USE ASSESSMENT INFORMATION FOR IMPROVING THE QUALITY OF EDUCATION

Cyprian Cele

cypriancele@yahoo.co.uk.

Abstract

It has been known for a long time that educational assessment is no longer carried out only to assess the progress and achievement of the learner and nobody else. Assessment outcomes provide a basis for making decisions at various levels about the progress of education. Decisions are made concerning the learner, teacher, school management and education planners and supervisors. This paper explores three areas where assessment results are used as part of a basis for decision making. These areas are: use of assessment outcome for improving the quality of instruction and learning; use of assessment outcome for accountability decisions; and use of assessment outcome for fostering an inclusive education system. The focus was to stimulate thinking and provide a framework for research and presentations. The assessments discussed here include classroom assessment, school level assessment, national examinations, and survey studies at national, regional, and international levels. These assessments generate a lot of data which may not be adequately utilised. Teachers may not be adequately trained in assessment and therefore cannot fully use the information provided for improving instruction. Holding someone responsible for assessment outcomes requires collateral data to inform decisions. Providing an inclusive education, particularly with mainstreaming, may be challenging to educators at all levels. The effectiveness of assessment utilisation is still limited by such factors.

ISSN 1021-559X /09/2019
Mosenodi Journal

© Cyprian Cele

Introduction

Every country aspires to give the best education to its citizens. Curricular are developed, implemented, and revised from time to time. Revision has become frequent in this digital era which results in knowledge and skills explosion. For a country not to remain behind in a fast changing global world, it has to keep an eye on the fit between its education system and the need for tomorrow's world. That is why the world has embraced The United Nations Development Programme Sustainable Development Goals (SDGs) adopted in Rio 2012 to be implemented 2016-2030. SDGs comprise 17 goals for development. Although education is placed as goal number four, it is the major driver of all other goals. While in Millennium Development Goals (MDGs), Education For All (EFA) emphasised quantity in education, SDG number four is emphasising quality. Attainment of quality in education is like undertaking a long journey into a destination we do not know the route to, but with the curriculum as our guide. This curriculum must capture the hopes and aspirations of a country as the ideal development position to be in after successful implementation. The curriculum must be serviced through fine-tuning and sometimes through drastic intervention as new knowledge is gathered.

The teachers and education managers drive the curriculum. Educational planners and supervisors are responsible for supplies, keep an eye on the implementation, and offer guidance to the implementers.

We spend money implementing the curriculum, and also check if that money is helping us get to the goal. The check comes through assessments which are carried out at various levels. Assessment is a teacher's instructional tool. As instruction takes place, the teacher has to evaluate the impact of her efforts and decide on individual learners as well as what should be adjusted overall for better learning. There are institutions dedicated to assessment and research. Their products provide information on the progress of education at a broader level than the assessment of the classroom teacher. They evaluate the system and the condition of the learners. Curricular evaluation and revision is a way of guiding education towards quality. The decision to revise curricula is based on perception of need for updated knowledge and skills which the current curricula cannot provide. Information necessitating revision must not be speculative or popular beliefs, and sayings, but must be based on data collected from educational research and assessment. Good decisions are based on different kinds of information (Nitko & Brookhart, 2011). Information provided through assessment and research must be accurate, otherwise we will keep making wrong decisions without guiding education towards the desired end. That is why we are here, to check ourselves as to whether we are providing useful information that is being used for advancing the quality of education.

The individuals and institutions represented here develop and implement curricula, carry out research and assessment and some oversee the implementation of educational plans. They are therefore pivotal in the provision of information needed for guiding decisions on the progress of education. We can therefore at this conference confidently focus on assessment as a tool for fostering the provision of quality education in an equitable and sustainable way. In particular, the conference is exploring whether the information we provide from research and assessments is adequate and is being used to support improvement of the quality of learning and decisions on who to blame when learner performance indicates there is inadequate learning.

We have used the word 'decision' several times already and shall use it very many times more in this paper and therefore a brief reminder on some components of decision theory may be helpful. Making a decision implies that

- one has opportunity of choosing among several lines of action;
- each line of action leads to one or more outcomes;
- values can be placed on each outcome;
- the alternative line of action with outcomes of the highest values is chosen.

Educational practice requires decision making at all times. The classroom teacher delivering a lesson is making decisions from moment to moment. The managers up to educational planners must of necessity make decisions at appropriate time to guide the frontline teachers. To be useful for decision making, the assessment information must be reproducible and evidence is adducible that its use is not associated with negative consequences. As we explore the conference themes on assessment, and the questions posed here, we are looking for lines of action that will provide education with the best values.

There are many uses to which assessment information can be put. Frechtling (1989) discusses three ways in which test information is used by administrators:

- reporting student achievement to the public;
- evaluating program or curricular effectiveness;
- enforcing educational accountability.

Do we provide the public with adequate information on what learners have achieved? Have we accurately documented effectiveness of the curricula we implement? Are we in Southern Africa using assessment information for policy formulation, evaluating the effectiveness of curricula and thereby making decisions for the advancement of the quality of instruction and learning? Information technology has resulted into devising various ways of doing and reporting assessment. Have we made adequate use of ICT in our assessments and what potentials of ICT could enable us to improve on the speed and quality of the information we provide? How do we use the massive data generated during an assessment? Is the information we provide suitable for assigning accountability in cases where performance of learners is less than adequate?

Learners are not equally gifted by nature. Some are physically or intellectually challenged or even both. They are in the minority, but are members of our society and information on their performance must be provided. Their experiences in educational institutions are different from the experiences of the unchallenged learners. Their learning requirements, learning styles and way of responding to assessment tasks are not the same as those of the majority of learners, even when accommodation has been made. Should we compare the performance of these learners with the performance of the unchallenged learners? SDGs are advocating mainstreaming of people with disabilities by 2030. What research and assessment information do we have to support this aspiration?

What assessments are we talking about? Assessment is carried out internally as well as externally. Internal assessment may be an individual teacher's classroom assessment or school wise assessment carried out periodically. A teacher's assessments may be oral questions, class discussions, practical work, or tests administered from time to time. They serve as feedback to the teacher as to how well learners have mastered concepts and skills. Resourceful teachers make use of these assessment results for improving the quality of instruction. Some schools carry out psychological testing, particularly to diagnose learning difficulties of some students.

Externally, every country represented here has end of cycle summative examination, often at the primary and secondary levels. Assessment at tertiary level is part of this discussion. The main interest in these results as far as learners and parents are concerned is whether the performance is good enough to take them to the next level of education. At school level, these results stir up a lot of thinking and activities. The teacher is keen to know if candidates did well in her subject compared to previous years and subjects taught by other teachers. Teachers may make notes of how the results can be used to better their teaching and understanding of the next set of learners. School administrators have the same interest, but more particularly look at the performance of their candidates in comparison to the performance of candidates from other schools. With increased public participation in the provision of education, the release of results is a time for publicity. It is also a time when schools may headhunt teachers they consider capable of enabling learners to produce good results. Are there ways in which these results are utilised effectively to focus on instructional and learning improvement rather than its selection, certification and league table functions?

Apart from achievement testing, some countries carry out national survey studies periodically. Examples of national assessments are Lesotho National Assessment of Educational Progress (LNAEP), and National Assessment of Progress in Education (NAPE) in Uganda. Some participate or have participated in large scale regional assessments, such as Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), Monitoring of Learning Achievement (MLA), and international studies like program for Progress in International Reading Literacy Study (PIRLS) and Trends in International Mathematics and Science Studies (TIMSS). These assessments provide system level information, with disaggregation into factors like gender, school location, and teacher factors. The assessments are useful in comparing countries, and making decisions on educational reforms. The use of item response theory (IRT) makes it possible to monitor trends in performance over time. They are low-staked assessments for learners, teachers, and even schools. Do the sampled learners put out their best on the tests to support a conclusion that the observed performance is authentic? How have we benefitted from these assessments in making policy decisions to improve educational quality? Have classroom teachers actually used reports from these assessments and improved on their teaching as well as student outcomes?

Proper test use is based on what the test is designed for. Very often educational curricular are prefaced with lofty goals like raising standards to world-class level, fostering equity, developing skills (Rudman, 1993). The assessment personnel has to go from the specific objectives or group of objectives and construct items the score on which can be interpreted in terms of those goals. Do we really assess and give information on the extent to which those goals are being achieved? Assessing what a learner knows and can do is not as easy as measuring height or assessing speed of running. In a paper presented at an Exploratory Seminar: Measurement Challenges, Pellegrino (2009) argues that we cannot really know what a student knows and can do. In designing assessment we have to ensure that Curriculum, Instruction and Assessment hang together so that an assessment does not become the servant tail that wags the educational dog. Frechtling warns that:

> *... a test score provides an estimate of what a student knows. Making a leap from this estimate to a conclusion regarding policies, institutions, or people requires caution and careful consideration of alternative explanations.*

That is why care is needed in interpreting and taking action on the basis of test scores. Frechtling (referred to above) points out that the accountability movement in the US was foreshadowed by performance contracting of the 1970. In this kind of contracting, a private education provider would agree with the Office of Economic Opportunity to teach underprivileged children to reach specified standards in

reading and writing. If the standards are not met, the company does not get paid. Most companies failed to meet performance standards and consequently lost money. Why did this happen? Learning effectiveness can be complex and needs time and care to achieve. Later on school effectiveness and school improvement projects came into being, with application of accountability. We shall consider accountability decisions made on the basis of assessment information, how widespread accountability decisions are, and whether they are fair and result in improving the quality of education.

The primary purpose for testing learners is to find out if teaching and learning are effective. We are gathered here to actively listen and learn from one another what we are doing, the successes and challenges we are meeting, new trends being experimented with, and the practices that have been adopted or need to be discarded, all for the purpose of making assessment more supportive of teaching, learning and curricula reviews. This contribution will concentrate on:

- use of assessment information for improving the quality of education;
- use of assessment information for accountability decisions;
- suitability of assessment information on special needs learners.

As you can tell from the paragraphs above, I am merely asking questions. Your papers and participation will provide the answers. Further, I am addressing only some sub-themes, and therefore these notes do not cover the conference comprehensively.

Using Examination Feedback to Support Engaging Learning Environment

Using assessment information to improve instruction and learning is not a new agenda. Gronlund (1985) notes that tests are no longer administered to evaluate learners and schools. Formative testing, diagnostic, and mastery testing are efforts integrated into instruction for improving learning. Computer use came in to provide more detailed scoring so that objective level and even item level information is available to aid the teacher. Thus National Assessment of Educational Progress (NAEP) used to report performance on specific objectives to facilitate use of assessment results for instructional improvement. They now report on an IRT scale with performance bands of *Basic*, *Proficient* and *Advanced*. Test information helps the instructor to know which concepts and skills the learners have not mastered. The data may come from oral questioning, classroom discussions, homework, formal tests, and other forms of assessment available to the teacher. It is not only learner weaknesses that can be identified from the school or classroom assessments, but the teacher may also discern how his instructional approaches can be made more effective in facilitating learning.

In this conference we go beyond the teacher's classroom assessment. We bring in the summative examinations and the large scale survey studies enumerated earlier. To understand whether these assessment data are being used to improve educational quality we need to look more closely at the information provided by these assessments, how they are communicated to the teacher, whether they are actually used in planning school level work schemes and classroom instruction. The main reports provided in terminal achievement testing contain both quantitative and qualitative information. Quantitative information provided includes:

- number taking the assessment;

- profile of performance per subject;
- percentage passing the assessment at specified levels;
- disaggregation of performance at various levels by gender, region/districts, etc.;
- comparison of performance with performances in previous years;

Item level information may be given, like in the case of computer scored objective components.

This numerical information is of great excitement to the public. Learners and parents value them mostly for what they convey about learners' chances of advancing in education. Education managers and providers value them for the feedback on learning progress, actions that can be taken to make learning more effective and judgment that may be passed on them about their performance. What about the classroom teacher?

Indeed a teacher is keen to see the performances of learners in her subject. Some teachers may get inspiration on how to prepare the next lot of candidates better. For otherwise, the main interest could be in searching for bragging rights for being a teacher that manufactures the highest number of distinctions in a school. Where a school provides rewards for teachers whose learners do well in examinations, this may be the major interest in the information provided. Qualitative information is obtained from examiners' reports at grading meetings. Examiners who conduct or moderate practical assessments often write reports on what they observe in the field. Otherwise background information captured in summative assessment is limited. Unless examination processing system is coded in such a way that content and background variables are captured, examination bodies are under pressure of turn-around time and may not provide elaborated information on performance correlates. Survey assessments enumerated before are system based and the sampling design may not give information specific to a learner or institution. However, Verger, Parcerisa, & Fontdevila (2018) carried out analysis based on data from National Large Scale Assessments (NLSA) and Programme for International Student Assessment (PISA) and noted that these large scale assessments are increasingly being used around the world to promote Global Education Reform Movement (GERM). GERM encourages use of standards, accountability and decentralization. In 'governing by numbers', these assessments are more and more providing a means of holding schools accountable.

Friedman et al. (2018) through the Australian Council for Educational Research utilized a contract with UNICEF to survey large scale assessments carried out in Southern and Eastern Africa, how the assessments are carried out and how the resultant scores are used to improve education at the primary level. The content areas of interest were literacy and numeracy. Contextual factors that were linked to cognitive performance included individual characteristics, gender, age, home language, socio-economic status, pre-schooling, school location and home resources. Teacher training, availability of teaching and learning materials and home based reading were found to be significant in most systems. Other than countries like Lesotho, Kenya, Ethiopia, Rwanda and Zimbabwe, most of the participants had not applied Item Response Theory to the data. These assessments can be effective if an assessment is followed by analysis, dissemination and action. What is the outcome of these large-scale assessments that can be of value to the classroom teacher in our context? The information is similar to what is provided in summative assessment:

- profile of performance per subject. Not many subject areas are assessed in any one cycle;

- percentage passing the assessment at specified levels;
- disaggregation of performance at various levels by gender, region/districts, etc.;
- comparison of performance with performances in previous assessments (trend).

Scoring or coding of the work of learners can be done in such a way that the strengths and weaknesses of learners in each item or group of objectives can be reported. Moving from regional to international assessments gives individual countries curricular information, but individual schools get general information. A national curricular overlap with the agreed content for an assessment becomes reduced. These assessments would seem to be more suitable for curriculum planners, education providers, and administrators than for teachers.

Whether we are dealing with national examinations or large scale surveys, the important point here is the extent to which we are ploughing the massive information generated into uplifting the quality of education. Are quantitative and qualitative information received, understood and used by classroom teachers to improve instruction? While it is important to know the performance of subgroups, do we actually make use of this feedback to improve instruction? Papers describing how large scale assessment reports are impacting the quality of education will greatly illuminate the usefulness of these studies.

There is one aspect of large scale assessments I want to mention: **dissemination.** A format often used in disseminating the results is to call education representatives to a central place to go through the report. Thereafter dissemination is supposed to take place at regional up to school level. Do we succeed in getting down to school level and enabling school teachers to understand the report and extract what is useful for classroom instruction? Do schools that were not sampled take any interest in the report? What benefits have classroom teachers enjoyed from these large scale assessments? Curriculum developers and education providers must interact with the information provided by these assessments in a professional, not defensive way, to ensure that information feeding into reform effort is authentic and useful.

When quantitative information is provided, the assumption is that the teacher will be able to read, understand, and take corrective action. This may be the case. A straightforward item analysis report may look simple, but not every teacher can interpret the output. Consider the multiple-choice item analysis output (*based on Iteman*) below.

Option	*N*	*0-33%*	*33-67%*	*67-100%*	*Color*	
A	1771	0.10	0.03	0.01	Maroon	
B	36943	0.74	0.93	0.98	Green	**KEY**
C	1993	0.12	0.03	0.01	Blue	
D	729	0.04	0.01	0.00	Olive	

(On the basis of total score on the test learners are divided into bottom, middle and top scorers. N is the total number of learners who selected each option and the percentage of learners in each group who selected each option is indicated).

Can all our teachers get meaning and make use of the information portrayed here for their instructional improvement?

Some of us may have moved into modern test theory and are applying item response theory. The outputs from item and test analysis based on IRT may be much harder for the bulk of teachers to interpret than classical output. The information therefore may not be very useful for helping a teacher improve his instructional approaches. What mechanism do we have in place to ensure that assessment feedback to schools are understood and used to support instruction and learning?

An assessment is an evaluation exercise. Evaluation report provides information and conclusions. In their *fourth generation evaluation*, Guba and Lincoln (1989) argue that evaluation reports tend to be tailored to the interests of the client and the evaluator. The stakeholder may not be served by these reports. That implies holding back vital information from the stakeholder who may be a pivotal link in achieving the objectives of a program. Are there useful aspects of our assessment information not released to the public because they are considered unsuitable for public consumption? In order to effectively use assessment information as a feedback we need to:

- provide information it generates;
- assess the comprehensiveness of information provided;
- ensure all stakeholders understand the information;
- identify who should act on each source of weakness;
- set up a mechanism for overseeing that action is taken;
- explore innovative ways of providing and disseminating assessment information.

Using Assessment Information for Accountability Purposes to Improve the Quality of Teaching and Learning

Parents provide home education to their children as far as they can, but they acknowledge the limitations of family education and accept that organised institutions are places that can take their children beyond what they can teach at home. Educational institutions are built to pass on relevant knowledge and skills which learners can use for the betterment of lives. Not only that, but the schools and institutions are places of enquiry to discover new knowledge and ways of doing things. A person who has been to an institution should come out different from the one at the gate of an educational institution. It should be obvious that she has acquired the relevant knowledge and skills and has better behaviour repertoire than the one who did not receive that level of education.

It costs money to build, and equip these institutions with educational materials and recruit and maintain manpower to operate them. Never the less, countries still spend their money on education because ignorance would not help them to reduce poverty and disease. With all the money, time, and energy spent on education, should it be business as usual if learners do not show in their performance that they have acquired the knowledge and skills they were sent to the institutions for? In establishing a curriculum and entrusting an institution with the responsibility of implementing it, we expect every child to attain the highest standard possible, without extraneous circumstances barring any child from maximising his potential. The education system is supposed to remove all obstacles to instruction and learning. The teacher

is expected to exert himself to the maximum so that the learner achieves what he is in the institution for. The learners are expected to do their best to ensure they actually learn.

If children who have gone to school are found wanting in their performance on assessment; if they cannot do anything with the knowledge they are supposed to have acquired in school and do not show that they are better prepared to live in their communities, education can be regarded as having failed to deliver the required results. Should someone be blamed? Take selection as one use to which assessment information can be put. A senior secondary school receives students with good grades from a junior secondary school. These students are aiming at becoming lawyers, doctors, engineers and all sorts of professionals. Their parents pay all the fees required for their education. At the end of it all, these children do not do well enough to be selected for the courses they want. Limitation of places requires that selection and placement be made. Failing to be selected implies that the opportunity to get new knowledge or skill or to better the present status of knowledge and skills is lost. Learners who perform well rejoice while their counterparts are in distress. This rejection may close all ways for the learner to develop a career that would improve his economic and social standing. Should questions not be asked as to why these students did not do as well as was expected of them? Accountability seems to be a justifiable line of action if there is evidence of inadequate learning and change in behaviour. Accountability is about hanging: '*who gets hung and by who*?' (Mehrens & Lehmann, 1969).

Are the assessment reports we issue suitable for making decisions that can affect the lives of education implementers? A measure of the quality of assessment is the extent to which the test addresses the objectives of education specified in the curriculum. To enhance the value of assessment information, various approaches to assessment are implemented besides timed paper and pencil tests. Coursework, portfolio assessments are some of the ways of varying sources of information. In some subjects, authentic assessment requires that the end performance in that field be assessed. Physical Education, Foods and Nutrition, Music, Dance and Drama are some of the subjects that require that samples of performance be observed. Visiting assessors are often used to obtain these scores. Some examining bodies may have resorted to statistical moderation to reduce costs. These procedures have their demands. Sieving information from portfolios is time consuming and can be expensive.

Some countries have expressed a concern that a single summative examination does not reflect the true achievement of a learner and therefore continuous assessment need to supplement summative assessment. Tanzania is one country that took continuous assessment very seriously. Gandye (1993) points out that the need for urgent educational reform led the ruling party to issue the 'Musoma Directive' in 1974 to the effect that emphasis on written examinations must be reduced by incorporating a learner's classroom progress, other performances and character to constitute his achievement. These school based assessments were to be carried out on a continuous basis. Though there were some administrative hitches at the beginning, Gandye reports that the reform eventually worked. Some other countries could have more information on the success in incorporating results of school based continuous assessment, including non-intellective measurement, into final summative assessment. Multiple measures are known to increase confidence in using the assessment result for decision making. Are these multiple measures sustainable?

Can assessment results issued by examining bodies be used for accountability decisions? Assessment results contain a lot of information and can therefore be used cautiously for that purpose. Chatterji (2019) offers guidelines for the use of test information for accountability purposes. If test

information is wrongly interpreted and used, it can block the attainment of education that the planners and stakeholders intended. He bases his guidelines on *Standards for Educational and Psychological Testing* (2014). If performance on tests is used for accountability, what would be the consequences on instructional practices, authenticity of school based assessments, and administration of tests? He provides extensive evidential considerations that must be taken into account before using test information for accountability, such as:

- Test Content: What does the test content cover; how are the items scored and how can we interpret the scores from these items? Summarising test information for reporting purposes is a data reduction process. A lot of detail is lost.
- What do the test scores tell us about the present and predicted status of a learner?
- Are there studies to validate the dimensionality of the test items so as to facilitate interpretation?
- What would be the consequences of using learner performance for accountability decisions?
- Can the accountability decision made be appropriate for all schools and for all types of learners?
- Are the scores reliable so that accountability decisions made in one year can be applied to scores from tests in subsequent years?

Assessment reports as we issue them, whether achievement or survey studies, are research products. We need to interrogate the products to decide whether they are providing relevant and sufficient information to support education process and decisions for accountability? We need to question what we do and the reports we give out. We may be able to deduce how much a person has learnt, and possibly what a person can do? But are we assessing whether the person has been transformed by going through learning institutions? Do our assessment reports capture all that schools do to provide a child with education? Does a school that excelled in co-curricular activities, and trained learners to be good citizens but done poorly in examinations deserve to be blamed? Are those assessment exercises tapping all outcomes an education system is designed for? We cannot meaningfully use results based on an assessment with validity questions for making accountability decisions.

The classroom teacher may be the front target for accountability purposes. Before she is hanged, further questions need to be raised. Were the children willing to learn? Were the parents supportive of the children's learning? In what community setup does the child live? Is the child equipped with skills to sieve what is important from a mess of activities going on around him? The school itself may be at fault. Is the teacher getting a living wage in time? Where does the teacher live? Is the teacher provided with requisite instructional materials? The teacher, as the one directly interacting with the learners needs to be understood if we are not to destroy the education system through blames. How was the teacher trained? Does anybody observe the teacher in action and provide corrective feedback? Other than content, was the teacher helped to acquire assessment skills? We want learners to develop their practical skills. What if the teacher is not provided with the necessary materials? Has the teacher understood the curriculum? What sort of in-servicing has the teacher undergone? Do we actually have information on teacher subject knowledge and her pedagogic approaches? Can we quantify the contribution of a teacher to the learning achievement of a child?

In reviewing the work of Chetty et al., Adler (2014) concluded that the claim that higher value-added teacher score adds to learner greater economic success in life is not supported by data. Using

assessment reports for making educational policies without interrogation can lead to undesirable consequences. Apart from proper preparation of learners to take assessment, valid scores require that test instruments be kept properly and the administration handled with integrity (Cizek et al., 2012). Holding one accountable has consequences (Smith & Rottenberg, 1992) that can be destructive. Teaching to the objectives specified in the tests, paying more attention to the learners a school or teacher considers capable of doing well at the expense of the poorer ones, giving up by teachers or learners, are some of the consequences that can arise if test scores are not used with care.

The quality of education is not going to improve if accountability ends with the teacher unless the information obtained indicates so. The learners, parents, communities in which they live, and managers of institutions also need to be assessed when learner performance is below expectation. At national level, the role played by upper administrative levels need also to be brought into focus when education is not achieving its objectives. We need to be sure that there is no obstacle above the teacher inhibiting her from enabling learners to do their best. Should we emphasise use of assessment information for taking corrective measures like training and supply of instructional materials? To be sure, failure to do one's job requires being held accountable, but we need to use multiple sources of information for decision making.

We must now turn to fairness.

4. Equity and inclusivity: being fair to everyone

We are familiar with the history and development of mental testing, spurred on by a realisation that in many activities, individuals differ. Kinnebrook (Cronbach, 1949) lost his job at the Greenwhich Observatory in 1796 because his record of when certain stars crossed the field of the telescope was consistently lower than that of his boss. He was considered incompetent, but 20 years later it was discovered that even among experts there are differences in the speed of response to stimuli. It would be going out of our way to become judgemental and take sides on the decision of International Amateur Athletic Federation for trying to remove a source of individual differences through medication to lower the amount of progesterone in the body of some athletes. Perhaps if the reasons why Federer is so good at tennis, Wood can rise from the dead when it comes to golf, Messi can kick a football to follow a semi-circular arc, and Bolt tries to compete with lightening could be found, someone may want to remove what makes these stars shine. Individuals differ in many ways:

- physical abilities;
- mental capabilities;
- subgroup factors like gender, religion, places of abode, etc.

If individual differences did not exist, assessment industry would be redundant. Tests and assessments attempt to quantify or describe the amounts of these differences. Cronbach (1990) points out that testing of abilities has always been *intended* as an impartial way to perform a political function of determining who gets what. Individual differences unrelated to the construct under test must be controlled for if the measurements are to be considered fair for allocating rewards.

Not all tests can be fair to everyone. There are learners with diverse special needs. Some are blind, deaf, or cognitively challenged. Some have no arms and therefore use their mouths or legs to write and manipulate objects. A chemistry test to identify radicals in a compound would not be fair to a blind learner. Manipulating apparatus and observing colours are outside his capability. These handicaps make it a

challenge to implement adequate accommodations for everyone to participate in an assessment and obtain scores with the same meaning. Smarter Balanced Assessment Consortium (2018) provides extensive guidelines for valid assessment of learners with various disabilities.

For accommodation to be useful it should comply with some conditions:

- accommodation should make it possible to administer the intended test to the learner;
- it should cater for each learner's unique characteristics;
- accommodation should not alter the construct being measured.

Opportunity to learn is an important consideration in judging test fairness. Mainstreaming may require that deaf learners sit in the same class with the learners who can hear. There are not many schools for the deaf and therefore most of these are mainstreamed. The medium of instruction is verbal. The deaf may not know any sign language. Teachers for the deaf are not enough to be assigned to every school and every class within a school. Learning of this group is therefore greatly hampered. Should they take the same test which learners who can hear take provided it is accommodated? Similar considerations apply to learners with other disabilities.

Apart from physical and intellectual differences, there are group differences that work against uniform assessment. Learners live in various communities. There are learners who by necessity must disappear from school to go hunting, fishing, planting crops at the beginning of rainy season, or begging to make a living. Their skills are geared towards the way their communities make a living and schooling may not be seen as of immediate benefit to them. They miss the objectives which are taught when they are absent. How do we handle learners who lose learning time while honestly chasing other requirements for living?

Equity and inclusiveness of an assessment should not be considered only on the basis of accommodations made in the tests. Accommodation starts with the way we treat and feel about our disadvantaged populations. Do we provide these groups with suitable access routes, bathrooms, seats, and tools they need for learning to their best? Would these provisions make a difference in their learning achievement? Resource allocation plays a vital part in inclusiveness. These groups must be valued and provided with opportunities that are best for their learning styles. Testing should be consistent with the best practice in the ways they learn.

Assessment should give a learner an opportunity to display her best. There are learners who are specially gifted. They grasp concepts instantly. If a learner is so good in mathematics that she needs more advanced topics, do we have capacity to identify and develop such an individual? If their talents are not recognised, such persons can lose interest in learning and subsequently not display their best performance. Equity issues need not be confined to grouping by disability or other factors only. It applies equally well to the assessment instruments themselves. Most of our countries have one assessment provider for a given level of education. Are the assessment results consistent within a subject that offers alternative papers? Do between assessors, subject and year to year differences exist? Where more than one institution assesses the same subjects, a need for comparison arises. In England, the issue of comparability came into focus because of the existence of London, Oxford and Cambridge as providers of assessment (Newton, et al. 2007). The concern was in the comparability of demands of matriculation standards. Would a student fail to get into a course or into a university because she was assessed by institution X and not Y? Since then comparability

studies have been taken seriously. There may be hope for free movement of labour within and beyond our region. Is there a sound comparability procedure to ensure fairness in qualifications? What best practices on equity and inclusiveness do we have to share? Are there institutions providing alternative assessments that truly bring out the best in special learning groups?

Summary Remarks

In sharing our experiences on how assessment data are used for instructional improvement, we may agree that assessment information must be ploughed back into improving learning. Information from assessments helps us to give feedback that can guide the process of instruction. We need information on whether quantitative as well as qualitative reports of the assessment outcomes are received, understood, and used by teachers for instructional improvement. Is it feasible to, or do we have assessment information on line for the use of teachers?

We may have research reports on the validity, dependability, and adequacy of the information provided through assessments. Exploration could be done on innovative report formats that are more easily understood and can be attractive to the classroom teacher. Assessment should tell us how to improve pre and in-service teacher education. Well trained teachers who are adequately resourced may be an asset in attainment of quality education. Assessment for learning may be more productive in uplifting the quality of education than assessment of learning. This is not to condone irresponsibility. Education is expensive. It follows that when learners do not achieve the expected level of performance, someone should account for it. How do we use assessment information to partition accountability? What impact do such actions have on educational quality?

The demand for goal number four in SDGs aims at mainstreaming people with various disabilities. Validation should pry into adequacy of the ways we teach, test, and give feedback to the teachers and learners with disabilities. It is upon us to ascertain that the scores of people with disabilities can be given the same meaning as the scores of people in the mainstream. Should the performance of learners with disabilities be compared with the performance of the majority? Empirical findings are needed to address these questions rather than depending on theoretical rationales alone.

References

Adler, M. (2014).Review of measuring the impact of the teachers. *National Education Policy Centre.*

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Chatterji, M (2019). A Consumer's guide to testing under the Every Student Succeeds (ESSA): What can the Common Core and Other ESSA Assessments tell us? Paper prepared for National Education policy Centre (NEPC) at the University of Colorado, Boulder.

Cizek , G., van der Linden, W., Cook L. (2012). Statewide student assessment programs test and data integrity. National Council of Measurement in Education

Cronbach, L. J. (1949). Essentials of psychological testing. Harper & Row, New York.

Cronbach, L. J. (1990). Essentials of psychological testing.(5th ed.). Harper & Row, New York.

Frechtling, J. A. (1989). Administrative uses of school testing programs. In Linn, R.J. Ed. (1989). Educational measurement. (3rd ed., p475-484). American Council on Education. N.Y.

Friedman, T., Schwantner U., Spink, J., Tabata, N., & Waters C. (2018). Improving Quality Education and Children's Learning Outcomes and Effective Practices in the Eastern and Southern Africa Region : Main report. Australian Council for Educational Research (ACER).

Gandye, P. (1993). Continuous Assessment: The Tanzanian Experience. An edited aper presented at the 1993 IAEA Conference on school-based and external assessments. Mauritius Examinations Syndicate.

Gronlund, M. A. (1985). Measurement and evaluation in teaching. (5th ed). Macmillan Publishing Company, New York.

Guba, E.G., Lincoln, Y. S. (1989). Fourth generation evaluation. Sage Publications, New Delhi.

Mehrens, W A & Lehmann, I. J. (1969). Measurement and evaluation in education and psychology, Holt, Rinehart and Winston.

Newton, P., Baird, J., Goldstein, H., Patrick, H., & Tymms, P. (eds). (2007). Techniques for monitoring the comparability of examination standards. Malta. Qualifications and Curriculum Authority.

Nitko, A. J. & Brookhart, S. M. (2011). Educational assessment of students. Pearson. Boston.

Pellegrino,J. W. (2009).The design of an assessment system for the Race to the Top: A Learning Sciences Perspective on Issues of Growth and Measurement. Paper presented at the Exploratory Seminar: Measurement Challenges within the Race to the Top Agenda.

Rudman, H. C. (1993). National testing or political testing: Is there a difference? Educational measurement: Issues and practice, 12(3), 5-9.

Smarter Balanced Assessment Consortium: Usability, Accessibility, and Accommodations Guidelines. (2018). Prepared with the assistance of the National Center on Educational Outcomes

Smith, M. L. & Rottenberg, C. (1992). Unintended consequences of external testing in elementary schools. Educational measurement: Issues and practice, 10(4), 7-11.

Verger, A., Lluís Parcerisa, L., & Fontdevila, C. (2018): The growth and

spread of large-scale assessments and test-based accountabilities: a political sociology of global

education reforms. Educational Review.