FINAL

THE DILEMMA OF EQUATING EXAMINATIONS AND ASSESSMENT STANDARDS FOR THE NATIONAL SENIOR CERTIFICATE IN SOUTH AFRICA

Prof. Coert Loock, University of Johannesburg, coertl@uj.ac.za

ABSTRACT

Learner assessment and standard setting, has always been an issue of discussion, not only among professionals but also in public. Broadfoot (2002:5) indicates that the search for an unambiguous and dependable way of measuring "ability" is indeed one of the enduring themes of assessment research in the 20th century. Further to this, the tension between the scientific aspirations of assessment technologies to represent an objective reality and the unavoidable subjectivities injected by the human focus of these technologies is very much in evidence in most countries (Davies, 2002; 185 – 204). Indeed, as Moss & Schutz (2001: 37 – 70), argue, *it is the process of generating standards*, and in particular, the possibility of extremes that do not conform, that is an essential dynamic of education quality and innovation.

The generation and setting of *standards* becomes even more of an issue in countries where examinations and assessment mediate university entrance. Access to institutions that are both catering to a mass market, and are very important to people's life changes, is almost bound to be a vexed and a political issue (Bakker & Wolf, 2001; 285). Democracies demand fairness and objectivity, and equal opportunity for access. This is especially true in the *South African context*, where the emergence of a new democratic state in 1994 ushered changes in various social arenas including education, where "the dialectical dying out of the old and the birth of the new" (Lubisi & Murphy, 2002:255), has to account for the *diversity in standards* that still exist. The consequent introduction of a new exit level examination (NSC) the FET-phase complicates matters even further and questions are raised about standards and standard-setting...

BACKGROUND

The Senior Certificate (Report 550) for full-time candidates has been written for the last time in 2007 and is replaced by the National Senior Certificate in 2008. the National N3 (Report 190 and 191) and the National Senior Certificate offered by the Further Education and Training (FET) colleges will be phased out in 2009 and will be replaced by the National Certificate (Vocational), levels 2,3 and 4. (Umalusi, 2008). Although the establishment of a coherent national framework for educational qualifications, as indicated by the South African Qualifications Authority (SAQA), (2001:9), is well advanced and has been widely supported in principle, the implementation of this change remains a challenge, especially in the more political environment of secondary education and university entrance requirements.

Since a single certificate is issued to all learners in South Africa, it is also necessary to ensure that *standards between accredited assessment bodies* are comparable. The dilemma is that outcomes based assessment (OBA) should link up with a set of critical outcomes (Department of Education, 1998a: 7), in a 'continuous way' which has to be *externally moderated* at the two exit levels (GETC) and FETC). This implies that the contemporary view that standards for learner performance need to be established by providing benchmarks such as pre-established norms, should be reconsidered or at least enhanced, by a process of generating, discussing and using *"dynamic standards"* for moderation purposes.

The principle question really is, whether the current "equating" procedures (the so-called, "standardization" of examination marks), inherent in the previous Senior Certificate (Matric) Examination, can be applied to the new outcomes based NCS with its envisaged progressive moderation processes and how can the assessment bodies and Umalusi ensure that the assessment results are credible, objective and politically legitimate? The fact that marks are adjusted during the resulting process is not readily accepted by the public, in spite of the levels of sophistication developed over the years.

All *equating procedures* have major drawbacks and limitations; while it is difficult to make sense of the idea of "constant standards". Nonetheless as Bakker & Wolf (2001:289), indicates: "upper-secondary tests end exams remain by far the most legitimate and acceptable entry mechanism available for a modern university system. In South Africa a number of experts have expressed skepticism about the standardization process, whereby the results of an examination are compared with established norms and standards and adjusted if necessary. The process is premised on the probability that candidates with equal ability will obtain equivalent results if writing question papers under different circumstances (Loock & Grobler, 2005). Other countries (and well established democracies) are continuously trying to grapple with the same issues. Bellar (2001:315); Alberts, 2001:353) refers to the Israeli and Dutch concerns over whether examination standards are consistent and equivalent over subjects and over time.

The key argument to be presented is whether, the current "equating practices" are s the best way to ensure fairness, objectivity and equal opportunity to all learners in the system, after implementation of the new Outcomes Based Education System as defined by the National Curriculum Statement, and if not so, what are the alternatives?"

STANDARDIZED ASSESSMENT AND LARGE SCALE PUBLIC EXAMINATIONS

2.1 INTRODUCTION

The earliest evidence of standardized testing based on merit comes from China during the Han dynasty The concept of state ruled by men of ability and virtue was a permutation of the Confucian philosophy. The imperial examinations at the time covered the Six Arts, including music, archery and horsemanship, arithmetic, writing and knowledge of the rituals and ceremonies. Later the five studies (military strategies, civil law, revenue and taxation, agriculture and geography) were added to the testing.

It is important to note that **standardized testing** is not traditionally part of European pedagogy, which was influenced by the skeptical and open-ended tradition of debate inherited from Ancient Greece, and favored the essay. The use of standardized testing in the United States is a 20th Century phenomenon that has been driven in part by the ease of computer-grading of standardized tests, and the comparative difficulty of grading essays by computer. In the United States, the need for the Federal Government to make meaningful comparisons across a highly decentralized public education system has also contributed to the debate about standardized testing.

In the Netherlands the procedures for the construction of exit level examinations and for setting cut-off scores remained unchanged until the early nineties. During the 1990's, equivalence became a political issue. In parliament there was some doubt as to whether the increasing numbers of pupils opting for higher education have e not been accommodated by declining standards. A similar fear is often expressed in South Africa. While the content of education is changing, the expectation is that the general "outcome level", whatever that should be, should not change (Alberts, 2001). Obviously this calls for comparison and equating procedures even more than before, which brings us to the issue of equating.

2.2 EQUATING HIGH STAKES EXAMINATIONS AND TESTING

The current focus in South Africa, and many other countries is on high stakes testing, in which measurement of results will have significant impact on learners, as promotion, graduation and the qualification for scholarships and bursaries are now tied to performance in these tests (Agrey:2004).

The very term 'high stakes" embodies the hopes and fears these tests inspire only if the stakes are high, say their advocates on one hand – only if there is something valuable to be gained or to be lost – will teachers and students take the tests seriously and work hard to do their best, thus serving both their own interests and the public interest in higher achievement (Heubert & Hauser, 1999)

High stakes examinations and tests thus requires a clear standardized curriculum as well as the introduction of systems, processes and procedures that will eradicate inequities in students' opportunity to learn and perform accordingly. Skeptics of high stakes public testing and examinations, such as Black (1991) and Kohn, 2000a and 2000b), posits that public examinations are often seen as assurances of fairness and reliability that is quite unjustified and that the public demands for external assessments arise from three main considerations, namely: public distrust in the education system, the perceived inferiority of school-based assessment over external written examinations and testing and the desire to compare between schools. Kohn, even contends that high stakes examinations and testing marks a major retreat from fairness, accuracy and quality. According to him, these tests are often biased, either politically or socio-economically, since they require a set of skills more likely to be possessed by children from a specific political orientation or from a privileged society, thus a greater disparity is created and fairness and equity become a hollow promise.

The obsession with high stakes standardized testing has not arisen in a vacuum. The ideological basis can be found in the neo liberal value systems expressed in the current globalization phenomenon, a planetary unified global trading network operating according to a common set of rules (Smith, 2000, Agrey, 2004), where the market is the only factor to be considered in structuring our lives and our institutions. Based on the philosophy that 'better" education can be measured, standardization, easily quantifiable results and the willingness to reshape all intervening processes seem to characterize the path to success in both education and business (Ollman, 2003).

The history of equating Grade 12 examination marks in South Africa

Assessment in South Africa has been dominated by the Senior Certificate (Matric) examination, which doubles as a school leaving certificate and a university entrance qualification (Lubisi & Murphy, 2002:260). Certification for university entrance was for many years closely administered by the rules and regulations established by the Joint

Matriculation Board (JMB) who by 1921, assumed the role of quality controller for matric examinations.

The JMB exercised quality control, firstly by scrutinizing all examination papers set by state departments (four 'national" departments controlled by the Department of National Education) and checking of marked scripts by subject moderators; secondly by scrutinizing the statistics of matric examination results; thirdly by annually reviewing the right of any department to run the matric examination and lastly, to exercise control over school syllabi (Trümpelmann, 1991).

From 1921 up to 1953, the JMB granted permission to various provincial departments of education to run school leaving examinations and thus become examination bodies, however, the JMB was to ensure that these exams were of a *comparable standard* to that set by the JMB itself. As far back as 1933, a standing committee of the JMB would supervise examination statistics because of the considerable variation in the failure rate of various examinations in the course of time, "the only conclusion one can come to is that the variation must be in the standard of the examination" (Trümpelmann, 1991:106)

The argument was that an obligatory standard distribution curve ought to be applied per subject in order to adjust the marks to a standard score before the comparative process could be applied.

In order to improve the application of the standard distribution curve, a further sub-committee for standard distribution was appointed in 1975. They immediately attempted to structure the issue concerning standard distributions and formulate an equation and demanded rectifications from time to time; however, according to Gledhill (In Trümpelmann, 1991:192), concern was voiced about the propriety of the *adjustment of examination marks* on the basis of standards and norms. Such views were based on the feeling that examinations itself should be the measure of success – the examination paper, set by competent examiners and moderated by experienced moderators should be the final criterion. Fact is, according to Gledhill, that practice proved different, as it has

been repeatedly demonstrated, by having several examiners mark the same script independently, that the marks given by different examiners to the same answer to the same question may differ widely and a variation of 10% is common, even in the so-called objective subjects. It was true then and it is still true today.

1992, was marked by the demise of the JMB. Matriculation was to be controlled by the Matriculation Board, which was to be a sub-committee to the Committee for University Principals (CUP), now called the South African University Vice-Chancellors' Association (SAUVCA). A new statutory body, The South African Certification Council (SAFCERT), was established in 1986 under the South African Certification Council Act (Act 85 of 1986), with school leaving certification as its primary objective. Until 2002, SAFCERT would be responsible for the moderation and quality control of all school leaving examinations.

SAFCERT was transformed by the end of 2002, to cater for a new education and training system and will henceforth be known as The General and Further Education and Training Quality Assurance Council (Umalusi), in accordance with the General and Further Education and Training Act, 58 of 2001). Umalusi took over the responsibilities of SAFCERT until the current Senior Certificate (Matric) exam will be substituted by a new structure, ushered by the government's national curriculum reform project, Curriculum 2005, which will be characterized by the division between General Education and Training (GET) and Further Education and Training (FET). Umalusi continued to build on both SAFCERT and the JMB's approaches on controlling the standard of the SCE. However, it was becoming increasingly evident that the context in which the examination was being written was continuously changing (Lolwana, in Reddy, 2006).

While most of the examination bodies in countries, other than South Africa, utilize statistical data to standardize results; few of them *apply pre-determined statistical norms* or desired distributions. However, not only is the standardization of examination raw

marks recognized as an educationally sound practice, it proved to be a cost effective, reliable and appropriate process for the South African scenario (SAFCERT, 2002:18).

It is clear that the concept of high stakes standardized testing and the issue of equating examinations and scores, using one common scale is not a simplistic one and needs retrospective research.
Equated exams enable us to compare the performance of learners over a number of years and are thus important in situations where exams are used to control the performance level of certified learners and their consequent progression.

The equivalence of marks obtained in the matriculation examinations in South Africa prior to 2008 changes in the curriculum always require careful consideration. We cannot simply assume that the traditional reference norm or reference instrument can still be applied, or that established equating procedures, such as those currently adopted in South Africa, will provide meaningful information (Agrey, 2004).

According to Greatorex & Suto (2006), a major theme of research currently is the nature and use of human judgment in the marking of school examinations. Grade 12 examination marking is a diverse activity, encompassing a wide range of subjects with a variety of question styles, rubrics and marking schemes.

Psychologists have constructed multiple models of judgment and decision-making, which have yet to be applied to examination marking, and one potentially useful theoretical approach is that of dual processing. Such models distinguish two qualitatively different but concurrently active systems of cognitive operations:

System 1 thought processes which are quick and associative, and
System 2 thought processes which are slow and rule governed (Kahneman & Frederick, 2002).
The intuitive judgments of system 1 can be described as automatic, effortless, skilled actions which occur in parallel and so rapidly that they can be difficult to elucidate.

System 2 judgments in contrast have been termed ‘reflective”. And the thought processes they comprise are characterized as slow, serial. Controlled, and effortful rule-applications of which the thinker is self-aware (ibid.002).

There may be question types or stages of marking , that involve system 1 processing; at times; simple and repetitive matching of a candidate’s single word response with the model answer given in the mark scheme may be all that is required. At other times examiners become more familiar with a particular examination paper and mark scheme, or more experienced in marking in general, some sophisticated processes may be transferred from system 2 to system 1, while others remain exclusive to 2.

An empirical exploration of human judgment by the Research division of Cambridge Assessment

To explore this further an investigation was conducted into two contrasting examinations In line with the creation of nine (as opposed to four) provinces as part of the democratic dispensation of 1994, each of the provincial departments of education is currently responsible for administering their own examinations, including the setting, marking and moderation of examination papers. The nine provincial examination boards jointly coordinate their arrangements for the Senior Certificate (Matric) Examination through a national statutory body, the Inter-Provincial Examinations and Assessment Committee (IPEC).

The Senior Certificate can be regarded as a "group certificate", which requires candidates to pass minimum levels of a prescribed combination of subjects. These subject packages not only entailed differences in terms of some of the subject content, they also entailed differences in the level of difficulty. For the purposes of the Grade 12, Senior Certificate, candidates could offer subjects at "Lower", Standard and Higher Grades (National Education Policy Initiative, 1992, p. 16).

All these subjects are internally (Continuous Assessment), as well as externally (Written) assessed. For the purposes of this paper, we will concern ourselves with the *external (written) part of the assessment which take place during November/December.*

The scale at which the examination is conducted cannot be underestimated. The 2003 Examination , for example, was taken by 662342, full-and part-time learners in 5558 examination centres and marked by 38,512 markers across the country (Department of Education, 2003) The problem that faced SAFCERT at the time, and currently inherited by Umalusi, is to ensure that candidates with equal ability, who write different examination question papers under different circumstances, will obtain equivalent results

in order to comply with the requirements for the issuing of a single certificate, as indicated above.

SAFCERT concluded that, *"statistical moderation is necessary to take care of the variation in the standards of marking that may occur from year to year, from one subject to another, or from one examining body to another"* (SAFCERT, 2000/01: 10). The conclusion, since 1933 was, that statistical data could be utilized much more effectively to bring greater reliability to the Senior Certificate (Matric) Examinations. It will be the case until and when the examination in its present format will be replaced by a total new way of assessment as proposed by the FETC Policy Document (South African Qualifications Authority (SAQA) April, 2001).

Statistical moderation of examination and assessment results

In the report to the Minister of Education, The South African Certification Council (SAFCERT 2002: 15), reiterated that there is an abundance of evidence, both in South Africa and elsewhere in the world, that "despite careful attention and diligence of competent and experienced examiners, moderators and markers, it is impossible to determine whether a question paper is actually of the required standard until it has been written and marked. Therefore it is essential to review the raw examination marks. These raw results should be adjusted if evidence indicates that the question paper did not produce a fair result".

Example of the "standardization processing South Africa, standardization is achieved by changing the actual marks awarded to a learner, in other systems raw marks are never changed and, instead, standardization is achieved by alterations to the Grade-boundaries. In short, the standardization of examination results is the process whereby the results of the examination is compared with established norms and adjusted if necessary.

The “standardization process” consists of a set of computer programs by way of which examination RAW MARKS are compared with the norms currently supplied by the General and Further Education and Training Quality Assurance Council (Umalusi) (Loock, 2002:3). If the examination papers of the examinations in one year are the same standard as those of the previous year, the results should theoretically compare closely with the norms, as calculated by the Quality Assurance body (Umalusi). If this is not the case, adjustments will be suggested by way of the standardization process. The suggested adjustments are analyzed and evaluated by the National and Inter-provincial Standardization Committees

These committees have to decide whether adjustments suggested by way of the standardization process are acceptable, and if not, what other adjustments, if any, should be affected (to a maximum of 10%). The committee will take decisions on various reports, namely:

- Examiners/moderators reports;
- Pairs analysis;
- The difference between the raw mark averages per subject and the recommended adjustments, and
- The relative frequencies of the desired distribution as plotted on a smooth graph.

On the strength of all these factors the following decisions can be taken, namely:

- To accept the recommended adjustment;
- Take the raw mark as is;
- Do a block adjustment;
- Adjust the recommended adjustment by a certain percentage, or adjust between symbol intervals

During the standardization process a graph and statistical data are produced separately for each subject – the graph and statistical data reflect the outcome of the examination accurately in comparison to the norm.