

# “A tail of many dogs”

---

## Assessment and the scramble for standards

Presented at the 4<sup>th</sup> Sub-Regional Conference on a Assessment in Education, hosted by Umalusi from the 26<sup>th</sup> to the 30<sup>th</sup> June 2006

*Sandile Ndaba, Umalusi*

### **Abstract**

Educational assessment increasingly focuses on uniform educational standards and accountability for students and teachers in order to meet common educational standards. Examinations are used for a variety of purposes: to help learners, parents and prospective employers to make educational and career choices, to provide a basis for selecting individuals for specific sorts of further or higher education, or for certain jobs, and for qualifying individuals to undertake specific vocational or professional activities as well as to exclude those that do not qualify. It is for this reason that those involved in examinations often refer to examinations, pejoratively, as ‘the tail that wags the dog’. However, this ‘tail’ that is examinations has many dogs (the various contending purposes to which examinations are put) and it wags them in accordance with the various agendas of educational technocrats.

This paper discusses the issue of standards and the use of examinations in the ‘scramble for standards’. It begins by discussing the term ‘standards’ in general and then proceeds to consider standards in the South African education system. It argues that examinations, while they serve extremely important purposes of licensure, certification and accountability, cannot, however, ensure quality education. The author argues that too much is invested in the ability of examinations to drive standards. In the course of this, important quality levers are forgotten.

### **Introduction**

This paper explores the pervasive and hegemonic dominance of assessment in reform efforts to improve the standards and quality of education. Current educational reforms in many countries around the world call for the establishment of clear performance standards for schools, and results of large-scale assessments are increasingly reported in terms of the percentages of learners ready to meet minimum standards. A crucial question is whether standards-based reporting is adequately understood by non-technical audiences—such as the

press, policy makers, and the public—that are among the most important consumers of assessment results.

A stream of reports and pronouncements has fuelled the popular perception that the South African education system is in crisis. Real and imagined declines over time in performance in various forms of assessment; such as the Senior Certificate Examination, the systemic evaluation tests, cross national comparisons on tests like the South Eastern and Eastern Consortium for Measuring Educational Quality (SACMEQ), and comparisons to benchmarks such as TIMSS and MLA have been publicized as evidence that our educational woes continue unabated.

The use of test scores to index educational success or failure is rarely ever questioned. Low scores are bad news; and high scores are good news. In the rhetoric of education reform, it often sounds as if improving the education system is synonymous with improving test scores. In such a climate, the logic of high stakes testing is compelling. The logic is so seemingly simple: Test learners and see what they can do and hold schools accountable if they fail to make the grade. Rather than micro-manage schools, policy makers can dictate that content standards and performance standards be created to codify expected educational attainments.

The question posed in this paper is whether this obsession about testing as a means to improve educational quality and standards works.

## **The issue of standards**

The term 'standards' has become an important rallying point for educational technocrats. It has become emblematic of reform efforts aimed at improving quality and standards in education. The concept is, however, notorious for its elasticity, elusiveness and malleability. The use of the term in the public rhetoric often masks its real complexity. It can have several meanings:

- A standard can relate to a criterion score on a test. Both the methods of scoring and the decision about setting the criterion level depend on somebody's judgement of what is satisfactory or what constitutes a high or, worse still, a world class standard.
- A high standard could refer to the height of the hurdle or to the proportion of students who can jump it and there is an obvious trade-off between the two. A common restriction of meaning is to equate high standards with test performance.
- A national standard might be a minimum of performance that all should attain, or a scale of performance levels on which each individual will be placed (Black, 1998; Linn, 2003; John, Hoover and James, 2004).

Conceptions of standards can, therefore, be based on content standards which describe in specific detail what content should be covered, how and at what level of complexity. Sometimes standards are perceived as measures of outcomes where specifications about the competencies to be displayed by learners and their levels of demonstration are specified. Standards are also conceived as the quality of practice inside the classroom; often referred to as 'opportunity to learn'. Finally, standards may be conceived in terms of the quality of

inputs or resources devoted to the system in terms of schools, teachers and instruction. Yet when we talk about standards, we talk as if we all understand the same thing.

The drama on standards plays itself out in the public psyche in deeply symbolic ways on a variety of stages, enacted by different characters and it demonstrates varied texts and subtexts that are nonetheless all embedded in the meta-text of quality education. Although punned for the ostensible reason of improving the quality of education, standard-based assessments serve mainly to hold those involved in education to account.

## The tail and its many dogs

The main aim of assessment is to provide evidence of the quality of performance in a given outcome. This evidence is based on judging a carefully selected sample of that performance. This does not seem to raise too many problems. However, the moment we begin to think about the reasons for assessing and why we assess in one way rather than the other and the purposes to which the outcomes of assessment are used, many problems arise.

Assessment can be used for a variety of purposes: to help learners, parents and prospective employers to make educational and career choices, to provide a basis for selecting individuals for specific sorts of further or higher education, or for certain jobs, and for qualifying individuals to undertake specific vocational or professional activities as well as to exclude those that do not qualify.

All the aforementioned, in my perjorative and rather irreverent inversion of the common reference to assessment as the **tail that wags the dog**, are the 'dogs' that the tail that is assessment wags.

The main complexity that arises out of the purposes to which assessment is used is when it is used to hold those involved in education to account. The internal uses of assessment for diagnostic purposes, or its formative uses to improve the learning of those assessed are less problematic than when it's used for accountability. Assessment for accountability is a 'high stakes' affair.

I will now consider how assessment has been used as a pawn in the game of standards.

## Assessment and the scramble for standards

To add insult to injury, assessment is frequently used to draw conclusions on standards and the quality of education. Assessment becomes the main driver of standards and quality in education: the tail that wags the dog, as the common parlance among assessment professionals refers to it. It is used interchangeably in both formative and summative ways to make judgements on standards. Assessment scores are often interpreted through making a complicated set of comparisons:

- Comparisons over time
- Comparisons between and among examining boards
- Comparisons across different subjects

- Comparisons across and among schools
- Comparisons across and among districts and provinces
- Comparisons across and among racial groups
- International surveys
- National surveys.

This desire to use assessment to drive the quest for standards and quality education is a worldwide phenomenon manifested in the surveys, conducted internationally, regionally and nationally to measure the quality of educational attainment.

In the late 1950s the International Association for the Evaluation of Educational Assessment (IEA) was formed. It initiated what would be major studies aiming to measure the quality of cognitive achievement at various levels of education in several countries and to identify the main causes of differences in achievement of educational outcomes. Twelve countries joined its first mathematics study. By 2000, fifty countries were participating in surveys covering mathematics and science, now called the Trends in International Mathematics and Science Study (TIMSS), science and reading, known as Progress in International Reading, Literacy Study (PIRLS).

The IEA surveys precipitated a scramble for standards. Strongly influenced by the IEA experience, several other such studies, usually of regional focus, have since been established. These include the Programme for International Student Assessment (PISA), set up by the OECD in 1998 and now covering fifty-nine mainly first and second world countries; the Southern Eastern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) which, since its first survey in Zimbabwe in 1991 has expanded to fifteen African countries; the Latin American Laboratory for the Assessment of Quality Education (LLECE), which began in 1997 and covers sixteen countries; the UNESCO Monitoring Learning Achievement (MLA) project and the survey in French-speaking Africa known as PASEC. (EFA Global Monitoring Report, 2005: 44-46)

Not to be outdone, South Africa has joined the fray to participate in some of these surveys typifying the scramble for standards: TIMSS, SACMEQ and MLA. As if this was not enough, two national surveys called systemic evaluation at grades 3 and 6 have so far been conducted and more are in the pipeline.

The question that is still begging is: **Has this madness about testing helped to improve the quality of education?** Not really. That would be my answer. The immense potential that assessment has to drive instruction is undisputed. However, many educationalists are less confident of its potential to improve educational quality (Chisholm: 2004, Muller: 2004, Kanjee: 2005, O'Neil: 2005, Wolf: 2005, Lolwana: 2006). Tests of cognitive achievement are incomplete proxies for the quality of education.

The results of these surveys are, however, very instructive on factors that influence educational quality and standards. I will now turn to some of these factors.

## **Factors influencing standards and quality of educational attainment**

Each study made great efforts to identify the most salient and significant factors that influence standards and quality education (EFA Global Monitoring Report: 2005). All the studies referred to above show that nearly all education systems, pupils' home background proved to be crucial. Learners from higher socio-economic backgrounds where parents had higher levels of education and were thus materially well off tended to perform better than their counterparts from poorer homes. In Africa and Latin America, the urban-rural divide exacerbated the problem.

In many developing countries, the material resources in schools are inadequate. This comes out very clearly in the SACMEQ studies where the average child was in a school with 8.7 of the twenty two desirable school resource items, with the disparities ranging from 4.3 items in Malawi to 16.7 in the Seychelles, this complicated by wide variations in urban-rural variations within countries. Even in those countries where a relative degree of equity in the distribution of material resources had been achieved, the teachers in urban schools tended to be better qualified than their rural colleagues.

The gender of primary school teachers was found to have a profound influence on performance; particularly of girl learners. The SACMEQ studies showed wide discrepancies by country. For all countries, 53% of pupils surveyed were taught by female teachers, on average, but the share ranged from 17% in Uganda to 99% in the Seychelles. Learners taught by females scored, on average, three tenths of a standard deviation higher than pupils with male teachers.

Other items prominent in the African studies were behavioural problems of pupils and teachers: late arrival, absenteeism and learner dropout were all correlates of poor performance. In the PISA studies where socio-economic school climate, teacher morale and commitment, school autonomy, teacher-pupil relations and disciplinary regime had some compensatory influence towards greater equity. In Latin American countries covered in LLECE studies, pupil socio-economic background and classroom climate appeared to be the most important predictors of achievement.

Forty years of research conducted through the IEA studies reveal some major findings relating to the quality of educational attainment. Three of these findings are particularly instructive for policies aimed at improving educational quality. The first is that the distribution of cognitive abilities in the population has a significant impact on average achievement levels. The greater the overall proportion of children enrolled, the lower the average achievement levels tend to be. The achievement levels of particular cohorts of ability, however, are not affected—the cognitive achievement levels of the most able cohort remain unchanged by expansion. The second is that time spent actually working on particular subjects, either in school or as homework, affects performance, especially in mathematics, science and languages. The third is that although socio-economic status is influential in determining achievement in all contexts, textbook availability and school

resources appear to be capable of countering socio-economic disadvantage, particularly in low income settings.

## Conclusion

Reform efforts aimed at raising the standard of education have relied too heavily on using summative assessment as a lever. However, these efforts have not been as successful as it was hoped they would be because they ignored the formative purposes of assessment. A small, but growing group of researchers are now turning their attention to assessment embedded in the context of the classroom. They argue that more emphasis should be laid on formative assessment as this allows for feedback to the learner and thus enhances the opportunity for learners to improve their performance. The cumulative effect of this will be a steady improvement of the learners' overall performance which would in turn improve the quality and standards in education. While standards-based assessment may tell us how many learners can hurdle the bar we set for them, it contributes little to the strategies for improving teaching and learning in the classroom.

## References

- Black, P (1998) *Testing: Friend or Foe? Theory and Practice of Assessment and Testing*. London: Falmer Press
- Chisholm, L. (2004) 'Public Debate Lacks Quality Information and Analysis'. In *HSRC Review*, 2 (1) March 2004
- John, J, et al. (2004) 'Differentiating Standards-Based Education for Students with Diverse Needs'. In *Remedial and Special Education*, 25 (2) March/April 2004, pp 74-78
- Linn, R.L. (2003) Performance Standards: Utility for Different Uses of Assessments. In *Education Policy Analysis Archives*, 11(31). Retrieved 19 May 2006 from <http://epaa.asu.edu/epaav11n31/>
- Lolwana, P. (2004) 'What is the Standard of the Senior Certificate Examination?: Umalui's Report in Context'. A Paper Delivered at the HSRC Matric Colloquium on the 3rd of November 2004.
- Muller, J. (2004) 'Assessment, Qualifications and the National Qualifications Framework in South African Schooling'. In Chisholm, L (ed). *Changing Class: Education and Social Change in Post Apartheid South Africa*. Cape Town: HSRC Press
- O'Neil, O. (2005) 'Assessment, Public Accountability and Trust'. A Paper read at the First Cambridge Assessment Conference, 18 October 2005
- Wolf, A. (2005) 'What can and should we be Measuring?' A Paper read at the First Cambridge Assessment Conference, 18 October 2005