

¹REFLECTIONS ON ASSESSMENT AND RESEARCH

Omaze Anthony Afemikhe,
Institute of Education, University of Benin, Benin City, Nigeria
e-mail: tonyafemikhe@yahoo.co.uk, omaze.afemikhe@uniben.edu

Abstract

Research and assessment are two terms which are related but different as they ask different types of questions. This paper presents their similarities and the differences between them. In addition it examines how assessment results are used in research. As part of our reflection, issues considered included development and validation of assessment instruments, assessment irregularities, school-based assessment. It also looked at high stake assessment and large scale assessment, paper and pencil versus e-assessment and security issues and cybercrime in assessment. The intent is to map out current issues that are of concern in assessment and research. Towards this end the paper also explored teaching/school effectiveness in terms of teaching practice, value added and created, monitoring and improvement of learning, student comprehension and skills development and accountability related issues.

Introduction

Should we call our efforts to learn about programs and outcomes “research” or “assessment”? The name can make all the difference. And trying to adhere to the standards of research may get in the way of doing effective assessment (Upcraft&Schuh, 2002: 16).

Research and assessment are two concepts that have a lot in common and very often they are seen as one and the same thing. Research simply could be seen as ‘searching again’. It is a process of finding answers to questions to which up till that point in time there is no solution. Experience shows that research is basically classified as basic and applied. Applied research uses results derived from research to solve real word problems by examining a specific set of circumstances. Basic research on the other hand focuses on fundamental principles and testing theories. There is however a symbiotic relationship between them. That is why it can be stated that basic research feeds applied research and applied research feeds basic research. In actual

¹Lead paper presented at the Educational Assessment and Research Network in Africa (EARNiA) Conference, November 23-27, 2015 at the University of Buea, Buea, South West Region, Cameroon.

fact basic research produces the theory on the basis of which practical problems are solved by applied research and applied research provides data to support theory and guide and revise the development of theory.

Assessment is the collection of information for making decisions. In education the decisions are about the student, school, curriculum and policies. The assessment could be executed by and within the school or done by an agency external to the school. Within the school according to Afemikhe (2015) and reemphasized by Afemikhe and Imobekhai(2015) it

can be seen from the angles of the learners and the authority. Assessment information can be used to choose students into programmes; while assessments can also highlight the strengths and weaknesses of students in the learning process; and yet assessments could also be used to determine those who qualify and therefore worthy of certificates. Hence assessment focuses

on choosing, learning and qualifying from the learners' angle. These assessments include those for improvement of learning and those for certification which have been technically referred to as formative and summative assessments respectively. From the perspective of the authorities, assessment serves an important function of who gets admitted or selected into an educational programme. Schools use assessment information composed of test scores and other affective measures for selecting potential students. Assessment information can equally be used to track the functioning of the components of the educational system which is a monitoring function. In this case the questions asked include how are the students performing, what are the problems hindering optimum performance of the students and the school? The schools belong to the community and it invests in them. As a result the community could be interested in how well the schools are doing the job assigned to them. Thus assessment information is useful in holding accountable those responsible for the different components of the school system if it is to achieve its goals.

The similarity between Assessment and research begins with both of them asking the right questions, collecting data, analyzing results, and using the results to provide evidence and/or make decisions. They both utilize qualitative and quantitative data. According to McGillin (2003) 'good assessment/evaluation can be expanded into good research. Good research should lead to even better assessment procedures. Good assessment makes use of the best conceptual and theoretical models and the best research measures or methods'. Erwin (1991) argues that 'although they share many processes, research and assessment differ in two important respects:

- Assessment guides good practice, whereas research guides theory and tests concepts.
- Assessment typically has implications for a single institution, whereas research typically has broader implications for higher education'.

Assessments use research methods, but they have very different reasons for being conducted. Upcraft and Schuh (2002) have provided a distinction between research and assessment. According to them assessments are undertaken to provide information to aid decision-making; that is they guide practice and that can influence how it is executed. They influence policy. Theory is used to frame research. The assessment study's findings, can lead to the theory being re-conceptualized, affirmed, or perhaps even rejected until another investigation is undertaken. Limitation in resources is a salient difference between research and assessment. Research usually has enough time to develop studies which are tight and fits the available resources or find same unlike assessment that may not have that luxury of time.

The time to influence policy and practice may be limited and so assessments must be modified to fit realistic expectations. Researchers however have time to plan and complete their studies. Assessments are dependent on organizational context which are not static but dynamic. The assessment programme can be altered because of change in leadership occasioning new problem, different evidence and devaluation of old problems. Research is almost independent of the organizational context. When there are flaws with design due to problems with implementation assessment can hardly do something but research can start all over again. Research is apolitical but the political context affects assessment.

Having shown the differences between them and highlighted what they have in common the next part of this paper examines how assessment results are used in research. Thereafter we shall examine:

- i. Development and validation of assessment instruments
- ii. Assessment irregularities
- iii. School-based assessment
- iv. High stake assessment and Large scale assessment
- v. Paper and pencil versus e-assessment

vi. Security issues and cybercrime in assessment

Use of Assessment Results in Research

The theme of this conference has described the relationship between assessment and research as conjugal. According to (Muhwava, Nyirenda, Mutevedzi, Herbst&Hosegood, 2008) “conjugal relationships are defined as “married or regular sexual partners who are members of the same household, regardless of their place of residence”. Thus research and assessment are intimate bed fellows. As indicated earlier research and assessment share a common methodology; the differences have already been alluded to. Assessment results have found great use in research decisions: guidance, administrative and instructional decisions. This categorization has been seen as arbitrary and overlapping. One can ordinarily look at instructional decisions as those which affect what happens within a particular classroom; administrative as those which affect activities within the school building. Guidance can be seen as involving guiding students in their vocational choice and other personal decisions using self reports. All in all, research use cuts across all three areas.

Assessment use in research can be looked at from the angle of the student, teacher and policy maker. At the core of all these uses in quality provision of education. Quality is of the essence in education; it is a ubiquitous concept which has attracted much concern (Afemikhe, 2013). Lot of confusion are attendant in its definition with its being seen as multifaceted (Fraser, 1992), elusive (National Association of Vocational Education (1994) and slippery and value-laden (Harvey & Green, 1993). Afemikhe (2003) saw quality as value for money; this looks at quality of provision, processes, or outcomes, judged against monetary cost and even returns on investment by stakeholders. UNICEF (2000) sees quality in education as encompassing quality learners, quality learning environment, quality content, quality process and quality learner outcomes. The quality of education provided is a function of many variables. The need for emphasis on quality is attendant on the need for organisations to survive in the face of increasing competition and the demand for better services and products (Teeroovengadum, Kamalanabhan&Seebaluck, 2010). Thus education provided must be qualitative; educational quality is important to many stakeholders (children, families, teachers, schools, education system and international agencies). The interests according to Furniss (n.d) are as shown in Table 1.

Table 1

Interest groups and their interests in education *

Interest Groups	Interests
Children	⇒ Am I passing?
Families and communities	⇒ Are they learning?
Teachers	⇒ What are they learning?
Schools	⇒ Are we doing a good job?
Education systems	⇒ Are results consistent with national priorities? ⇒ Is schooling efficient?
International agencies	⇒ How does this country compare with others?

* Adapted from E. Furniss (n.d)

Looked closely, one realizes that important issues include teaching/school effectiveness in terms of teaching practice, value added and created, monitoring and improvement of learning, student comprehension and skills development and accountability related issues. The readiness, interest and learning profile of the students are very important here. The readiness includes skills, concepts and concepts that indicate that the learner is ready to learn. Learning profile focuses on strengths and weaknesses, preferences, and self awareness by the learners. Information on all these is obtainable from assessment. Quality assurance is germane in this regard and many models have been generated in ensuring that

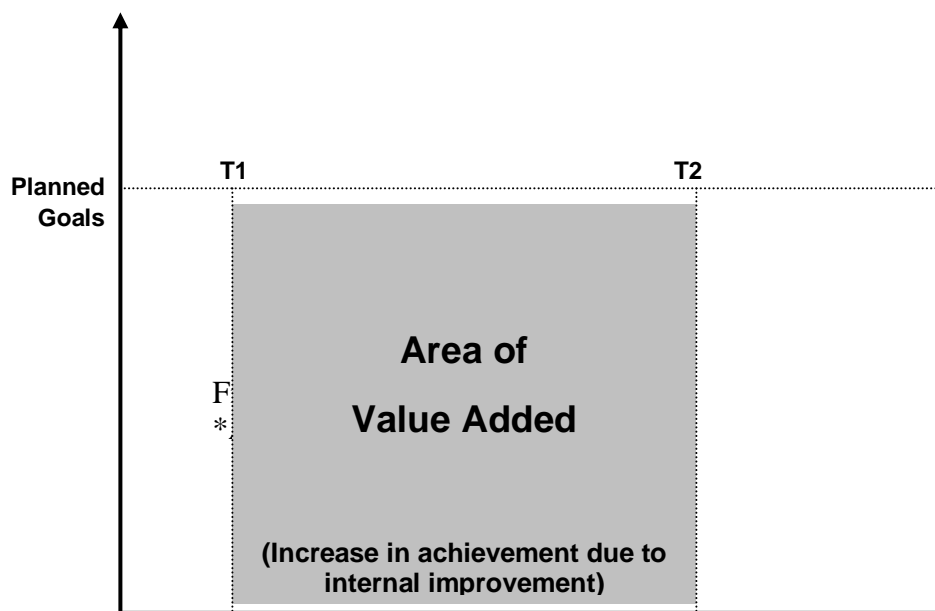
there is quality assurance in education and indeed in any service provision. These models according to the literature include goal and specification, process, absence of problem, satisfaction, legitimacy, organisational and quality management models (Cheng, 2001).

School /teaching effectiveness involves specification of the goals and objectives of any activity; these are however value laden. The outputs are looked at against these to determine the effectiveness of the school or teaching. According to Firestone (1991, p. 2)

effectiveness is not a neutral term and defining the effectiveness of a particular school always requires choices among competing values...criteria of effectiveness will be subject of political debate.

It is a composite of achievement and character formation. The factors involved are content and organizational and process variables. Content and organizational issues normally examined include effective teacher characteristics, supportive school leadership, effective instructional strategies, good home-school-community relationship and positive external relationship with board and its personnel. The related process variables include clear goals, objectives and mission, decentralized decision-making, high student expectation and strong school culture. In measuring effectiveness different approaches have been used. An important approach used is the statistical control method where achievement is regressed on socio-demographic variables of the student. Effectiveness is then defined as difference between observed mean and predicted mean. Unfortunately, it sometimes happens that not all relevant variables are included in the regression equation and some important variables may be excluded. Frameworks which have equally been used include indicators approach that focuses on goals, resources, satisfaction and legitimacy among others.

The Contextual Assessment has equally been used. It posits that learner outcomes and school effectiveness are embedded in national contexts. It identifies three factors of goals, school accountability and strong central control. Lastly, school effectiveness criteria for support include, inter alia, adequate time, financial and human resources as well as a culture of decentralization (Van Damme, Opdenakker & Van Landeghem, 2008). An important variable in school effectiveness measurement is student achievement represented as raw scores. These however do not describe how well a school has performed. How well the school has performed in promoting student achievement is important; this is described by what has now come to be called value-added. It is described as measuring relative progress made by students in a school over a period of time when compared to other students from other schools in sample. This progress involves what school adds which is seen as above what is expected considering the students' background and their prior learning. The value added has been represented by Cheng (2001) as in Figure 1.



Accountability is important in any organization. It involves making good commitment and this is usually looked at from two angles. These include basic standards and institutional mission that must be focused on in programme execution and those in the organizations such as teachers and administrators assuming responsibility for the task they have been assigned. In this regard Shulman (2003: 1) has indicated that:

My point is that excellent teaching, like excellent medical care, is not simply a matter of knowing the latest techniques and technologies. Excellence also entails an ethical and moral commitment-what I might call the “pedagogical imperative.” Teachers with this type of integrity... inquire into the consequences of their work with students.... A professional actively takes responsibility, she does not wait to be held accountable.

Assessment has been used greatly in assessing the teacher and school, enhancing student learning but it’s used by policymakers has been limited (Spillane, 2005). One reason for this which has been alluded to in the introduction to this paper is the time required to execute assessment for results to be available timely. Evidence based approach can be utilized in ensuring that assessment is used by policymakers. Chuderwood, Zapata-Rivera and VanWinkle (2010) have clearly indicated responsibilities that policy makers bear to make decisions. These responsibilities are shown in Figure 2.

The responsibilities include school improvement plans that focus on information about students strengths and weaknesses, goals, and community pressure to mention a few; others are professional development, programme selection and evaluation, curriculum selection, improvement of student achievement, staff allocation and communication. Professional development includes all approaches to improve teachers’ and principals’ effectiveness to raise pupil achievement. Programme selection encompasses efforts to direct funding to the pressing needs while curriculum selection concerns curriculum adoption, approach and efforts to link

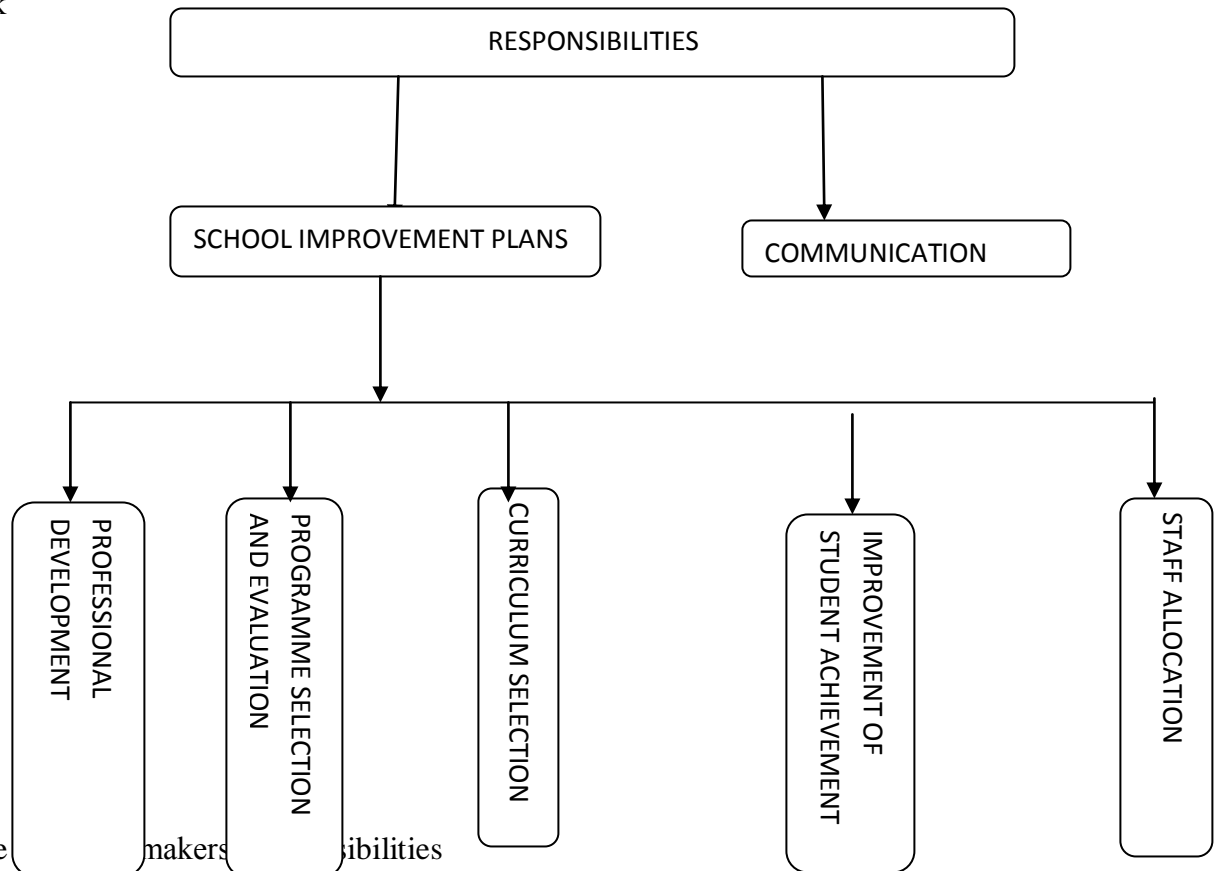


Figure 2: Responsibilities of policymakers

curriculum and instructional standards. Students are the reasons for the establishment of schools and therefore the improvement in their achievement is important. In this regard programme placement, gains in achievement among others, is emphasized. Communication involves sharing information with stakeholders such as students, parents, community and outside agencies.

Development and validation of Assessment Instruments

The implementation of assessment requires instruments; the instruments should yield valid and reliable information. Instruments can be constructed, adopted or adapted. When the instruments are to be adopted or adapted, it is assumed that the instruments have been constructed and this is the development phase of the instruments. It is also important to ensure that one can have confidence in the interpretation and use of the scores obtained from the assessment instrument. The need to develop the assessment instrument can emerge when none exists for the particular purpose or the existing ones are not adequate for one reason or the other.

Planning the assessment entails specifying (a) the construct or domain to be assessed, (b) the population for which the assessment instrument is being constructed, (c) the objectives of the assessment within the requisite framework, and (d) the concrete means through which the behaviour sample would be gathered and scored. Thereafter a preliminary plan of the instrument is put in place. When this has been done the development then involves generating item pool, submitting the item pool to reviews for qualitative analysis, and trying out the items on a representative sample of the population. The results of the trial administration are evaluated, items are added, deleted or modifies as needed and cross validation is carried out.

The classical test theory (CTT) was initially the only method used in the analysis of items in the construction of assessment instruments. It focused on the examinee's total score on the assessment instrument. This was utilized in establishing the criterion-related validity evidence, construct-related validity evidence of the instrument. Factor analysis has also been applied in determining construct-related validity evidence. With respect to affective measures, trait theories, factor-analytic strategy, criterion-keying and a combination of construction strategies can be used in their construction (Aiken, 1997).

The item response theory (IRT) is a more recent theory in the development of assessment instruments. Its emergence is to address some of the shortcomings of the classical test theory. The item characteristics obtained from CTT are group dependent but in the case of IRT they are invariant. Tests are also dependent on the specific items selected for the test in CTT but this is not the case for IRT if the assumptions are met (Urbina, 2004). Indeed both theories can be utilized but IRT is currently gaining more groups in validation of assessment instruments.

School-based assessment

Large scale assessment is seen to be good but takes a disruptive character when it is equally of high-stakes. This is particularly true for public examinations as examinees can go all length to ensure that they pass. The result craze to pass by all means is the preponderance of irregularities in assessment. With an interest to overcome the disruptive effect of high-stakes assessment and a desire to capture a broader range of outcomes beyond those that can be measured by traditional tests, the interest in school-based assessment (SBA) has come to the fore and now greatly emphasized in almost most national education systems. SBA, which is also seen as continuous assessment is defined as assessments administered in schools and evaluated by the teachers, marks from which, in some countries, could count towards the

students' external/public examination results. It is a move towards reducing the over reliance and over dependence on one-shot public examinations for certification. Through it meaningful assessment is ensured. It is a component of many national examination systems such as in Nigeria, Ghana, Botswana, South Africa, etc.

It is characterized by its authenticity and robustness besides being holistic, integrated, low stake and comprising of quality assurance. It is a formative and diagnostic task geared towards improving the quality of teaching, learning and the mode of assessment itself. There cannot be any effective teaching and learning without valid, reliable, well-structured and coordinated assessment and evaluation. Teachers are at the core of SBA as they are involved in the planning, gathering evidence, interpreting evidence and using results for decision making. According to Yusof (2013: 28)

School Assessment emphasises on collecting first-hand information about learners' learning based on curriculum standard, whereby teachers plan their assessment, prepare the instruments, administer the assessment, examine learners' responses and report their progress. Teachers could conduct formative assessment and provide constructive feedback and feed forward to learners. Teachers could also conduct summative assessments. Teachers are encouraged to develop various forms of assessments tasks that are authentic, contextualised and could enhance learners' higher order thinking skills and the 21st century skills such as creative, innovative, problem solving and decision making apart from instill moral values.

With SBA being embedded in the teaching and learning process it is expected to reduce the fear usually associated with public examinations and leading to the production of good quality students. Its characteristics apart from the teacher being an important figure in its implementation include collection of many samples of student performance over a period of time. It is carried out in normal classrooms and can be adapted and modified by the teacher to match the goals of instruction. Feedback is immediately provided by the teacher and hence it stimulates continuous evaluation and consequently leads to adjustment of the teaching learning programme.

In Nigeria, as in most other countries, SBA is executed through the use of a variety of assessment types. The assessment tasks may include tests, essays, examinations, reports, investigations, exhibitions, productions, performances, presentations and demonstrations. The marks from this assessment process are usually submitted as part of entry of candidates for external examinations at completion of senior secondary school. When they are used for this purpose they are expected to be moderated (i.e. adjusted) by various factors including the school group's performance at external examinations. This is to place marks in school-based assessment from different schools on the same scale. The implementation in some cases has been poorly executed. In Nigeria it has been described as a caricature by Obanya (1979) and Afemikhe (1989 &1990, 2000) because of the shoddy nature it has been implemented. Nonetheless it can lead to improvements as earlier on alluded to.

Let me elaborately use information contained in one of my textbooks (Afemikhe, 2014). Planning is of essence in any school-based assessment programme. It must consider the local needs, resources and skills possessed by those who would be entrusted with its implementation. A school testing programme must have a planning committee that will be saddled with planning and directing it. The committee should outline the objectives of the programme, keep a file of tests to be used, plan the budget, make arrangements for construction, adoption, adaption and procurement of the respective tests for the programme. Other activities to be planned for are information dissemination to students and parents, professional development and support for teachers and material and technical conditions.

In any testing programme, the frequency of testing is an important issue that must be addressed. This involves the test to be used, the class at which it should be used, and the time of year when it is to be utilized. Over testing should be avoided; one must be able to justify why each testing has to take place. Whatever test is used should enable students to learn and demonstrate knowledge, skills, attitudes or behaviour that is of interest. The assessment must consider the curriculum and consider how teaching and learning take place. The students must be able to explain the why of the achievement as well as indicate how the results would be used. This involves the students knowing whether the testing is for diagnosis, formative or summative purposes.

The students should be fairly and consistently treated such that they have opportunities to demonstrate abilities; some students should not be given preferential treatment. The same criteria should be used in grading the assessments. All grading should be based on previously defined criteria. For example, multiple-choice questions should earn a mark when correct while essay questions that take into considerations other factors beside content should so indicate. The teachers who teach the same class should understand the criteria for scoring and apply them in the same way. Students should be made aware of how assessments are to be scored.

The school testing programme definitely yields multiplicity of assessment scores. These have to be combined and given a value judgment. This grading should be in accord with set standards. Achievement and efforts should as much as possible be graded separately. The grade should not necessarily be based on all work; it should be based on a representative sample of content and activities performed and related to curricular aims.

One must decide about the information that should be provided students or parents or even to policy makers. Whatever the situation, the information must be timely, clear and useful to the user. It must be provided in a language easily understood by teachers and students. Regular information must be given to parents. Different types of reports are normally used. Some of the assessment forms include those for affective domain, report for a term's work, and transcripts.

One major problem is that of comparability of scores across teachers and schools. While some teachers may be lenient in award of marks others can be very strict. In a situation like this one cannot clearly say whether scores from different sources are really comparable. Can one say for example that a score of 40% given in one school is the same as 40% obtained in another school? Is 55% in one school really a better performance than one of 45% in another school? This is the comparability issue that needs to be addressed. This is particularly so when these scores are to form part of marks in certification examinations. One way to address this problem is the introduction of social moderation. This involves 'use of trained-interviewers who meet formally to assure the quality of assessment instrument and ensure that judgments of standards are comparable from school to school through a process of investigation' (Queensland Studies Authority, 2010:4). The real issue is the extent to which school assessment are in agreement with curricular requirements.

School-based assessment involves a collection of different assessment information. With time these become very massive and therefore there would be a problem of data storage. In a situation where the information is not properly stored it could be lost and this is a bad omen for implementation of school-based assessment. To take care of this problem, technology should be relied upon. The computer easily becomes handy here. With the use of computer comes the problem of availability of teachers who can handle computer processes. This however is not a problem that cannot be surmounted as teacher development programmes can be put in place to address it.

The implementation of school-based assessment requires the use of different types of instruments. It is doubtful whether all teachers are capable in construction and use of these

instruments. To obviate this problem, schools can work as a group to generate the relevant instruments for use in their schools.

High stake and Large Scale Assessment

Large scale assessments are those assessments taken by many examinees over a geographic area. These assessments are usually external to the school and they include public examinations, national assessments and international assessments. The assessments could also be termed high-stakes if they are used to determine the future of individual students, such as promotion to a higher class or are for the award of a certificate. Thus high stakes assessments are designed to hold individual students accountable for their own test performance. While public examinations, national assessments and international assessments are all external to the school and are large scale assessments, it is only the public examinations that are both high-stakes and large-scale. The high stakes assessment is hinged on some theories including motivational theory, theory of alignment, information theory and symbolism theory.

With respect to the motivational theory the external rewards attendant on classification for employment and certification for higher education can motivate teachers to improve performance. Motivational theory is the predominant theory underlying test-based accountability. According to this concept, the extrinsic rewards and sanctions associated with the high-stakes test serve to motivate teachers to improve their performance. With respect to intrinsic motivation researches are inconclusive on effect of reward on it. Cameron and Pierce (1994) have found that reward does not decrease intrinsic motivation, while others have concluded that tangible rewards often undermine internal motivations (Deci et al. 1999).

The theory of alignment posit that when the major components of the educational system such as standards, curriculum, and assessments are aligned by educators that system-wide improvement is most likely to occur. Alignment here involves synchronizing both external and internal components of the school and alignment between the external accountability of schools and schools' sense of internal accountability (Abelmann and Elmore, 2004). According to the information theory students' performance data can be used by teachers and administrators to make decisions about students and programmes; when the same data is given to policy makers and giving them incentives to improve their performance can guide classroom and organizational decision-making.

The symbolism theory is manifested in the notion of "public answerability" — that is, the idea that the public has a right to expect its resources to be used responsibly and that public institution are accountable for caretaking the public trust. Through high-stakes assessments evidence is provided that public education is, in essence, responsible and rigorous and further provide evidence of the system's performance.

The interest in assessment with particular emphasis on tests is as a result of the effects on students, teachers and principals, parents (attitude toward education, engagement with schools and participation in child's learning), policy makers (belief about system performance, judgment about programme performance and allocation of resources (Stecher, 2012). In particular, tests will prompt schools to reform policy, encourage teachers to adopt more effective practices and motivate students to learn. One looks out for changes in policymakers goals of reform. From testing it is also possible to know the extent to which changes in behaviour due to gain scores are indicative of improvement in scores.

All said it may be necessary to examine the positive and negative effects of high-stake learning. Table 2 taken from Stecher (2012) seems to contain most of what needs to be known in this regard.

Koretz, McCaffary and Hamilton (2001) examined teacher responses to high-stakes testing. The results indicated positive, negative and ambiguous responses. The positive

responses include providing more instructional time, working harder to cover more material within a given time and working more effectively. The negative response is cheating and the not clear responses are reallocation of instructional time, aligning instruction with standards and teaching to the test. High-stakes assessment however measure limited sample of behaviour and hence there is usually teaching to the test. Importantly they can be used to promote changes in school practice. This can usually be utilized in measuring value added.

All in all the data from high stakes tests can be seen to be useful to policymakers for assessing school and system-level performance. They are however insufficient for individual-level accountability and provide meager information for instructional guidance. Above all these certain issues becloud high-stakes assessment. Among these are reliability, validity, basing decision on a single score and considerations of non-instructional variables that negatively impact on achievement. As indicated by Michigan Association of School Psychologists (MASP, 2004) in large scale assessments, sampling error

Table 2*

Potential Effects of high-stakes Testing

Positive effects	Negative effects
Effects on students	
Provide students with better information about their own knowledge and skills	Frustrate students and discourage them from trying
Motivate students to work harder in school	Make students more competitive
Send clearer goals to students about what to study	Cause students to devalue grades and school assessments
Help students associate personal effort with reward	
Effects on Teachers	
Support better diagnosis of individual students' needs	Encourage teachers to focus more on specific test content than on curriculum standards
Help teachers identify areas of strengths and weaknesses in the curriculum	Lead teachers to engage in inappropriate test preparation
Help teachers identify content not mastered by students and redirect instruction	Devalue teachers' sense of professional worth
Motivate teachers to work harder and smarter	Entice teachers to cheat when preparing or administering tests
Lead teachers to align instruction with standards	
Encourage teachers to participate in professional development to improve instruction	
Effects on Administrators	
Cause administration to examine school policies related to curriculum and instruction	Lead administrators to enact policies to increase test scores but not necessarily increase learning
Help administrators judge the quality of their programmes	Cause administrators to reallocate resources to tested subjects at the expense of others subjects
Lead administrators to change school policies	Lead administrators to waste resources on test

to improve curriculum or instruction	preparation
Help administrators make better resource allocation decisions e.g., provide professional development	Distract administrators from other school needs and problems
Effects on Policymakers	
Help policymakers to judge the effectiveness of educational policies	Provide misleading information that leads policymakers to suboptimum decisions
Improve policymakers' ability to monitor school system performance	Foster a "blame the victims" spirit among policymakers
Foster better allocation of state educational resources	Encourage a simplistic view of education and its goals

*Source: Stecher (2012: 86)

reflects the reliability of testing a different group of students each year, while measurement error refers to the variation in scores associated with testing students on a particular occasion.

Fluctuating scores on large scale assessments are typically the result of sampling error which could emanate from the low sampling of content. Reliability estimates are generally not provided for various subgroups of the population and hence according to MASP (2004) no data is provided regarding consistency of the assessment for these subgroups.

Even though three main types of validity-related evidence are generally emphasized there are however six forms of validity-related evidence that should be considered in large scale assessment: content, substantive, structural, generalizability, external, and consequential. Structural, generalizability and consequential validity are most crucial. While structural validity refers to the fidelity of the scoring structure, generalizability refers to the ability to generalize results to the population and across populations, and consequential validity involves the social consequences of the assessment to the society. Each of these is of great importance in examining the validity of large scale assessments.

There are many factors that are non-instructional and have negative impact on achievement. The fact that students are not randomly assigned to schools and because some of these factors such as family stability and mobility, parental involvement and expectations regarding school success, parent income level, parent education level, family ethnicity, student motivation, student absenteeism, and student capacity for learning, which have all been shown to influence test performance (Simner, 2000) blame for poor performance cannot be placed solely on the school. We must however give proper consideration to each of the factors.

The final weakness is that of drawing conclusions from a single data point and the method of reporting the data to the public. The use of a single test score to make far ranging decisions and judgments about the status of education is really inappropriate. This is because a single assessment does not provide enough information to make accurate or responsible comparisons.

Paper and Pencil Versus E-assessment

The earliest mode of assessment probably was the oral type and with the introduction of writing, paper and pencil tests (PBT) emerged. Oral assessments are still used but most assessment use the paper and pencil mode. In its utilization, examinees are given an opportunity to perform at their best following usability principle of taking into consideration factors which facilitate administration of the assessment. It takes different forms of essay, multiple-choice, fill-in, matching. With the advent of the computer and other technology related materials assessment are now presented in e-mode. The e-assessment take the form of

computer based test (CBT) or mobile based test (MBT). In the CBT or MBT, the assessment is provided through a device such as a computer or a mobile phone and the examinee is expected to respond through the same medium.

According to Nikouand Economides (2014) CBT has advantages over the PBT. These include the fact that in CBT there is a random choice from a large questions pool; innovative and sophisticated item formats can be used; examinee can obtain immediate feedback; there is automated scoring and advanced security is also possible. According to Chu, Hwang, Tsai & Tseng (2010) mobile learning devices improve learners' attitudes, learning achievements and motivation.

The use of different modes of assessment raises the issue of whether the different modes produce different assessment results. This issue of comparability of test administration modes is one that has been of concern to assessment experts, researchers, practitioners and users (Wang, Jiao, Young, Brooks & Olson, 2008). If differences are observed there is what is called 'test mode effect'. The results from comparability studies have been inconsistent; some have shown performance in CBT to be better than in PBT, some others with PBT being better than CBT and others showed no better performance in any. Students with mobile devices have been found to achieve a better score than CBT students and lower with PBT (Wang, Jiao, Young, Brooks & Olson, 2008). Bauner, Roded&Gafni (2009) found equivalence between Information based test (CBT and MBT) and traditional PBT version of the test used in their study.

There are however some issues with CBT which include interruptions to power supply, non-standardized computers in different locations, internet server problems such as heavy traffic on server, items becoming known to other examinees and handling problems during administration. Despite all these problems, CBT is assuming an improved position due to technological growth. Studies have also shown that mode effect is attributable to incessant use of computers which leads to improved performance in both PBT and CBT and males performing better in both CBT and PBT (Hwang & Chang, 2011).

Assessment Irregularities

An assessment whether conducted at the school level or externally to the school is not in all cases conducted according to prescribed rules and regulations. The extent of the non-compliance is usually most felt in large-scale examinations with high-stakes implications. The irregularities have been seen as examination malpractice. To avoid malpractice examination bodies whether school-based or external to the school must make efforts to put in place measures to ensure that malpractices are minimized. According to Afemikhe (2010) public examination bodies commit large amount of monetary investment in material and financial outlay to ensure that their role as gate-keepers is not compromised. He opined that this is usually 'effected through consciously putting in place control mechanisms at all stages of the certification process from the syllabus development, item generation, production of question papers, distribution of question papers to custodian points and eventually schools and centres where the examinations take place'.

Assessment irregularities include events which cause assessed performances not to be related to actual performance; they are process variables which could be intentional or not that create bias of outcomes (Glidden, 1996). TechTalk (2005) described assessment irregularities as acts which involve non-compliance with stipulated guidelines for the conduct of the examination. Manipulations, falsification, forgery or alterations of documents are also aspects of irregularities (Information Systems Audit and Control Association, 2005:1). Indeed it involves the use of improper means by a candidate in an assessment. This is collaborated by the World Bank Group (2001) when it indicated that malpractice is "a deliberate act of wrongdoing, contrary to official examination rules, and is designed to place a candidate at an unfair advantage or disadvantage".

It is now known that malpractice is most common in large scale examinations with high-stakes. In such examinations success in a public examination can have ‘profound, immediate and long-term impact on a candidate’s life’(World Bank Group, 2001). This is particularly so when success in such examinations open opportunities for work and progression in educational ladder. According to Fibersima (2001) as stated by Afemikhe (2010) ‘examination malpractice leads to retarded educational growth and development, false sense of value and impression of candidates’ capabilities and loss of confidence in candidates. Resources are wasted leading to devastating, grave and serious repercussions’. Afemikhe (2010: 142) also quoting Chileshe (2010) indicated that:

Individuals involved in leakages lose their moral direction. They no longer recognise ethics as a value....Their policies will be warped and visionless. If they become medical doctors, they will wrongly diagnose their patients and send them to the grave. If they are lawyers, they would corrupt justice and promote unjust causes. As professionals, they will fail to comprehend the complex rules of business and lead a country into underdevelopment.

Assessment irregularities are exhibited in many ways. These according to Afemikhe (2010) and Gijima, Govender&Cesare (undated) include leakage, external assistance, smuggling, copying, collusion, intimidation and substitution. An examination of each of these malpractices shows the involvement of examiners, printers, proofreaders, school administrators, invigilators, candidates, supervisors and even teachers working outside the examination room.

The penalties to discourage malpractice have been features of public examination systems since the sixteenth and seventeen century, when in China penalties for malpractice included the death penalty and the exile of corrupt examination officials. The World Bank Group (2001:4) listed some control measures in this regard. They include among others paying examination officials slightly higher salaries than personnel in comparable positions in the Ministry of Education (Uganda), paper setters set individual questions rather than complete papers and reducing human access by replacing candidates’ names with examination numbers. In some cases original identification numbers are replaced by other numbers (fictitious roll numbers) and a record of matching numbers are stored on computer file (Lahore Board in Pakistan).

In addition to all these, other measures have been put in place to reduce irregularities ‘The candidates and other persons who are directly implicated in an irregularity affecting the validity of examination scores are usually subject to sanctions including: the exclusion from examinations, withholding or non-reporting of results and even decertification’ (ABEM, 2004; Linn, Baker & Dunbar: 1991). In South Africa there is usually an irregularity hearing (Gijima, Govender&Cesare, undated). In Nigeria, Act 33 of 1999 defines examination malpractice and sanctions for offenders. At the University of Benin, Benin City, malpractice handling involves receipt of report from invigilator by Dean/Director, report is reviewed, investigation panel is set-up, student is notified, panel decides on prima facie case through interviewing student and examining evidences and make recommendations based on extant rules. The recommendations are then forwarded to the Vice-Chancellor for approval and student is communicated about the outcome. Some examination bodies have also gone to the extent of embossing photo on candidates’ certificates.

Cyber-crime in Assessment

Since the inception of information and communication technology in almost all facets of human endeavour, cybercrime has become rampant. Cybercrime and computer-related crime are not the same. Whereas cybercrime is an activity in which computers or networks are a tool, a target or a place of criminal activity, computer-related crime could involve stand

alone computers (Charney, 1994; Goodman, 1997). Computer crimes encompass unauthorized or illegal activities perpetrated via computer as well as the theft of computers and other technological hardware. To the extent that computers may not be linked to any network, computer-related crimes are more than cybercrime.

Gercke (2012) has outlined a typology of cybercrimes for which the categories are however not mutually exclusive. It is composed of four components:

1. Offences against the confidentiality, integrity and availability of computer data and systems.
2. Computer-related offences.
3. Content-related offences.
4. Copyright-related offences.

The approaches utilized in perpetuating these varied crimes include hacking and cracking with hacking involving unlawful access to a computer system. This could involve breaking of password-protected websites; circumventing password protection of a computer system. This is usually possible when there is inadequate and incomplete protection of a computer system. There is also the issue of illegal data acquisition which can emanate from computers connected to the internet where attempts are made to access information via the internet. This is done through use of software to scan for unprotected ports or circumvent protector measures. Social engineering, which involves tricking people to release normal security procedures, can also be utilized.

There is the use of illegal interception of communication infrastructure and internet services. Data interference could also happen through deleting, suppressing or altering computer data through computer virus. There could also be system interference.

Security Issues in e-assessment

E-learning appears to be a precursor of e-assessment. E-assessment involves the utilization of computer screen or any Information and Communication technology related device for the purpose of assessment. Security is very essential for valid e-assessment delivery. Web security and e-assessment security are germane in this regard. While web security is a necessary condition but it is not a sufficient condition for security of ICT related assessments. Web security deals with the securing of the server/s running web applications as well as the application itself. This however does not guarantee that an e-assessment will be secure. Before the advent of computers, assessment were mainly of the paper and pencil type with the students well seated in a conducive environment and each given the opportunity to perform at his/her best. The security issues associated with e-assessment would now be presented. This section is based principally on the presentation by Marais, Argles & von Solms (undated)

Assessment of students whether assessment of learning, assessment as learning or assessment for learning presupposes that a particular student takes the assessment. It is only if this happens that we can have confidence in the assessment information which is generated. With paper and pencil tests difficulties have been noticed and that is why many suggestions for control of who writes the examination have been suggested. Among these is use of multi-mode biometric solutions, finger prints and photo-on-certificate. The issue of authentication is also an important issue in e-assessment. Authentication in e-assessment allows the student assess as well as provide confidentiality; only intended student should get access. The options available include the use of passwords, challenge response questions, e-token authentication, smart card authentication, and biometric authentication. The password is cheap to implement but can easily be abused and the other methods though more expensive and cumbersome can be used.

The integrity of the electronic device is very important. Electronic corruption which is the change of information on server can take place. It could also involve use of resources not

specified in the test or helping another student. Students should not have the opportunity to log in twice as this facilitates double submission for him/her and then for another person needs to be blocked. The server can be made to deny two logins.

Very often one needs to be sure that the e-assessment was executed from the appropriate place and hence what is submitted by the student can be seen as his/her own work. This is more important as e-assessment utilises groups of computers connected to the intranet or internet. As a result other services are available to the one taking the assessment. In the same way examinees outside the examination hall could also have access to the test for some reasons. Therefore it is mandatory to have locations for the e-assessment controlled. Firewalls can be used to distinguish different examinees and password can be utilised to gain access to the assessment task.

Test visibility is another teething problem; the questions being attempted by one examinee should not be seen by another. To ensure this a large question bank is utilised and the questions from it can be randomly selected in formation of tests to be attempted by different examinees. Confidentiality and privacy are also of essence. Each examinee should have access to his/her report and responses provided. Other unauthorised persons should be denied such access. If this is not controlled plagiarism could emerge. This issue is also addressed using password by the examinees.

Another import issue is availability of secure client/server software; this focuses on the set-up of the e-assessment clients/computers that are used by the students doing an on-line assessment. In this regard the machines must be appropriately configured so that they cannot divulge information pertaining to previous sessions. A firewall needs to be enabled on the client machine to protect against attacks from a person wishing to disrupt the e-assessment. The machines should also be protected from viruses and other malicious persons and program (WebCT Security, 2005).

Deniability of e-assessment submissions can also be a problem; a student who has submitted must not be able to deny doing it. Authentication is useful in this regard with use of biometric device and use of electronic signature. The response can be digitally signed with the student's biometric information. The point of registration is the first point to ensure this identification.

Conclusion

Research and assessments are two related terms that share common methodology. The differences between them were highlighted. The use of assessment was examined; the essence of assessment is utility. Otherwise it is an exercise in futility. Assessment within the school system which previously was mainly of the paper and pencil type has seen innovations with the emergence of information and communication technology in the form of computer based testing. The efficacy of these two modes of testing has witnessed researches which have yielded inconclusive results.

School-based assessment plays an important role in schools and it is felt that it can assist in solving some of the problems of irregularities in public examinations. Large-scale assessment was examined and it was posited that it is not bad in itself. It only becomes a source of worry when it serves a high-stakes function. Cybercrime in assessment and related security issues were equally treated. It is the expectation that what has been present would act as a strong pad to launch this conference.

References

Ablemann, C.H., Elmore, R.F., Even, J., Kenyon, S. & Marshall, J. (2004). When accountability

- knocks, will anyone answer? In Richard F. Elmore (Ed.) *School Reform from the Inside Out, 133-199*. Cambridge, MA: Harvard Education Press.
- Afemikhe, O. A. (2015). *Assessment in schools: A look at the big picture*. Inaugural Lecture Series 151 University of Benin, February 5, 2015.
- Afemikhe, O. A. & Imobekhai, S.Y. (2015). *Nigerian teachers' utilization of test construction procedures for validity improvement of achievement tests*. Paper presented at the International Association of Educational Assessment conference which held in Kansas, October 10 - 15.
- Afemikhe, O. A. (2014). *Educational measurement and evaluation*. Lagos: Amfitop Books
- Afemikhe, O.A. (2013). *Public examinations, quality of education and challenge of examination malpractice*. Paper presented at the Summit on repositioning NECO for service delivery at Chelsea Hotel, Abuja, Nigeria October 8 & 9.
- Afemikhe, O. A. (2010). 'Strategies for managing examination malpractice in public examinations' *Journal of Educational Assessment in Africa* Vol. 5, 139-152.
- Afemikhe, O. A. (2000). Assessment in National Educational Systems: The first twelve years of formal education in Botswana and Nigeria. *Mosenodi: Journal of the Botswana Educational Research Association*, Vol. 8, No. 2, 39-46.
- Afemikhe, O. A. (1989 & 1990). Implementation of continuous assessment in Nigeria primary schools: Observations of the Bendel State case. *Benin Journal of Educational Studies*, Vol. 8, No. 1 & 2, 205-211.
- Afemikhe, S. S. O. (2003). *The pursuit of value for money*. Ibadan: Spectrum Books Limited.
- Aiken, L.R. (1997). *Questionnaires & inventories: Surveying opinions and assessing personality*. New York: John Wiley & Sons, Inc.
- American Board of Emergency Medicine. (2004). *Policy on Examination irregularities*. ABEM.
- Baumer, M., Roded, K. & Gafni, N. (2009). Assessing the equivalence of internet-based vs. paper- and-pencil psychometric tests. In D.J. Weiss (Ed). *Proceedings of the 2009 GMAC conference on computerized adaptive testing*. Retrieved 13th November, 2015.
- Cameron, J. & Pierce, D. (1994). *Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis*. *Review of Educational Research*, 64(3), 363-423.
- Charney, S. (1994). Computer crime: Law enforcement's shift from a corporeal environment to the intangible, *Electronic World of Cyberspace*, *Federal Bar News*, Vol. 41, Issue 7, 489 - 494.
- Cheng, Y. C. (2001). *Paradigm Shifts in Quality Improvement in Education: Three Waves for the Future*. Invited Plenary Speech Presented at The International Forum on Quality Education for the Twenty-first Century Co-organized by UNESCO-PROAP, National Commission for UNESCO of Ministry of Education, and National Institute of Educational Research, China Beijing, China 12-15 June.
- Chu, H.C., Hwang, G.J., Tsai, C. C. & Tseng, J. C. R. (2010). A two-tier test approach to developing location-aware mobile learning system for natural science course. *Computers and Education*, 55, 1618-1627.
- Chuderwood, J. S., Zapata-Rivera, D. & VanWinkle, W. (2010). *An evidence-centred approach to using assessment data for policy-makers*. New Jersey: ETS RR-10-03.
- Clariana, R. & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, Vol. 33, No. 2, 593-602.
- Deci, E.L., Koestner, R. & Ryan, R.M. (1999). A meta-analytic review of experiments

- examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627-668.
- Erwin, T. D. (1991). *Assessing Student Learning and Development: A Guide to Principles, Goals, and Methods of Determining College Outcomes*. San Francisco: Jossey-Bass.
- Frazer, M. (1992). Quality assurance in higher education. in Craft, A. (Ed.). *Quality Assurance in Higher Education (9-25)*. Proceedings of an International Conference, Hong Kong 1991 London: The Falmer Press.
- Gijima, C. L., Govender, P. & Cesare, J. *Dealing with Examination and Assessment Irregularities in an Era of Rapid Change: a South African Judicial perspective*. http://www.iaea.info/documents/paper_1162a1953f.pdf
- Glidden, G. (1996). *Assessment irregularities: A discussion guide for current issues in test administration*. Retrieved from: <http://www.natd.org/Glidden96.htm>.
- Goodman, M.D. (1997). *Why the police don't care about computer crime Harvard Journal of Law & Technology Volume 10, Number 3, 466-494*.
- Harvey, L. and Green, D. (1993). Defining quality. *Assessment and Evaluation in Higher Education*, 18(1), 9-34.
- Hwang, G.J. & Chang, H. F. ((2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievement of students. *Computers and Education*, 56, 1023-1031.
- Information Systems Audit and Control Association (ISACA). (2005). *Irregularities and illegal acts*. Illinois. USA.
- Koretz, D. McCaffary, D. & Hamilton, L. (2001). *Toward a framework for validating gains under high-stakes conditions (CSE Tech. Report No. 551)*. Los Angeles, CA: CRESST/CSE, University of California, Los Angeles, Graduate School of Education and Information Studies. Retrieved on October 10, 2015, from <http://cresst96.cse.ucla.edu/CRESST/Reports/TR551.pdf>
- Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher* 20 (8), 15 -23.
- Marais, E., Argles, D., & von Solms, B. (undated). *Security issues specific to E-assessments* <http://nces.ed.gov/pubs98/safetech/glossary.asp>
- McGillin, V. (2003, December). Research versus assessment: What's the difference? *Academic Advising Today*, 26(4). Retrieved from <https://www.nacada.ksu.edu/Resources/Academic-Advising-Today/View-Articles/Research-versus-Assessment-Whats-the-Difference.aspx>
- Michigan Association of School Psychologists (2004). *Position statement on the use of the Michigan Educational Assessment Program (MEAP) for high stakes decisions for students in Michigan*. Retrieved January 2, 2005 from the MASP Web site: <http://www.masponline.org/High Stakes Testing in Michigan.htm>
- Muhwava W, Nyirenda M, Mutevedzi T, Herbst, K. & Hosegood, V. (2008). Operational and methodological procedures of the Africa Centre Demographic Information System. *Monograph Series no. 1 African Centre for Health and Population Studies*. Somkhele: South Africa.
- Obanya, Pai (1979). Improving science teaching and learning through continuous assessment. *Journal of the Science Teachers Association of Nigeria*, Vol. 18, No.1, 29-39.
- Queensland Studies Authority (2010). *School-based assessment: The Queensland system*. Brisbane, Australia: The State of Queensland (Queensland Studies Authority).
- Simner, M. L. (2000). *A joint position statement by the Canadian Psychological*

- Association and the Canadian Association of School Psychologists on the Canadian press coverage of the province-wide achievement test results.* Retrieved from http://www.cpa.ca/documents/joint_position.html
- Shulman, L.S. (2003). *No drive-by teachers.* *Carnegie perspectives.* Retrieved from <http://www.carnegie foundation.org/perspectives/no-drive-teachers> October 16, 2015.
- Spillane, J. P. (2005). Primary school leadership practice: How the subject matters. *School Leadership & Management*, 25(4), 383-397.
- Stecher, B. M. (2012). *Consequences of high-stakes, large-scale testing on schools and classroom practice.* Retrieved from: <https://larrycuban.files.wordpress.com/2012/01/mr1554-ch4.pdf>
- TechTalk, (2005). *Reporting irregularities and discrepancies.* Retrieved from: <http://www.michigan.gov/techtalk>. Accessed.
- Teeroovengadum, V., Kamalanabhan, T.J. & Seebaluck, A. (2010). *Service quality in the education sector: Moving towards a holistic and transformative approach.* Paper presented at International Research Symposium in Service of Management, Le Meridien Hotel, Mauritius, 24-27 August.
- The World Bank Group. (2001). *Public examination system: Topics – Malpractice.* The World Bank Group. 2001. Retrieved from: <http://www.worldbank.org/education/exams/malpractice/asp>.
- UNICEF (2000). *Defining Quality in Education.* A paper presented by UNICEF at the meeting of The International Working Group on Education in Florence, Italy.
- Upcraft, M.L., & Schuh, J. H. (2002). *Assessment vs. Research: why we should care about the difference.* About Campus / March–April, 16-21.
- Urbina, S. (2004). *Essential of psychological testing.* New Jersey: John Wiley & Sons, Inc.
- Van Damme, J., Opdenakker, M. & Van Landeghem, G. (2008). *Educational effectiveness: An introduction to international and Flemish research on schools, teachers and classes.* Leuven: Acco.
- WebCT Security, (2005), 'How to protect your identity and use WebCT at the same time', *College of Charleston Academic Computing*, http://www.cofc.edu/~webct/faculty/WebCT_Security.pdf.
- Yusof, H. N. M. (2013). *School-Based assessment: transformation in educational Assessment in Malaysia.* Paper presented at the Cambridge Horizons seminar on School-based Assessment: Prospects and Realities in Asian Contexts which held in Subang Jaya Malaysia on 3 June. (<http://www.africafiles.org/article.asp?ID=24118;accessed 23/9/10>).