

## **High Stakes Testing in South Africa: friend or foe?**

Sarah J Howie  
Centre for Evaluation and Assessment  
University of Pretoria

The Jomtien conference in 1990 on Education for All is seen by many as a turning point for the introduction of increased monitoring and evaluation of the quality of education systems around the world. Internationally debates have arisen about nature and frequency of assessment and its impact on education systems with its intended and unintended consequences.

The phenomenon of large-scale testing is a relative latecomer to South Africa education. Since 1994 large scale assessments have been implemented in core subjects such as mathematics, science and language in national and international assessments. In this paper, various forms of the large-scale assessments in South Africa are discussed in relation to high stakes testing and their effect on the education system as a whole in the light of international experience.

## **High Stakes Testing in South Africa: friend or foe?<sup>1</sup>**

The Jomtien conference in 1990 on Education for All is seen by many as a turning point for the introduction of increased monitoring and evaluation of the quality of education systems around the world. Internationally debates have arisen about nature and frequency of assessment and its impact on education systems with its intended and unintended consequences.

The phenomenon of large-scale testing is a relative latecomer to South Africa education. Since 1994 large scale assessments have been implemented in core subjects such as mathematics, science and language in national and international assessments. In this paper, various forms of the large-scale assessments in South Africa are discussed in relation to high stakes testing and their effect on the education system as a whole in the light of international experience.

High stakes testing, large-scale assessment, national assessment

## **Introduction**

Governments in developing countries have increasingly turned to assessment since the 1990 Jomtien World Conference “Education for All”, where the emphasis in education significantly shifted from measuring inputs to an increased emphasis on

---

<sup>1</sup> The idea of testing being either a friend or foe was first mooted in Black, P. J. (1998). *Testing, Friend or Foe? The Theory and Practice of Assessment and Testing*. Falmer Press: London

educational quality outcomes, in order to ascertain the extent to which their education systems are meeting the need to deliver quality in education. (Kellaghan and Greaney 2001). The term assessment in education in this paper is defined as “any procedure or activity that is designed to collect information about the knowledge, attitudes or skills of a learner or group of learners” (Kellaghan and Greaney, 2001 , p.19).

The impact of the 1990 Jomtien conference and the 2000 Dakar World Education Forum has been considerable as drivers behind the expansion of assessment internationally (Ross and Genevois, 2006). Both these events called on nations to broaden their concentration on access to education by aiming to “improve all aspects of the quality of education and ensure excellence so that recognised and measureable learning outcomes are achieved by all (Ross and Genevois, 2006, p.26). Furthermore, as the interest in measuring outcomes increased, explicit linkages between educational outcomes and quality were made to the notion that these were essential for educational development within a global economy (Ryan and Feller, 2009) and the increased political interest has led to direct links between educational quality and economic imperatives. High profile international agencies such as OECD and UNESCO are promoting this thinking through statements suggesting that quality of education has an influence as to how quickly societies can become richer and how individuals can improve their productivity and income (Ross and Genevois, 2006)

There are a number of different purposes that assessment serves and different characteristics dependent on the level of information required (see Table 1). For instance, at the student level, it can be used to describe students learning and to diagnose learning problems whilst at the system level the main purposes would be to

reach a judgement on the effectiveness of an education system or part thereof, which is primarily the interest of governments and policymakers. It is essential that the nature of the assessment should follow from the intended purpose (Macmillan, 2001, p.4).

<Insert Table 1 about here>

Serious questions have been raised (Nichols & Berliner; Ryan and Feller, 2009 ; Tamassia & Adams, R. J. (2009) ; Torrance (2009) ) about the relative merits of system assessment (also known as national assessment, learning assessment and assessment of learning outcomes (see Kellaghan and Greaney, 2001, p.33) or testing on a large scale (which uses standardised assessments on district, province, state, national or international levels (Macmillan, 2001) being employed across the Western world, where national assessment of some kind has been in place for many years. A national assessment is designed to describe the achievement of students in a curriculum area aggregated to provide an estimate of the achievement level in the education system as a whole at a particular age or grade level (Greaney and Kellaghan, 2008 , p.7) and is normally conducted on either a sample or a whole population of students. System assessments (also known as systemic evaluations in South Africa) are primarily concerned with quality in education, a dynamic concept (Ross and Genevois, 2006, p.41) and “frequently mentioned and rarely defined” (Kellaghan and Greaney, 2001, p.22). Here it is taken to mean assessing “the adequacy or appropriateness of objects or processes for the purposes for which they were intended” (Ibid, p.22). Assessment in terms of the kind used by the reform movements (post Jomtien) suggest that the type of assessment that is likely to impact

on quality is the one that focuses on outcomes, is conducted externally to the school and the expectation that the assessment will act as a lever of reform (ibid, p.29)

Furthermore, concerns about the consequences of high-stakes testing (namely “those assessments that have serious consequences attached to them” Nichols and Berliner, 2008, p.xv) have arisen in the USA (Bracey, 2000 ; Clarke, Haney, Madaus 2000; Nichols and Berliner, 200; Jones, Jones & Hargrove, 2003 ; Kohn, 2000; Orfield & Kornhaber, 2001 ; Ryan, 2004 amongst others) which has a long history of standardised testing. These concerns have grown since the advent of the No Child Left Behind policy. In the USA, high stakes testing can result in certain sanctions being implemented at school level. For instance, these can include school choice with students being able to move to another school, or to schools being forced to close. In a number of states rewards have been introduced even linked to teacher salary increases (e.g. Tennessee). This is also the case in Argentina where high performing schools received additional money or other rewards and in Chile where honoraria are given to the teaching staff based on increased test scores. There are a number differences between Chile and the USA, with one notable difference being that Chile provides additional resources to lower performing schools and the USA does not (Ryan and Feller, 2009). The situation in the USA has led to what some call high-stakes educational accountability systems (Linn, 2008; McDonnell, 2008; in Ryan and Feller, 2009) with undesirable and often unintended consequences. Outcomes-based educational accountability systems are those that involve some mandated form of systematic assessment based on educational standards (e.g.: content) (Ryan and Feller, 2009, p.176). Conversely low-stakes outcomes-based educational accountability systems (Linn, 2008; McDonnell, 2008; Ryan, 2005, in Ryan and

Feller, 2009) are being promoted by the critics of high stakes testing in the USA. Ryan and Feller (2009) refer to low-stakes outcomes-based educational accountability systems providing descriptive information about student achievement levels and school performance suggesting that this information may actually motivate teachers, principals, communities to improve schools performance (p.177). They quote the example of Uruguay's national assessment where the results are used to describe student achievement and the public receives the descriptive information about the national performance whilst the school-level information is reported only within the education community and not to the public resulting in a "low stakes" position. Linn (2008 ) claims that low stakes alternatives are receiving considerable attention in the USA and Ryan and Feller state that these alternatives have a number of attractive features but feel that it might be difficult for the USA to move away from its current high stakes position.

This has also been the case where prominent educationists in the UK (see Assessment Reform Group; Black and William, Torrance, 2009) have stirred debates about the use of high stakes testing versus the benefits of a more low stakes testing approach and the combination of school self-evaluation (where schools have the opportunity and means to evaluate their own teaching and learning (e.g.: Macbeath, 2002) methods being primarily used across Europe (see Ryan and Feller, 2009). Torrance observed a plateau effect in the National Testing for the age 11 tests in English, mathematics and science between 2000 and 2006 and noted that Linn (2000) had noted this previously in the context of the USA and its implementation of minimum competency testing (2009, p.489). For many years, the Assessment Reform Group in the UK comprising Paul Black, Mary James, Gordon Stobbart and Dylan William amongst others had

written about the unintended consequences of national testing (Black, 1998) and had advocated instead an increase in enhancing formative assessment in the classroom and indeed had shown research results about the improved teaching and learning as a result (Black 1994, Black 1998).

Recently, Torrance (2009) argues that a review of the evidence from “both sides of the Atlantic” is that ‘testing does impact on the curriculum but that it narrows the curriculum to that which is tested and, in doing so, probably lowers rather than raises educational standards (p.488). However, Torrance (2009) does make the point that the negative consequences of the test-based reform does not mean that “there is no validity or potential benefits in using test results in evaluation studies and international comparative studies. Good quality outcome measures are important if we can produce them” (p.491).

To a large extent, South Africa has been absent from many of the concerns and debates given its short history in large-scale testing and late entry to international comparative assessments and national assessments. South Africa achieved its first democratic government in 1994 and conducted its first international comparative study (TIMSS 1995) and its first national assessment was administered in 2000. High stakes testing has been limited to its national final year school exit examinations (at Grade 12) but with the increased emphasis on testing being given by the latest Ministry of Education under the current President Zuma, this situation may change. Given the international debates on testing and the intended and unintended consequences of high stakes testing in particular, how can South Africa utilise its expanded national testing programme optimally whilst avoiding some of the problems

noted in the international lessons. With reference to the international debate, this article addresses the question to what extent can large-scale testing in South Africa be considered as a friend or foe and how can South African benefit most from large-scale testing?

In this article, the various forms of the large-scale assessments in South Africa are discussed in relation to high stakes testing and their effect on the education system as a whole in the light of international experience. Given the resistance growing in Western countries to high stakes testing (see Nichols and Berliner, 2008, Torrance, 2009), to what extent can large-scale testing in South Africa be considered as a friend or foe and how can South African benefit most from large-scale testing?

This article is structured in the following way: firstly a brief overview of the South African education system is provided, followed by a description of the types of large-scale assessment being implemented in South Africa starting with the national assessments. This is followed by the experience of the international assessments is described and thereafter the national examinations. Finally, there is a discussion of these assessments in terms of the questions outlined above.

## **South African Education context**

South Africa emerged from an oppressive political regime during the 1990s and had to undergo significant reconstruction of its society. This coincided with the materialization of South Africa as a significant economic power in Africa which is also seen as an emerging economy and is classed as an Upper Middle Income country (World Bank 2008). Despite the seeming prosperity, there are great economic

inequalities within the country (in addition to the social disparities) (see Table 2).

South Africa is considerably ahead of most countries in Africa with a GNP per capita of US\$ 3690, but yet 34% of South Africans live on less than US\$2 per day (UNESCO 2007:229).

<Insert Table 2 about here>

In terms of the amount spent on education, South Africa spends a relatively high percentage of its GNP (4.1) which compares well to most African countries, with the exception of Kenya (7.1%) and Lesotho (7.3%). South Africa has also, since its new democracy, traditionally spent a large percentage of the budget on education although it has dropped from 22% to 19% in the past few years (UNESCO 2007, p.320).

The number of children in primary education in South Africa has gradually been increasing over the past couple of decades and the Net Enrolment rate is 89%. (UNESCO, 2007). The South African school population grew faster between 1970-1995 than those in Sub-Saharan Africa as a whole, Europe or the USA from 30% to well over 90% in just 25 years (Crouch and Vinjevold, 2006). Overall, 89% of eligible pupils in primary school attend school, which is a higher rate than that of most developing countries. South Africa has more than 26 000 ordinary schools. Of these, about 19 000 are primary schools, nearly 6 000 are secondary schools, and there are over 1 000 combined schools (Grades 1 to 12) and intermediate schools (both primary and secondary grades, but not all the grades). Class sizes regularly exceed 40 pupils in a class at secondary school level which contravenes government policy (Howie, 1997, Howie 2001).

South Africa faces many infrastructural challenges in its government schools. Approximately 40% of schools are classified as poor or very poor. Forty percent of South African schools do not have electricity and 33% have no telephone lines (Department of Education, 2006). However, South African independent schools (private schools) are well-resourced and furthermore use latest technology in their instructional activities.

It is against this background that the challenges regarding educational quality and the need to monitor the education system can be understood.

## South Africa's experience of large-scale assessment

Although the Human Sciences Research Council had a programme to develop standardised instruments in key curriculum subjects and psychological tests between the 1960s through to 1990s, large scale testing in South Africa can be traced back to the arrival of the new political dispensation under the leadership of Nelson Mandela, in South Africa's history in 1994. With the new dispensation came the controversial outcomes based education approach (Jansen, 1997) and a new curriculum aiming to provide all children in South African with equitable, quality access to education. In order to monitor changes occurring in the system formally, a system of Systemic Evaluations (internationally known as National Assessment) was introduced at the turn of the new century for Grades 3, 6 and 9 where large-scale testing (sample-based) in mathematics, science and language . According to the National Education Policy Act 27 of 1996, monitoring and evaluation should be done with the objective to assess

progress (at system level) that corresponds with the provisions of the Constitution of the Republic of South Africa and the National Education Policy.

In addition to the act mentioned above, provision for conducting systemic evaluation on a nationally representative sample of learners and learning sites, was made in Section 48 of the Assessment Policy for General Education and Training (DoE, 2001). This is done in order to evaluate all aspects of the school system and learning programmes contained therein; i.e.: to assess the effectiveness of the entire system and assess the extent to which the vision and goals of the transformation process are being realised.

However, even prior to the introduction of the Systemic Evaluation, a number of international comparative studies had taken place (Third International Mathematics and Science Study (TIMSS) 1995, 1999; Southern African Consortium for the Monitoring of Educational Quality (SACMEQ) 2000; and Monitoring Learning Achievement (MLA) also inspired by the changes taking place in the system but also because of the international recognition of political changes in South Africa and the reintegration of South Africa into the international community.

The international comparative assessments arose in a vacuum of testing in the mid-1990s where South Africa had traditionally only included national exit examinations in the final grade of school (commonly known as matriculation exams), now Grade 12 and known as National Senior Certificate (NSC) as a form of external large scale assessment. The stakes for this examination are very high and thus for some time it was regarded as the only form of high stakes testing that was available in the country.

In the following sections, each of form of large-scale assessment, namely national assessments, international assessments and public examinations are discussed more in-depth in relation to their implementation in South Africa.

## **National Assessments**

Systemic Evaluation in South Africa was introduced in 2000 as a series of sample-based national assessments every three years that take place alternatively at Grades 3, 6 and 9. To date these have been administered at Grades 3 and 6 but not yet at Grade 9 level. Systemic Evaluation aims to measure the effectiveness of the education system by assessing the components of the education system at the aforementioned Grade levels. The Systemic Evaluation provides and implements a national framework for the evaluation of the education system as well as to develop benchmarks from which performance can be interpreted. The focus of the assessment is language (literacy at Grade 3), mathematics (numeracy at Grade 6) and science.

According to the existing legislation in South Africa, the Minister of Education is decreed with the task of monitoring the provision, delivery and performance of education standards. Systemic Evaluation endeavours to measure both learner performance and monitor the teaching and learning context. It therefore assesses the extent to which the education system achieves the desired social, economic and transformational objectives. According to the Assessment Policy, Systemic Evaluation should be conducted in Grades 3, 6 and 9 of the education system and is a response to the National Education Policy Act 27 of 1996, which requires the Minister to monitor

and evaluate standards of education provision delivery and performance throughout the country.

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

The necessity for appropriate and adequate quality assurance systems in education to ensure the attainment of the goals of transformation was highlighted in the conception of the new post-1994 Department of Education. The policies on Systemic Evaluation aimed to evaluate the performance of the entire education system (DoE, 2001).

Systemic evaluation is generally guided by the following principles:

- integration with other quality assurance initiatives;
- practicality of the design of the programme;
- collaboration between the provincial and national departments of education;
- capacity-building for systemic evaluation and self-evaluation;
- gathering and using information to improve education provision and delivery; and
- ensuring inclusivity through the active participation of learners with special education needs.

None of the principles above suggest that South Africa's national assessment are of a high stakes nature. In the past decade, since the Systemic Evaluation Systemic Evaluation programme was initiated the National Department of Education together with provincial departments, managed to conduct two cycles of Systemic Evaluation at Grade 3 in Literacy, Numeracy and Life Skills and one cycle at Grade 6. In addition to the assessment of learner achievement in the three Learning Areas using a set of assessment tasks, structured questionnaires were administered to learners, parents, principals, school-based educators and district managers in order to gather information about the context in which learning and teaching takes place. The questionnaires were based on 26 agreed upon education indicators designed to inform policy formulation.

The first national systemic evaluation was conducted at the Grade 3 level in a sample of schools across the country in 2001. The study was thereafter also extended to the nodal areas, which had been identified in the Rural Development Integrated Strategy. Primary schools in the nodal areas were evaluated at Grade 3 in 2002 using the national Systemic Evaluation instruments. A second national Systemic Evaluation study was conducted on Grade 6 learners in 2004 in Language, Mathematics and Natural Sciences. The collected information from these two evaluations at Grade 3 and Grade 6 level served as baseline against which to assess the impact of a plethora of activities, especially in the nodal areas, that aimed at improving learner performance.

## International Assessments

At the time that South Africa entered its first international study in 1994 (TIMSS 1995), it was emerging from decades of political and social isolation and a legacy of inequity.. The South African education system was fragmented into 19 different systems with an unequal and vastly different quality of education being offered. This inequality was evident in the performance at the Grade 12 level with the exit matriculation examinations that were organised within different education departments and therefore different examination bodies. With the dramatic expansion of the education system in the years after 1994 and the focus on access to education (as already stated), there was an urgent need to ascertain the quality of education at all levels across the system particularly given that there was no systemic overview of the schooling system with regard to education quality was available in 1994. Having an “external assessment” such as TIMSS 1995 provided unique opportunities for South Africa at the time

South Africa entered its first international study (TIMSS 1995) without the ideal conditions and resources for conducting such a study. These conditions included that the provincial boundaries had not yet been established and that no central database existed containing all the names of the schools. The research team had no prior experience of conducting a large-scale assessment nationally. In short the learning trajectory was considerable and the challenges phenomenal. The Third International Mathematics and Science Study was conducted with many challenges between 1995 and 1998 with data collected in 1995 in South Africa<sup>2</sup> (see Howie 1997, Howie and Hughes, 1998). TIMSS 1995 has since been followed by a number of international

---

<sup>2</sup> Data was collected at the end of school year - in Northern Hemisphere in 1994 and in the Southern Hemisphere in 1995.

studies (TIMSS 1999, 2003; Progress in International Reading Literacy Study (PIRLS) 2006) and regional studies Southern African Consortium for the Monitoring of Educational Quality (SACMEQ) II and III.

However the introduction of the international and regional studies was (and still is) also controversial and a sensitive issue in many circles. The results of TIMSS 1995 produced outrage in different circles starting with the Department of Education who had difficulty publicly accepting the very low performance of the South African learners. The business community and public were shocked that the top performing learners in the country did not compare with the average learners of the top performing countries. TIMSS 1995 and subsequent international and regional studies revealed both the damage of past political policies, but also the difficulties of implementing effective change in teaching and learning in South African schools since 1994. However it provided South Africa with a set of educational quality indicators and benchmarks against which to review the performance of the system.

South Africa was the only African country to participate in TIMSS 1995 whereas by 2003, Botswana, Egypt Ghana joined South Africa, Morocco and Tunisia who participated in the TIMSS 1999 study. Then in 2007, it was announced that South Africa would not be joining the TIMSS 2007 study. Controversy surrounded this announcement and a number of differing reasons emerged from the previous national organizers of the study and the government. The lead research organization suggested that deferring the study was because of the timing of the research versus the implementation of interventions saying that

*“The South African education system has undergone radical restructuring in its recent past, as several initiatives and interventions have been introduced – each one sharing the*

*common objective of improving teaching and learning in all areas of the curriculum, but especially in mathematics and science. Bearing in mind the strain this intervention has put upon the education system (and more pertinently, the educators themselves), it is recommended that South Africa does not participate in TIMSS 2007, but rather does so in 2011 as this will allow the interventions to become embedded within the education system. This achieved it would then be more reasonable to measure South African performance in TIMSS 2011 to see how far the country has progressed” (Reddy, 2003, p.120).*

On the other hand, the press suggested that the government did not want to expose itself to further humiliation. Members of the broader education community regretted this decision as a missed opportunity to continue the external monitoring the quality of education.

Subsequently, South Africa participated in the Southern African Consortium for Monitoring Education Quality second study in 2001. Seven African countries participated in the first study (implemented in the mid-1990s) which published its studies within individual reports. This avoided the possibility of data from the seven countries being compared. It had been explicitly agreed that the data would not be published in a comparative form as this was not the purpose of the SACMEQ studies. South Africa was only able to participate in the second study (as one of 14 countries) after a Quality Assurance Unit had been established within the Department of Education, as only government-related organizations can participate in the SACMEQ studies. Although not initially intended as a comparative study, comparisons across African countries were made and South African pupils performed compared to most of their African peers raising further questions about the quality of education in South Africa.

In 2004, South Africa embarked on the second study of the Progress in International Reading Literacy studies, PIRLS 2006. It was one of only two African countries in this study (the other being Morocco who was previously the only African country participating in the PIRLS 2001). As with TIMSS, South Africa's performance was very poor and the South African Grade 5 pupils achieved the lowest scores compared to all other participating countries that had actually tested Grade 4 pupils. This provoked harsh criticism of the system nationally leading to the critical questioning of the Minister of Education in a parliamentary session in May 2008 (Government of South Africa, 2008). As South Africa's participation in the PIRLS 2011 is imminent, it remains to see what lessons will emerge.

## **National Examinations**

“Assessment in South African schools has been dominated by the Senior Certificate examinations” (Lubisi and Murphy, 2002, p.260). These are national examinations administered at the end of schooling (Grade 12) where almost half a million students write annually. The results have significant and far reaching implications for students' opportunities after school in terms of their future careers. Even more difficult is what has been termed, the annual South African circus (Jansen, 2008) where the media generate a feeding frenzy amongst the public and communities around the schools annually before, during and after the examinations. The annual pass rate - the subject of much speculation and media frenzy - has fluctuated greatly from 58% in 1994 down to its lowest post-1994 of 47% in 1997 up to an unbelievable and discredited 73% in 2003 (Taylor, 2009) and has dropped back to around 60% in 2009. The university exemption rate allowing students to apply to university however hovered

between a low of 12% (1999) and a high of 19% (2003) between 1994 and 2007. In 2008, 554 664 Grade 12 students wrote the first National Senior Certificate examination based on the new curriculum that had been introduced in Grade 10 in 2006 and therefore the class of 2010 is the third cohort that have written this examination with the results due in January 2011 with an increase of 10.4% enrolled for examinations (642 154 pupils) compared to 2009 (DoE, 2010). When the pass rate of 62% was announced for the 2008 examinations there was widespread criticism (Reyneke, Meyer and Nel, 2010, p.277). In 2009, the pass rate was 60.7%, the lowest in 10 years but based on a different and new curriculum and the examinations that were intended to uphold standards and require increased cognitive demands.

The poor performance was attributed by Reyneke, Meyer and Nel (2010) to a lack of resources, lack of student discipline, lack of student commitment, lack of educator discipline, commitment and morale, ineffective policies at school level: weak organisational structures and undeveloped managerial skills, teacher union disturbances and problems in implementing collective agreements, problems in implementing government policies, poor organisational structures and inadequate parental involvement. In response to some of the identified problems in the system, a multitude of national intervention strategies were initiated including the provision of learning and teaching materials, increased monitoring, targeted support programmes for schools, targeted guidelines for specific subjects on national level in addition to a variety of provincial strategies (DoE, 2010).

National Examinations have the highest stakes at school level and have been used as the only measure of the quality of education prior to the inception of national

assessments and international studies described above. The Department of Education is planning to introduce examinations at the end of grade 9 (end of compulsory education). The key knowledge areas, skills, attitude and values are assessed in line with assessment practices of Outcomes Based Education (DoE, 2003c) however changes to the curriculum mean that this is currently under review.

## **Discussion**

What have we learnt about the state of health of the South African education based upon over a decade of large-scale assessments? In the following section, this health is discussed in relation to some key functions of large-scale assessments mentioned by Howie and Plomp (2006), viz. *mirror, monitoring and enlightenment*.

To date the *mirror* (see Howie and Plomp, 2006) presents the country with an “ugly face” of systematic and systemic failure in education. The problems of the past still haunt the present and the new government has yet to see positive effects (on achievement) of the policy developments since 1994. Perhaps the conventional wisdom that it takes 20 years to change an education system is valid, also for South Africa. It would appear that our expectations are not realistic enough. Nonetheless with the extensive interventions put in place (see DoE, 2010) what we should be seeing however is the attainment of key indicators indicating that the basic and essential conditions have been attained so that ultimately learners’ achievement will improve and on the infrastructural side, one sees progress (see Howie, 2008).

*Monitoring* (see Howie and Plomp, 2006). our education system through international studies has revealed that there is a danger of the mathematics results as measured in the TIMSS studies declining. The monitoring (also revealed through the Systemic Evaluation SACMEQ, TIMSS and PIRLS) even more importantly has failed to reveal any improvement despite all the activities since 1995, hence the comment above about the time needed to observe change in an education system. What the monitoring has revealed however is that from TIMSS 1995 through to the latest study, PIRLS 2006, that the apartheid system produced two different education systems in terms of quality and in terms of having conditions that represented both a developed country and a developing country “two countries in one” (Howie, 2002; Fleisch, 2008) and the bimodal distribution of the data in each study (Howie 2002) is evident. A small cluster of learners at the very top end achieving the highest benchmarks in each study and the majority (80% and more) who do not achieve the lowest international benchmarks or lowest proficiency levels (see Howie, 2002). It is therefore very important that studies such as SACMEQ III, TIMSS 2011 and PIRLS 2011 be undertaken to fulfil this monitoring function.

Decision-making taking place based upon the international studies and national assessments has been difficult to discern. Whilst there are decisions made and events that follow the international studies and their outcomes, it is not always easy to categorically link these to the studies themselves. For example, the curriculum revision was underway when the TIMSS 1999 results were released. Those involved in the curriculum revision process requested the reports and findings and fed these into the proposals and decisions made regarding the General Education and Training band curricula for mathematics and science.

The dissemination of the national reports and the subsequent secondary analyses in particular serve the purpose of *enlightenment* (see Howie and Plomp, 2006). The broader community, politicians, the business community and industry, the media, education organizations, non-government organization as well as the education system become involved in the discussion about the results and education in general.

In summary, the main findings of the international studies, national assessments and national examinations suggest that the country has not been able to overcome its deprived legacy; that the new policies have not yet been implemented effectively or widely yet and that the country has not yet seen the fruits of its priorities in terms of education quality. However it does seem that to a large extent its priority of access has been achieved but given the economic and social conditions, continued monitoring of access is crucial.

Despite achieving access, South Africa is struggling to attain its goal of equity. The non-achievement of the international benchmarks across all the international studies by South African learners reveals that the level at which South African learners are able to operate educationally is two years and even up to four years behind what the curriculum stipulates. After more than ten years of participating in the international studies in mathematics and science, South African learners have not been able to close the gap on their peers in even the other developing countries participating.

Secondary analyses of these international studies also reveal the predictors of achievement in South African schools and provide deeper insights for the poor

performance or alternative better performance (for instance Howie, 2002; van den Berg and Louw 2007). The impact of the international studies in particular has differed over the years as has the reaction from various parts of the broader society. The results from the first international study “TIMSS 1995” led to a national concern and were interpreted as a real problem with the secondary school. It was seen as a “sign of weakness in secondary school mathematics and science teaching rather than as a symptom of the crisis in primary education” (Fleisch, 2008). The Department of Education at first adopted an adversarial role as this was a new experience to have the defects (of albeit an inherited system) so publicly revealed by an external assessment. However, as the Department recognized the benefits of such large-scale assessments, their initial rejection of the role of international studies converted to an acceptance and collaborative stance. Whilst Prof Asmal, then Minister of Education, declared (at a national launch of the systemic evaluation report in 2001) that comparisons (of achievement) were odious, nonetheless apparently note was taken of the findings from TIMSS 1999 where there was an enormous variation in number of school days (between 120 and 280 days at different schools) and school days as short as four hours. Shortly afterwards, Prof Asmal gazetted 200 school days per year with teachers having to be on the school premises for 7 hours a day. Furthermore, the TIMSS 1999 data on the international and national curricula for mathematics and science was available at the time that the curricula were being revised and was used by the committees were able to feed some of the findings related to the international curricula directly into the revision of the RNCS

It was in fact the third Minister of Education after 1994 (Minister Pandor) who appeared to recognize the value of various types of large-scale assessments nationally,

regionally and internationally. She of all the ministers supported these assessments whilst recognizing the difficulties related to their implementation and associated publicity. The national department now manages the SACMEQ studies directly and gave support to the PIRLS 2006 study. Researchers nationally have come together increasingly in order to improve the quality of the large-scale studies (e.g. PIRLS 2006 had a national committee comprising leading organizations in the field and experts in reading and assessment). The very poor results in PIRLS 2006 were lower than could have been predicted and were a severe blow to the Minister and the department, who a year after the study started with important initiatives in reading particularly. However, the initiatives would not have been visible at the time of the testing in 2005 and need to be monitored in the future. PIRLS 2006 now serves as a critical external baseline of reading literacy for grades 4 and 5. With the advent of all the initiatives listed, by 2010 one might expect to see the impact of these on teaching and learning and perhaps even an effect on the reading achievement itself.

The poor PIRLS results released at the end of November 2007 resulted in a national outcry reaching the front pages of every daily newspaper in the country. This placed pressure on the national department of education. In Parliament questions were raised about this poor performance as mentioned earlier and the Minister of Education was interrogated and in response provided a list of initiatives launched (Howie and Venter, in press).

PIRLS 2006 also revealed at the end of 2007 that more than half of the primary schools tested in 2005 had no school libraries or classroom libraries and revealed that more than 50% of children had no access to books at home. In the first quarter of 2008,

Minister Pallo Jordan announced that the public library budget would be doubled. The extent to which the PIRLS 2006 had contributed to this directly is unknown. However, in 2008, as one of the initiatives that the Minister of Education had announced in Parliament, the Ithuba Writing project distributed 2.3 million books in all 11 languages to the schools.

To what extent can large-scale testing in South Africa be considered as a friend or foe and how can South African benefit most from large-scale testing?

In addressing the question above, it is clear that apart from the final year examinations, South Africa has not yet, taken the path of high stakes testing of one compares South Africa to the USA and the UK and the large-scale assessment may still be more of a “friend” than “foe”. There are signs in two provinces of a possible link of the systemic evaluations to low outcomes- based educational accountability (Ryan and Feller, 2009) of the schools and teachers. However, should this link be made, we could expect similar behaviour to that reported in the American press (Nichols and Berliner, 2008) about the No Child Left Behind and the associated high stakes testing programme. To some extent some indications of a characteristic of high stakes testing is manifesting around the National Grade 12 school exit exams. Already at the turn of the century, reports about substantial rises in the senior certificate pass rates between 2000 and 2003 (Umalusi, 2004) were attributed to the introduction of Standards Based accountability interventions (Taylor, 2009). Forceful measures were taken against underperforming schools in the final year examinations. This accountability resulted in

a rapid rise in the examination results achieved by manipulating the results by “eliminating high-risk candidates, encouraging candidates to register at a lower examination level of standard grade, lowering the standard of examination questions and raising raw scores during the moderation process” (Taylor, 2009, p.341). The country also has a history of corruption in certain provinces where everything from leaking the examination papers, bribing data capturers, corrupt teachers selling examination papers to pupils and markers changing results have been found for more than a decade (Lubisi and Murphy, 2002; Umalusi, 2009). Furthermore, these behaviours reported above mimic the behaviour reported by Nichols and Berliner (2008) regarding the antics that schools, teachers and education authorities resort to when serious consequences are attached to testing programmes.

On the basis of the international experience and the observed local behaviour, high stakes testing should be avoided for as long as possible in South Africa. Apart from the international cautions surrounding the negative and unintended consequences listed earlier in this article, as a newly integrated and emerging economy and democracy, South Africa is not yet at a mature enough level to increase the testing stakes. Furthermore, if the experience of the USA (Nichols and Berliner, 2005; Nichols and Berliner, 2008) and the UK (Mortimore and Stone, 1990; Black, 1994) is to be correctly interpreted, as a nation South Africa cannot afford to repeat the expensive and damaging mistakes of these two countries. Mortimore and Stone (1990) remind us that the social context of the measurement of achievement (whether through examinations, coursework or assessments), however, has to be taken into account”. They also cited Airasian, who they said amongst others, has described how “if the stakes are high - teachers, as well as pupils, are likely to go to extreme lengths

(including cheating) to secure positive assessments.” Indeed high stakes testing in South Africa would be more likely to lead to possible “narrowing the curriculum (Nichols and Berliner, 2008) or “distortions in the curriculum” (Taylor, 2009); is less likely to take account of the value that schools add (Abernathy, 2007 in Taylor, 2009) and could result in raising aggregate student scores but would lead to huge variability among schools in their response to the standards-based reform initiatives (Elmore, 2003 in Taylor, 2009) and weaker schools simply would not be able to cope as well as stronger schools. Given these international experiences, it must be concluded therefore that high stakes is clearly a potential foe for the South African context.

However, low stakes testing for the sake of monitoring and informing our curriculum and pedagogical practices can be an education system’s friend. This has been extensively argued previously in Howie, 2000, 2002, 2008, 2009, Howie and Plomp, 2006, Plomp, Howie and McGraw, 2003) regarding the benefits as well as acknowledging the limitations. It is often asked why these studies should be done, especially in a country of limited resources when they cost a relatively large amount of money. The usefulness of international and national assessment studies can be categorised in terms of five broad areas of recipients who would make use of this information: (1) data is being demanded by policymakers and decision-makers worldwide at all levels of the schooling system (Kellaghan, 1996; Plomp, 1998; and Postlethwaite, 1999); (2) teachers and schools may learn from what is taught and how it is taught across systems (Howie and Plomp, 2006); (3) researchers are exposed to latest methodological developments and capacity is developed (Beaton *et al.*, 2000; Howie, 2000; Keeves, 1996; Postlethwaite and Ross 2000; Ross, 1994); (4) community are made aware of the outcomes of the system and the quality of education their children

receive and; (5) the consumers of the education system, the employers and institutions involved in further and higher education (Howie, 2001; Howie 2008).

Clear examples of significant impact are available internationally. The results of studies, as well as the more general OECD indicators, have been used in many publications and discussions about the functioning and possible improvement of educational systems (e.g., in Australia, Hungary, Ireland, Japan, New Zealand, U.S.A.) (Howie, 2000; Keeves, 1996; Kellaghan, 1996; Ross et al., 2000). Since policy decisions are not normally documented or published, the direct impact of international comparative achievement studies may not be clearly visible (Kellaghan, 1996). This will especially be so if findings serve an “enlightenment” function in discussions about the quality of education. If this is so, international studies will not “necessarily mean supplying direct answers to questions, but rather in enabling planning and decision-making to be better informed” (Howie, 2001).

## **Conclusion**

In conclusion, whilst South Africa has not yet embarked on high stakes testing in the sense of the USA and UK, a distinction can be made in terms of the impact of the large-scale testing. Currently it would appear that the National Assessment whilst not intended as high stakes assessment at this stage, also does not have a high impact on the system although they have some impact on the policymakers. On the other hand, the international assessments are not considered high stakes tests, but do have a seemingly high impact on the decision makers, the researchers, and a lower impact on

the community and the consumers of education. Finally, it would appear as if it is really on the Grade 12 National Senior Certificate examinations that could be conceived as high stakes testing and these certainly have a high impact on the decision-makers, schools, community and consumers of education. South Africa can learn from the lessons of other contexts of high stakes tests and postpone their implementation until such time as the system has matured sufficiently to utilise them appropriately.

## References

- Abernathy, S. (2007). *No child left behind and the public schools*. Ann Arbor MI: University of Michigan Press.
- Beaton, A.E., Postlethwaite, T.N., Ross, K.N., Spearritt, D., & Wolf, R.M. (2000). *The benefits and limitations of international educational achievement studies*. Paris: International Institute for Educational Planning/International Academy of Education.
- Black, P. J. (1998). *Testing, Friend or Foe?: The Theory and Practice of Assessment and Testing*. Falmer Press: London
- Black, P. (1994). *Performance assessment and accountability: the experience in England and Wales*. *Educational Evaluation and Policy Analysis*. 16(2), pp191-203.
- Black, P, & William, D. (1998). Assessment and classroom learning. *Assessment in Learning*, 5(1), 7-74.
- Bracey, G. (2000). *High Stakes Testing*. Tempe: Educational Policy Studies Laboratory, Arizona State University.
- Clarke, M., Haney, W, & Madaus, G. (2000). *High Stakes Testing and High School Completion*. Boston: National Board on Educational Testing and Public Policy.
- Crouch, L. and Vinjevd, P. (2006). South Africa: Access before quality and what to do now? *Profesorado 10*(1). Retrieved January 30, 2009 from <http://www.ugr.es/recfpro/rev101ART6ing.pdf>.
- Department of Education. (2001). *Assessment Policy for General Education and Training*. Department of Education: Pretoria.
- Department of Education (DoE). (2002a). *Revised National Curriculum Statement Grades R-9*. Pretoria: Department of Education.
- Department of Education (DoE). (2002b). *C2005. Revised National Curriculum*. Pretoria: Department of Education.
- Department of Education (DoE). (2003). *Framework for Systemic Evaluation*. Pretoria: Department of Education.
- Department of Education (DoE). (2005). *The National Protocol on Assessment for Schools in the General and Further Education and Training Band. (Grade R – 12)*. Pretoria: Department of Education.
- Department of Education. (2010). *Evidence-Based report: National Senior Certificate (NSC) intervention improvement strategies*.
- Elmore, R. (2003). *Accountability and capacity*. In M.Carnoy, R.Elmore & L.Siskin (Eds). *The new accountability: High Schools and high stakes testing* (pp.195-209). New York: RoutledgeFalmer.

- Fleisch, B. (2008). *Primary Education in Crisis*. Cape Town: Juta.
- Government of South Africa. (2008). National Assembly International Question Paper 9/5/2008.
- Government Gazette No.: 23406, Vol. 443. Pretoria: Department of Education.
- Greaney, V., and Kellaghan, T. (2008). *Assessing National Achievement Levels in Education*. National Assessments of Educational Achievement series, Vol. 1. Washington. D.C: World Bank
- Howie, S.J. (1997). *Mathematics and science performance in the middle school years in South Africa: A summary report on the performance of the South African students in the Third International Mathematics and Science Study*. Pretoria: HSRC.
- Howie, S.J. & Hughes, C.A. (1998). *Mathematics and science literacy of final-year school students in South Africa: Third International Mathematics and Science Study*. Pretoria: HSRC.
- Howie, S.J. "TIMSS-R in South Africa, a developing country perspective". Paper presented at *American Education Research Association Annual Meeting*, New Orleans, April, 2000.
- Howie, S.J. (2001). *Mathematics and Science Performance in Grade 8 in South Africa 1998/1999: TIMSS-R 1999 South Africa*. Pretoria: HSRC report., p.112)
- Howie, S.J. (2002). *English language and other factors influencing grade 8 pupils achievement in mathematics*. Enschede: University of Twente Press.
- Howie, S.J. (2002). *English language and other factors influencing grade 8 pupils achievement in mathematics*. Enschede: University of Twente Press.
- Howie, S.J., (2007). *Reading Literacy in South African schools*. PIRLS 2006 Encyclopedia: A guide to reading education in the forty PIRLS 2006 countries. Boston.
- Howie, S.J. *Measuring the health of the education system: lessons from international studies for South Africa*. Keynote address, presented at Association for the Study of Evaluation and Assessment national conference, Pretoria, July 2008.
- Howie, S., & Plomp, T. (2006). International comparative studies of education and large scale change. In N. Bascia, A. Cummings, A. Datnow, K. Leithwood, & D. Livingstone (Eds.), *International Handbook of Educational Policy*. Dordrecht, Netherlands: Kluwer Press.
- Howie, S., Venter, E., van Staden, S., Zimmerman, L., Long, C., Scherman, V., Archer, E. (2007). *PIRLS 2006 Summary report: South African children's reading achievement*. Pretoria: Centre for Evaluation and Assessment, University of Pretoria.
- Howie, S.J and Venter. E. (in press). The Impact of PIRLS in South Africa. K. Schwippert. (Ed). *The impact of PIRLS around the world*. Waxmann Publishers.
- Jones, G.J., Jones, B & Hargrove, T. (2003). The unintended consequences of high stakes testing. Lanham MD: Rowman and Littlefield.
- Keeves, J.K. (1996). *The World of School Learning: Selected Key Finding from 35 Years of IEA Research*. Amsterdam: IEA.
- Kellaghan, T. (1996). IEA Studies and Educational Policy. *Assessment in Education*, 3(2): 143-160.

- Kellaghan, T., and Greaney, V. (2001). *Using assessment to improve the quality of education*. Fundamentals of Educational Planning. Paris: UNESCO.
- Kohn, A. (2000). *The case against standardised testing: raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.
- Linn, R. (2008). Educational Accountability systems. In K.E. Ryan & L.A. Shephard (Eds). *The future of test-based accountability*. Pp/3-24. Mahwah NJ: Lawrence Erlbaum Associates.
- Linn, R. (2000). Assessments and accountability. *Educational Researcher*, 29, 4-16.
- Jansen, J.D. (1997). Why education policies will fail. *Indicator SA*, 12, 56-58.
- Jansen, J. (2008). *The Times newspaper*. 18 December 2008
- Lubisi R.C. & Murphy R.J.L. (2002). Assessment in South African Schools. *Assessment in Education*, 9(2), 255-268.
- Macbeath, J. (2002). Scotland: Schools speaking for themselves. In D. Nevo (Ed). *School-based evaluation: an international perspective*. Vol. 8, pp.243-259. Amsterdam: Elsevier.
- Macmillan, J. (2001). *Essential assessment concepts for teachers and administrators*. Experts in assessment series. Thousand Oaks: Corwin Press.
- McDonnell, L.A. (2008). The politics of educational accountability: Can the clock be turned back? In K.E. Ryan & L.A. Shepard (Eds.), *The future of test-based accountability* (pp.25-46). Mahwah, NJ: Lawrence Erlbaum Associates.
- Moloi, Q.M. and Strauss, J., (2005). *The SACMEQ II Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education*. South Africa Working Report. SACMEQ Educational Policy Research Series.
- Mortimore, P. & Stone, (1990). *British Journal of Educational Studies*, VOL.XXXIX No. 1.
- National Education Policy Act 27 of 1996,
- Nichols, S. & Berliner, D. (2007). *Collateral damage*, MA: Harvard University Press.
- Nichols, S.L. and Berliner, D.C. (2008). *How high stakes testing corrupts America's schools*. Harvard Education Press: Cambridge.
- Orfield, G & Kornhaber, M.L. (2001). *Raising standards or raising barriers? Inequality and high stakes testing in public education*. New York: Century Foundation Press.
- Plomp, T. (1998). The Potential of International Comparative Studies to Monitor the Quality of Education. *Prospects*, XXVIII(1): 45-59.
- Plomp, T., Howie, S.J and McGaw, B. (2003) International Studies of Educational achievements, In Kellaghan, T. & Stufflebeam, D.L. *The International Handbook on Evaluation*. Pergamon Press, pp.951-978.
- Postlethwaite, T.N. (1999). *International Studies of Educational Achievement: Methodological Issues*. Hong Kong: Comparative Education Research Centre, University of Hong Kong.

- Postlethwaite, T.N. and Ross, K. (1992). *Effective Schools in Reading: Implications for Educational Planners*. Hamburg: IEA.
- Phelps, R.P. (2000). Trends in large-scale testing outside the United States. *Educational Measurement Issues and Practices* 19(1), 11-21.
- Reddy, V. (2003). *Mathematics and Science achievement at South African schools in TIMSS 2003*. Pretoria: Human Sciences Research Council.
- Reyneke, Meyer and Nel, 2010, p.277.
- Ross, K. (2000). *Some Background Information on the Southern African Consortium for Monitoring Educational quality*. UNESCO IIEP Website.
- Ross, K. & Postlethwaite, N. (1992). *Effective schools in reading: implications for educational planners: an exploratory study*. The Hague: International Association for the Evaluation of Educational Achievement.
- Ross, K.N. and Genevois, I.J. (2006). *Cross-national studies of the quality of education: planning their design and managing their impact*. Paris: UNESCO.
- Ryan, J. (2004). The perverse incentives of the No Child Left Behind Act. *New York University Law Review*, 79, 932-89.
- Ryan, K.N. (2005). Making educational accountability more democratic. *American Journal of Evaluation*, 26(4), 443-460.
- Ryan, K.E. and Feller, I. Evaluation, accountability, and performance measurement in national education systems. In Ryan, K.E. and Cousins, J.B.(Eds.) (2009). *The Sage International Handbook of Educational Evaluation*. Los Angeles: Sage.
- Reyneke, M., Meyer, L., and Nel, C. (2009). School-based assessment: the leash needed to keep the poetic 'unruly pack of hounds' effectively in the hunt for learning outcomes. *South African Journal of Education*. Vol 30:277-292
- Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal Educational Change*. 10. pp.211-227.
- Tamassia, C. V. and Adams, R.J. International assessments and indicators. In Ryan, K.E. and Cousins, J.B.(Eds.) (2009). *The Sage International Handbook of Educational Evaluation*. Los Angeles: Sage.
- Taylor, N. (2009). Standards-based accountability in South Africa. *School effectiveness and School Improvement*. Vol. 20, No.3, September, 2009, p.341-356.
- Taylor, N. (2007) Equity, Efficiency and Development of South African Schools. In Townsend, T. (Ed). *International Handbook of School Effectiveness and Improvement*. Dordrecht: Springer.
- Torrance, H. Pursuing the wrong indicators. In In Ryan, K.E. and Cousins, J.B.(Eds.) (2009). *The Sage International Handbook of Educational Evaluation*. Los Angeles: Sage.
- Umalusi, 2006. Report on the Quality Assurance of the Senior Certificate Examination. Pretoria, South Africa: Umalusi.

- Umalusi, 2004. *Have the standards of the Senior Certificate examinations declined?* Summary report based on the evaluation for the senior certificate examination. Pretoria, South Africa. Umalusi.
- UNESCO (2007). *Global Monitoring Report*. Paris: UNESCO.
- Van Der Berg, S. and Louw, M. (2007). *Lessons learnt from SACMEQ II: South Africa student performance in regional context*. Stellenbosch Economic Working Papers: 16/07.
- World Bank. (2008). *World Development Indicators*. Washington: Word Bank.
- Young, M. (2000). *An International Perspective on Educational Reform*. Paper presented at the HSRC Round Table, 24 October, Pretoria.

Table 1. Differences between classroom, system assessments and public examinations

<b>Type</b>	<b>Classroom assessment</b>	<b>System assessments</b>	<b>Public examinations</b>
<b>Purpose</b>	Multiple, primarily diagnostic and focused on teacher decision making	To provide feedback to policymakers	To certify and select students
<b>Frequency</b>	continuous	Country dependent but for individual subjects offered on regular basis (annually-every four years)	Annually and more often where the system allows for repeats
<b>Who is tested?</b>	Individual students	Usually a sample of students at particular grade or age level	All students who wish to take this examination at the examination grade level
<b>Coverage</b>	Tailored to individual classes	Generally confined to one or two subjects	Covers main subject areas

Based on Macmillan, 2001, p.7; Greaney and Kellaghan, 2008, p. 18

Table 2: International development indicators from the Global Monitoring Report  
2007

	<b>SA</b>
Population	47 million
GNP per capita 2004 in U\$	3 690
Societal Inequality (Gini Index)*	57.8
Total public expenditure on education as % of GNP (%)	5.5
Adult literacy 2004 (%)	82
Primary Education population	7 470 000
Nett Enrolment (%) primary	89

Source: UNESCO, Global Monitoring Report 2007, p.225-320

\*Source: World Bank (2008) World Development Indicators; pp.68-71

Note: Gini Index is a summary of the degree of inequality in the society. Data for South Africa collected in 2000.