

To What Extent Is Testing a Barometer for Educational Quality? Unveiling the Intended and Unintended Consequences of High-Stakes Testing on Student Learning

S. Machingambi¹, C. Maphosa², A. Ndofirepi³, E. Mutekwe⁴ and N. Wadesango⁵

¹Walter Sisulu University, Republic of South Africa

²University of Venda, South Africa

³University of the Witwatersrand, South Africa

⁴University of Johannesburg, South Africa,

⁵Walter Sisulu University, Republic of South Africa

KEYWORDS Assessment. Achievement. Measurement. Testing. Educational Quality. Learning

ABSTRACT This conceptual paper explores a centrally important element in the teaching and learning process, namely, assessment. The paper argues that assessment is the mainstay of any quality learning process. The discourse specifically focuses on high-stakes testing as an assessment instrument that has gained wide currency the world over. The authors use a critical analysis to stimulate an intellectual conversation on the impact of high-stakes testing on student learning. The discourse unfolds with an exposition of conceptual terms that constitute the building blocks of the discussion and then highlights the major intended and unintended consequences associated with the use of high-stakes tests. The authors suggest that educational decisions that affect students learning in profound ways should not be based on the narrow results of high-stakes testing perse as these may not be representative enough. The use of eclectic assessment strategies that afford students multiple opportunities to demonstrate their understanding is therefore recommended.

INTRODUCTION

After being viewed as an industrial problem for most of the past decades (Ross 1995), the issue of quality has now become a service issue as well. Schools and educational institutions have not been spared by this imperative as teachers and schools are now more than ever expected to work extra hard so as to continuously improve the quality of their services (Aamodt 1999). Corby and White (1999) add that the accountability imperative places particular pressure on schools to be more accountable to their clients by publicly demonstrating high standards of performance. This, in essence means that schools have to demonstrate that they are able to deliver what is required of them. In responding to this call, schools have to continuously reflect critically on questions that affect teaching and learning.

The need for quality in educational standards places student assessment at the core of the teaching and learning process. This is so be-

cause through assessment of student learning, teachers and schools are able to evaluate the extent to which educational goals and standards of learning have been met (EdutopiQ 2012). Measuring how well students learn is a critical component of teaching that necessitates the design of educationally sound testing instruments administered to students at regular intervals. Assessment of students is particularly important as it affects decisions about grades, placement, advancement, instructional needs, curriculum and funding (Nichols et al. 2005). Assessment of student learning is considered by Engel (2007) as being important for three important reasons namely improvement, accountability and accreditation. He goes on to add that the results from an assessment can act as an impetus to faculty and staff to engage in self-reflection on learning outcomes and service delivery. For instance, assessment results may challenge teachers and educational institutions to reflect critically on the following crucial questions:

- Are teachers really teaching what they purport to be teaching?
- To what extent are students learning what they are supposed to be learning?
- How can schools explore new ways of promoting better learning?

Address for correspondence:

Dr Severino Machingambi,
Centre for Learning and Teaching Development,
Walter Sisulu University, PO Box 19712,
Tecoma, East London, RSA
E-mail: smachingambi@wsu.ac.za,
machingambi@yahoo.com

Tests, along with other relevant assessment methods, have generally become the most acceptable means of measuring students' knowledge, skills, abilities and competencies (American Psychological Association 1999). To many educational practitioners, testing is now widely viewed as the linchpin to quality teaching and learning and yet its outcomes on student learning is rarely interrogated critically.

This study therefore intends to explore the issue of testing so as to stimulate an educational conversation on the use of testing as barometer for measuring educational quality. The discourse will focus specifically on one type of testing called high-stakes-test. First, a review of the concepts of assessment and testing in general will be given after which a conceptualisation of high-stakes-testing will ensue. The discourse will unfold by bringing the major arguments for and against the use of high-stakes testing in teaching and learning in the spotlight. In doing this, the authors will not try to prescribe definitive answers or solutions but to provoke debate and stimulate further research around the use of high-stakes-testing.

Objectives of the Study

The major objective of this study is to stimulate an intellectual conversation regarding the utility of high stakes-testing in the teaching and learning situation. The study intends to specifically explore the perceived merits and demerits associated with the use of high stakes testing in the learning of the students. The study thus acts as an advocacy for the use of multiple assessment strategies in the teaching and learning process so as to afford students multiple opportunities to demonstrate their understanding and competencies.

DISTINCTION BETWEEN ASSESSMENT AND TESTING

Assessment and testing are critical processes in the teaching and learning trajectory because they enable the teachers to reflect on their own performances as well as knowing how well their students are progressing in their learning. In an educational context, assessment and testing are used to measure how much of the assigned materials students are mastering, and the extent to which the stated goals and objectives have

been met. Hersh (2004) contends that schools have an educational, professional and ethical duty to assess the impact of their activities on student learning. He goes on to argue that this is one way in which schools can defend themselves educationally as learning institutions and properly defend curricular and pedagogical practices and changes.

While the two terms assessment and testing generally denote the process of figuring out how much students know about a given topic or concept, each term has a slightly different focus. In order that educational practitioners get the best from each term, it is important that this paper attempts to explain what each term entails.

Assessment is conceptualised by Pennsylvania State University, (2004) as the process of documenting knowledge, skills, attitudes and beliefs in measurable terms. In an educational context, assessment can be perceived as the process of describing, collecting, recording, scoring and interpreting information about learning (WestEd 2000). Assessment plays a key role in building the student's confidence in his/her ability to learn as well as in developing lifelong skills (Pennsylvania State University 2004) Karter (2012) makes a cogent observation that unlike tests which are normally done at the end of a unit, semester or year, assessments can be done at any time. For instance, teachers can conduct assessments after a lesson, after teaching a particular skill or at the same time tests are done.

Eichstad (2002) refers to a test as a 'thing' or 'product' that measures a particular behaviour or set of objectives. He goes on to quote the American Psychological Association (1999) that defines a test as "an evaluative device or procedure in which a sample of an examinee's behaviour in a specified domain is obtained and subsequently evaluated using a standardised process. What comes to the fore from this definition is that taking a test involves gathering information in a standard way, under the pretext that the information collected is most accurate. In the same vein, Karter (2012) conceives a test as a single examination or set of questions to which the student is asked to respond or give answers. Going by this definition, administering a test can therefore be likened to a snapshot of an individual's or group's performance at a particular time (Eichstad 2002).

Wootton (2002) adds that a test usually inflicts stressful experiences on the test taker thereby predisposing many individuals into disliking being tested and start to associate the word “test” with a somewhat negative connotation.

Another critical difference between a test and an assessment lies in the formats that they can take. Karter (2012) observes that a test tend to follow a narrow format in which questions are asked and students provide answers. Their form could range from being multiple choice questions, essay questions, fill-in items, true and false and matching items. In contrary, assessments can have varied formats such as graded assignments, presentations, class projects, or questions that need answers from students (just like tests). An assessment can be in the form of a teacher talking to students about what they know, or a teacher’s observation of a student working or talking about a subject (Karter 2012).

Pennsylvania State University (2004) argues that another major difference between tests and assessments lies in the results of the two. He postulates that tests tend to show students’ ability to memorise facts and figures without much evidence of clear understanding of those facts and figures. However, assessments done in and outside class can actually tap into the student understanding of the facts and figures or particular processes.

Eichstadt (2002) regards an assessment as a more general process of gathering data to evaluate an examinee. Unlike a test which is based on a snapshot of an individual’s or group’s performance, in assessment information from a variety of sources such as tests, interviews, class projects, presentations and other measures is put together so as to make an informed judgement about a student. The authors therefore agree with Wootton’s (2002) conceptualisation of assessment as a systematic method of obtaining information from tests and other sources so as to draw inferences about characteristics of people, objects or programmes.

From the above it is clear that an assessment process is much broader than a testing exercise. It is important to point out that while a single test is a means towards a broader assessment process, it cannot, on its own constitute an assessment process while the later can also not be reduced to a single test. Thus, teachers can assess using tests as assessment instruments in conjunction with other assessment measures as

indicated earlier on in order to make overall intelligible judgments about a student or group of students.

HIGH STAKES TESTS DEFINED

A test is a high-stakes test if the results have perceived or real consequences for staff, students or schools (Togut 2004). These consequences can be intentional or unintentional. At the same time Edutopia (2012) refers to high-stakes testing as a practice in which the outcome on a standardised test is used as a determining factor in decisions concerning students (Kelly 2006). Typical examples of consequences for students may range from being held back in a grade or not being allowed to graduate until the test is passed (Brower and Beach, 2005). High-stakes tests involve rewarding or sanctioning schools, educators and students on the basis of test results and they enjoy wide use in the United States and Europe (Wootton 2002).

In California, students have to pass the Californian High School Exit Examination in order to graduate (Drummond 1996). In Texas, students from grades four to eleven take various forms of the Texas Assessment of Knowledge and Skills and students must pass the eleventh test to graduate (Kelly 2006). Greene (2003) refers to high stakes tests as standardised tests that are used to reward or sanction schools for their academic performance. The same author goes on to argue that the idea behind such tests is that rewarding or sanctioning schools for their performance provides schools with incentives necessary to improve academic achievement. Smith (2003) cites standardised tests administered under the No Child Left Behind (NCLB) programme, high school exit exams, and the use of test scores to determine whether or not a school will retain accreditation as common examples of high-stakes testing used in the United States.

The logic behind the use of standardised achievement tests in the American context is that they are believed to provide a merit-based method for comparison among students (Wootton 2002). Such tests were initially introduced in the American education system to help eliminate the predominance of individuals who were rewarded because of their personal connections or their wealth (Kelly 2006).

Results from high stakes tests provide guidance in making educational decisions about instruction, placement and promotional opportunities for students (Togut 2004). The critical issue is that failing a high-stakes test can have huge impact on students and schools (Eichstadt 2002).

MERITS OF HIGH STAKES TESTS

According to Education Commission of the States, ECS (2012), using high stakes tests is one way of determining how public schools are performing. Thus, policy makers perceive high stakes tests as a way to raise standards and achievement and hold students, teachers and schools accountable for teaching and learning. The Commission goes on to cite the following as major benefits of high-stakes testing:

- It assures that standards are taken seriously, motivates teaching of the standards and holds schools accountable for student performance
- Motivates students to learn and increases emphasis on student achievement
- Provides the same high expectations and the same basis of evaluation for all students
- Provides information that can inform policy makers on the quality of education
- Monitors school improvement efforts
- Allows recognition to schools and teachers of students who perform high and/or improve performance.

Proponents of high-stakes testing believe that its main objective is to hold schools, teachers and students accountable.

A research carried out in Florida showed that if properly designed, a high-stakes test can act as an impetus for schools to improve not just test scores but also real learning. As Togut (2004) asserts, some critics believe that high stakes testing will improve curricula as schools, teachers, and students attempt to meet the standards imposed by such testing. A study carried out in Texas by Heubert and Hauser (1999) showed that teachers believed that tests had helped them to set clear instructional goals for their students. As the ECS (2012) puts it, high stakes tests are a mechanism of ensuring that standards are taken seriously as schools are held accountable for student performance. This way, high stake tests serve as a mechanism with which to monitor school improvement efforts.

As argued by American Psychological Association (1999) when used properly, tests are among the most sound and objective ways to measure student performance. The Association goes on to point out that test results give classroom teachers important information on how well individual students are learning and provide feedback to the teachers themselves on their teaching methods and curriculum materials. WestEd (2000) reports on the successful use of high-stakes test results in Californian schools to determine whether all students are mastering key content knowledge and attaining the skills necessary for future success in education in education and the workforce.

Lauer (1996) argues that schools as part of communities must meet political demands of education to be more accountable and publicly demonstrate high standards. Accountability means informing parents and the public about how well a school is educating its students and about the quality of the social and learning environment. The essence of accountability entails providing evidence of student achievement to accreditation groups, state legislators and other stakeholders in education. In other words the onus rests on the schools to provide convincing evidence of the continuous improvement of student learning to all stakeholders. In this way, high-stakes testing programmes are therefore taken as an important mechanism for quality control.

Perceived Demerits of High Stakes Tests

Whilst high-stakes assessment has gained popular currency in many educational institutions the world over, it is not without its share of limitations and challenges. While this paper will not attempt to isolate all the challenges that are associated with high-stakes testing, it will examine the consequences of such testing on students, teachers and the curriculum.

According to ECS (2012), high-stakes testing places too much emphasis on a single test score which in most cases might not be an accurate representation of a student's achievement. The Commission further points out that State-wide assessments tend to lead to a narrower curriculum. Critics also raise the concern that high-stakes tests result in too much time spent on preparing students to take the tests without necessarily contributing to the improvement of instruction (Drummond 1996).

Fair Test (2011) argues against the use of high-stakes tests positing that they cause a great deal of harm to individual students and the education system in general. For instance high-stakes tests have been shown to lead to increased grade retention and dropping out. Grade retention can be counterproductive in the sense that students who are retained tend to be affected emotionally and this may dampen their self-esteem and interest in school thereby predisposing them into dropping out of school. Heubert and Hauser (1999) postulate that a number of students tend to be affected by test anxiety, a condition that is likely to lead to low test results. This way, high-stakes tests can be regarded as unfair to students as they do not promote broad and equitable access to educational opportunity and advancement to all students (American Psychological Association 1999).

Reynolds and Trehan (2000) observe that high-stakes tests are usually administered in the form of an exam, the results of which may be used to influence decisions relating to retention, tracking and graduation. Hersh (2004) raises the concern that basing decisions on the results of a single test, that does not include other relevant and valid information will not likely give a true reflection of a student's performance.

Greene (2003) makes a cogent observation that high-stakes testing often leads to the narrowing of curriculum and instruction as teachers will restrict teaching to test specifications. In other words teaching will revolve on that which can be tested with paper-and-pencil tests and anything that is not testable will either be ignored or eliminated from the curriculum irrespective of how important it may be. Further, as succinctly argued by Engel (2007), high-stakes tests restrict learning to memorisation of facts, simple solutions and vacuous thinking. In this way, high stakes tests have been blamed for reducing pedagogy to test coaching without adequately equipping students with skills vital for further learning (Smith 2003). The same author adds that in an environment where teaching is driven by tests, the role of teachers is reduced to that of test coaches thereby narrowing the opportunities that students should be exposed to. This has the ultimate effect that instead of teaching skills that are generally useful in life, teachers may concentrate on teaching the particular skills that relate to a particular test. This may lead to students scoring high in the test but without real mastery of the subject matter. This

therefore raises serious concerns about the validity and credibility of high stake tests.

Hersh (2004) adds that the tendency to reduce teaching to what can be easily measured through testing can lead to inappropriate coaching and even cheating on the part of teachers, schools and students. This way, high stake tests tend to distort information about the general level at which students are performing (Greene 2003). He further postulates that what is truly worth learning is not necessarily measurable through marks or test scores. Greene (2003) argues that high-stakes tests are distorting since the results may not necessarily accurately reflect student achievement and proficiency. Thus, the idea of teaching to the test conflicts with the notion of deep learning where students should be engaged in sustained, deep, meaningful and high quality learning activities which will help them acquire skills for lifelong learning (Fair Test 2011). Smith (2003) succinctly points out that there is no evidence that standardised high-stakes tests are able to measure critical thinking, problem solving, creativity and other more abstract aspects of intelligence. As WestEd (2000) notes, the nature of most high-stakes tests makes it difficult for them to address important curriculum goals that require generative thinking, sustained effort and effective collaboration. Such problem-solving skills demand teamwork and can easily be assessed through portfolios, oral presentations and projects.

Fair Test (2011) points out that good and creative teachers are often disgusted by an environment that reduces all teaching to test preparation. This often leads to demotivation with the ultimate effect that many teachers may leave the school or profession (Fair Test 2011). It is therefore apt to argue that high-stakes tests tend to have punitive effects on students and teachers. Since high stakes tests rarely encourage active and meaningful student engagement, they tend to drive students and teachers away from real learning.

CONCLUSION

Measuring how well students learn is an ideal that schools the world over have to uphold in terms of their professional, moral and accountability mandate. This among other things implies that schools need to design and implement valid and appropriate assessment strategies to measure and improve student performance.

Measuring and reporting on the performance of students and schools has become imperative particularly in this global epoch where schools are being pressurized to demonstrate that they are able to deliver on their obligations. Tests, when used properly, are generally regarded an effective, sound and objective way of measuring the performance of students and schools.

While the use of tests to evaluate performance has gained popular currency, the world over, this paper has argued that we should not lose focus of the unintended and potentially negative consequences of such testing on individual students, schools and educational systems. High-stakes tests, by their very nature, are particularly susceptible to such unintended and negative consequences on students given the high stakes decisions that are made on the basis of high-stakes scores obtained by students. There is therefore need that critical outcomes that affect students' continued education, promotion, graduation or retention be informed by the results of a broad and eclectic assessment strategy that include tests, assignments, practical observations and oral presentations.

RECOMMENDATIONS

Arising from the findings of this study, it is recommended that high-stakes tests should always be used in combination with a variety of assessment instruments so that their negative effect on student learning is counterbalanced. The use of a variety of assessment methods is critical as it affords students multiple opportunities to demonstrate their understanding and competencies in a friendly and conducive environment.

Schools should gradually shift from their cherished and traditional way of using tests as assessment instruments and pay more attention to continuous assessment through non-threatening formative learning tasks.

Educational practitioners and schools should adopt a paradigm shift with regard to tests and their utility in the teaching and learning process. Instead of merely using test results to grade students, such results can be used to enhance current instructional practices.

REFERENCES

- Aamodt A 1999. *Applied Industrial/Organisational Psychology*. Brooks/Cole: Wadsworth.
- American Psychological Association 1999. *Standards for Educational and Psychological Testing*. Washington, D.C: American Educational Research Association.
- Brower RE, Beach BV 2005. *Transformation Leadership and Decision-making in Schools*. Thousand Oaks, CA: Corwin Press.
- Corby S, White G 1999. *Employee Relations in the Public Services: Themes and Issues*. London: Routledge.
- Drummond RJ 1996. *Appraisal Procedures for Counsellors and Helping Professionals*. Englewood Cliffs: New Jersey.
- Education Commission of the States 2012. Assessment High Stakes/Competency. From <<http://www.ecs.org/html/issued=12 & sub/issue ID 33>> (Retrieved September 25, 2011).
- Edutopia 2012. Why Is Assessment Important. From <<http://www.edutopia.org/assessment-guide-importance>> (Retrieved January 17, 2012).
- Eichstadt T 2002 Test Don't Diagnose... You Do: The Difference Between Testing and Assessment. From <<http://www.speechandlanguage.com/clinical-cafe/tests-don-t-disagnose-you-do-the-difference-between-testing-and-assessment>> (Retrieved 25 March, 2012).
- Engel A 2007. Exposing the Myths of High Stakes Testing. Fair Test. From <<http://www.fairtest.org/exposing-myths-high-stakes-testing>> (Retrieved April 03, 2012).
- FairTest 2011. The Dangerous Consequences of High-Stakes Standardised Testing. From <<http://www.fairtest.org/facts/dangerous%20Consequences.html>> (Retrieved June 25, 2012).
- Greene JP 2003. Testing High Stakes Tests: Can We Believe the Results of Accountability Tests? From <<http://www.manhattan-institute.org/html/ccr-33.htm>> (Retrieved January 25, 2012).
- Hersh R H 2004. *Assessment and Accountability: Unveiling the Student Learning Factor in Assessment*. Sage: London.
- Heubert JP, Hauser RM (Eds.) 1999. *High Stakes: Testing for Tracking, Promotion and Graduation*. Washington, DC: National Academy Press.
- Karter T 2012. Difference between Assessment and Testing. From <http://www.who.com/infor_8090151_differences_between_testing.html> (Retrieved June 25, 2011).
- Kelly M 2006. High Stakes Testing. From <<http://www.netplaces.com/new-teacher/high-stakes-testing/what-is-high-stakes-testing.htm>> (Retrieved January 25, 2012).
- Lauer HR 1996. *Social Problems and the Quality of Life*. Dubuque: Brown Publishers.
- Nichols SL, Glass GV, Berliner DC 2005. High Stakes Testing and Student Achievement: Problems for the No Child Left Behind Act. Arizona State University. From <<http://nepc.colorado.edu/publication/high-stakes-testing-and-student-achievement>> (Retrieved December 18, 2011).
- Pennsylvania State Unibversity 2004. Testing and Assessment. From <<http://istudy.psu.edu/First Year Modules/Testing Assessment/Materials.html>> (Retrieved November 29, 2011).
- Reynolds M, Trehan K 2000. Assessment: A critical perspective. *Studies in Higher Education*, 25(3): 267-278.
- Ross JE 1995. *Total Quality Management: Text, Cases, and Readings*. Florida: St Lucie Press.
- Smith SE 2003. What is High Stakes Testing. From <<http://www.wisegreek.com/what-is-high-stakes-testing.htm>> (Retrieved March 12, 2012).
- Togut TD 2004. High Stakes Testing: Educational Barometer for Success, or False Prognosticator for Failure. From <<http://www.harborhouselaw.com/articles/highstakes.togut.html>> (Retrieved 10 August, 2011).
- West Ed 2000. High-stakes Testing. From <<http://www.wested.org/online-pubs/po-00-01>> (Retrieved 16 June, 2012).
- Wootton S 2002. Encouraging learning or measuring failure? *Teaching in Higher Education*, 7(3): 353-357.