects (standard deviation in parentheses)

Table 10B: Individual CASS marks, examination marks and gaps, and correlations by school between CASS and examination marks in the poorest and richest quintiles of schools

Table 4: Mean CASS marks, examination marks and gaps for all candidates, selected subjects (standard deviation in parentheses)

	Western Cape		Northern Cape		Free State		Eastern Cape		KwaZulu-Natal		North West		Mpumalanga		Gauteng		Limpopo		South Africa (North West included)	
	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.
	Biology HG																			
CASS	60.4	(15.0)	65.7	(14.6)	53.3	(15.0)	47.2	(15.9)	56.9	(14.7)	47.9	(15.2)	57.1	(13.8)	60.7	(14.2)	52.5	(15.2)	55.9	(15.2)
Exam	52.7	(16.6)	46.4	(15.4)	43.3	(17.3)	36.7	(17.2)	36.8	(17.3)	29.1	(15.9)	24.0	(13.0)	44.7	(17.7)	36.8	(16.8)	35.1	(18.3)
Gap	7.7	(10.3)	19.3	(11.0)	10.1	(9.1)	10.5	(9.8)	20.2	(13.5)	18.8	(12.2)	33.1	(14.4)	15.9	(11.3)	15.7	(11.7)	20.8	(15.3)
	Biology SG																			
CASS	42.9	(12.5)	54.6	(13.3)	39.9	(11.3)	39.3	(11.4)	48.6	(12.9)	43.3	(13.4)	52.2	(13.7)	48.0	(12.1)	41.2	(12.1)	45.6	(13.4)
Exam	31.5	(13.7)	28.9	(12.8)	28.3	(13.3)	26.0	(12.2)	25.5	(12.6)	23.1	(12.7)	22.0	(11.6)	27.6	(13.6)	22.8	(11.8)	25.6	(12.8)
Gap	11.5	(11.1)	25.6	(13.5)	11.6	(10.4)	13.3	(10.6)	23.0	(14.5)	20.2	(13.8)	30.2	(15.0)	20.4	(12.6)	18.4	(11.9)	20.0	(14.3)
	English First Language HG																			
CASS	57.2	(12.9)	53.6	(12.9)	55.8	(11.5)	56.2	(12.8)	59.8	(12.3)	56.6	(13.4)	56.5	(11.9)	59.5	(12.1)	59.0	(13.0)	58.4	(12.5)
Exam	49.8	(13.7)	50.8	(12.4)	45.5	(12.6)	43.7	(13.1)	57.2	(13.1)	43.3	(12.6)	39.8	(11.4)	54.3	(13.3)	48.1	(13.9)	52.0	(14.1)
Gap	7.4	(8.6)	2.8	(8.4)	10.3	(8.6)	12.4	(7.8)	2.6	(7.0)	13.3	(8.1)	16.8	(12.0)	5.2	(8.5)	10.9	(8.7)	6.4	(8.9)
	English Second Language HG																			
CASS	51.3	(13.1)	50.9	(11.7)	46.7	(12.7)	43.8	(12.2)	50.2	(12.9)	48.7	(12.8)	49.3	(12.2)	52.8	(13.1)	48.1	(12.6)	48.9	(12.9)
Exam	43.1	(13.6)	43.0	(12.9)	37.9	(13.5)	32.6	(11.4)	33.8	(11.6)	34.9	(12.3)	34.4	(10.7)	44.7	(14.8)	36.3	(12.4)	36.2	(12.8)
Gap	8.1	(8.5)	7.8	(8.5)	8.7	(8.8)	11.2	(9.4)	16.4	(10.5)	13.8	(9.7)	14.9	(10.0)	8.1	(9.2)	11.8	(9.4)	12.7	(10.2)
	Geography HG																			
CASS	60.7	(13.9)	52.9	(15.4)	47.4	(14.2)	45.1	(15.1)	49.3	(14.8)	41.7	(12.2)	40.1	(10.7)	51.0	(15.0)	44.4	(12.2)	45.0	(14.0)
Exam	54.0	(15.9)	45.4	(16.0)	40.0	(15.3)	36.4	(16.2)	45.0	(19.9)	29.6	(14.5)	27.0	(11.3)	44.0	(18.3)	24.0	(11.8)	34.7	(17.7)
Gap	6.6	(10.9)	7.5	(10.4)	7.3	(10.0)	8.7	(10.7)	4.3	(14.2)	12.1	(10.0)	13.1	(9.2)	7.0	(12.2)	20.3	(8.9)	10.3	(11.8)
	Geography SG																			
CASS	46.2	(12.7)	41.6	(13.0)	38.3	(10.4)	37.8	(11.4)	44.1	(11.9)	38.6	(10.7)	37.8	(10.4)	38.9	(12.2)	42.9	(9.7)	41.1	(11.9)
Exam	44.8	(13.2)	32.7	(12.6)	34.8	(11.0)	29.4	(10.9)	44.6	(14.9)	28.3	(11.9)	28.7	(9.7)	40.1	(14.6)	32.9	(10.0)	36.9	(14.4)
Gap	1.4	(11.7)	8.9	(12.1)	3.6	(9.7)	8.4	(10.6)	-0.5	(14.2)	10.4	(10.3)	9.1	(9.7)	-1.1	(12.2)	10.0	(9.8)	4.2	(12.7)

(Continued on next page)

Table 4 (Continued): Mean CASS marks, examination marks and gaps for all candidates, selected subjects (standard deviation in parentheses)

	Western Cape		Northern Cape		Free State		Eastern Cape		KwaZulu-Natal		North West		Mpumalanga		Gauteng		Limpopo		South Africa (North West included)	
	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.	Mean	Std Dev.
	History HG																			
CASS	62.0	(15.7)	56.5	(12.7)	55.4	(14.4)	51.1	(17.2)	50.2	(15.9)	41.8	(15.2)	49.1	(12.9)	57.3	(14.5)	44.3	(16.0)	51.4	(15.6)
Exam	50.7	(16.3)	34.4	(12.4)	38.2	(13.1)	36.4	(18.4)	40.9	(18.3)	29.0	(17.5)	25.5	(12.8)	44.4	(18.0)	28.5	(16.4)	35.9	(18.3)
Gap	11.3	(11.2)	22.0	(10.3)	17.2	(14.4)	14.7	(12.5)	9.3	(13.9)	12.7	(13.3)	23.6	(13.3)	12.9	(13.6)	15.8	(12.5)	15.5	(14.5)
	History SG																			
CASS	47.1	(13.2)	47.4	(12.9)	45.9	(11.0)	40.5	(12.2)	43.8	(12.8)	36.2	(12.6)	44.2	(12.7)	43.9	(11.9)	37.2	(11.6)	42.8	(12.8)
Exam	39.3	(13.9)	33.6	(12.5)	27.9	(12.2)	26.8	(13.0)	34.5	(14.6)	22.8	(13.0)	23.0	(12.5)	32.0	(14.2)	27.7	(11.8)	30.9	(14.4)
Gap	7.8	(12.7)	13.8	(12.3)	18.0	(13.5)	13.7	(13.6)	9.2	(14.3)	13.4	(13.5)	21.2	(13.9)	11.9	(13.8)	9.5	(12.3)	11.9	(14.0)
	Mathematics HG																			
CASS	64.1	(17.0)	61.0	(14.5)	61.7	(17.4)	56.2	(17.9)	53.7	(17.9)	50.9	(18.0)	47.4	(16.0)	57.2	(16.6)	57.1	(17.0)	54.5	(17.8)
Exam	52.4	(20.6)	45.1	(18.4)	47.8	(19.2)	40.7	(22.9)	36.4	(23.8)	34.1	(22.4)	16.7	(17.9)	45.2	(21.3)	41.5	(20.3)	35.9	(24.3)
Gap	11.8	(10.5)	15.8	(11.6)	13.9	(10.4)	15.4	(12.8)	17.3	(12.9)	16.8	(12.9)	30.7	(13.3)	12.0	(12.1)	15.6	(11.1)	18.6	(14.5)
	Mathematics SG																			
CASS	44.8	(18.2)	45.0	(16.4)	41.4	(17.5)	37.7	(15.0)	40.1	(15.4)	36.8	(15.1)	42.7	(14.9)	41.6	(16.2)	35.2	(15.3)	40.1	(15.9)
Exam	36.0	(21.3)	31.6	(19.8)	33.9	(20.7)	24.9	(17.8)	23.2	(18.4)	23.7	(19.4)	21.7	(18.6)	29.9	(21.2)	23.7	(19.0)	25.8	(19.7)
Gap	8.8	(11.2)	13.5	(11.9)	7.5	(11.3)	12.9	(12.3)	16.9	(13.5)	13.1	(13.3)	21.0	(14.5)	11.7	(13.2)	11.5	(11.5)	14.4	(13.5)
	Physical Science SG																			
CASS	48.0	(12.4)	52.5	(12.1)	40.2	(11.3)	41.1	(11.2)	44.2	(12.5)	44.8	(12.2)	47.0	(11.8)	45.2	(11.1)	40.2	(11.3)	43.8	(12.0)
Exam	32.5	(13.2)	32.1	(12.8)	30.1	(11.8)	26.8	(10.9)	26.6	(11.1)	26.0	(11.2)	25.5	(10.3)	28.7	(11.8)	27.1	(10.8)	27.6	(11.5)
Gap	15.5	(9.6)	20.4	(10.7)	10.1	(8.3)	14.3	(9.9)	17.6	(12.4)	18.8	(11.6)	21.4	(12.3)	16.5	(10.7)	13.1	(10.0)	16.2	(11.3)
	Physical Science HG																			
CASS	63.8	(14.3)	61.9	(13.1)	50.0	(16.0)	54.2	(16.1)	49.0	(15.3)	48.6	(15.0)	50.8	(11.6)	58.1	(13.9)	51.4	(15.3)	52.8	(14.7)
Exam	47.2	(18.4)	39.4	(17.0)	35.9	(17.0)	35.5	(18.0)	30.2	(17.5)	27.2	(15.3)	21.4	(11.2)	38.3	(17.9)	33.6	(16.3)	30.8	(17.6)
Gap	16.6	(10.1)	22.5	(11.6)	14.1	(7.8)	18.7	(10.3)	18.9	(13.0)	21.5	(11.3)	29.4	(10.5)	19.8	(11.2)	17.9	(9.8)	22.0	(12.2)

Note: Mean gaps exceeding 20 percentage points are highlighted

Table 10B: Individual CASS marks, examination marks and gaps, and correlations by school between CASS and examination marks in the poorest and richest quintiles of schools

	Mean CASS marks					Mean Examination marks					Mean Gaps					Correlation: CASS vs. Exam	
	Quintile					Quintile					Quintile					Quintile	
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	5
Biology HG	53.8	52.0	52.5	54.5	61.0	29.2	26.4	28.0	32.6	47.0	24.6	25.5	24.5	21.9	14.0	0.500	0.697
Biology SG	45.1	45.2	45.1	45.9	45.8	23.7	23.7	23.8	26.4	30.9	21.4	21.5	21.3	19.5	14.9	0.398	0.508
English First Language	58.4	52.8	54.4	56.1	59.4	47.1	42.2	45.3	49.9	54.8	11.3	10.6	9.1	6.2	4.6	0.816	0.800
English Second Language	46.7	47.2	48.1	49.4	51.5	33.0	33.2	34.8	36.8	41.2	13.7	14.0	13.3	12.6	10.3	0.606	0.767
Geography HG	42.1	41.7	41.7	43.1	52.4	29.4	29.1	30.4	32.8	46.8	12.7	12.6	11.4	10.3	5.6	0.663	0.799
Geography SG	40.6	39.4	40.1	41.9	44.1	32.9	35.0	35.7	39.3	43.2	7.8	4.4	4.5	2.6	0.8	0.502	0.611
History HG	50.3	48.3	46.1	49.1	58.2	31.1	30.0	29.5	32.5	48.7	19.2	18.3	16.7	16.7	9.4	0.569	0.738
History SG	42.6	41.6	41.5	43.3	46.0	28.2	29.0	29.8	33.0	37.2	14.5	12.6	11.7	10.3	8.8	0.420	0.559
Mathematics HG	50.5	47.2	47.0	49.8	60.1	28.5	21.2	22.7	28.9	45.6	22.0	26.0	24.3	20.9	14.4	0.781	0.831
Mathematics SG	39.4	38.2	37.0	38.7	45.2	22.3	21.7	21.8	24.9	34.6	17.0	16.5	15.1	13.8	10.6	0.693	0.795
Physical Science HG	50.6	48.7	47.5	48.7	57.4	25.1	22.2	23.3	26.9	39.3	25.5	26.5	24.2	21.7	18.1	0.645	0.803
Physical Science SG	43.8	43.6	42.6	42.5	45.6	25.9	25.5	26.1	27.9	31.7	18.0	18.2	16.5	14.6	13.9	0.501	0.642

Note: Quintile information was only available for 5 044 out of 6 174 examination centres (schools).

37 General Van Ryneveld Street, Persequor Technopark, Pretoria
Telephone: +27 12 349 1510 • Fax: +27 12 349 1511
E-mail: Info@umalusi.org.za • Web: www.umalusi.org.za

UMALUSI
Council for Quality Assurance in
General and Further Education and Training