IMPLEMENTING QUALITY PERFORMANCE ASSESSMENT IN AGRICULTURE

BY

TRUST MBAKO MASOLE
(Senior Researcher)
Botswana Examinations Council
tmmasole@bec.co.bw or masoletm@gmail.com

It is imperative to realize that quality education is neither measured by the number of children attending school nor confined to cognitive development only. Rather it is as much the accumulation of particular values, attitudes, and skills as it is the development of cognitive ability. The acquisition of skills depends much on the instructional materials being used together with associated assessment practices. The development of assessment materials in collaboration with practitioners is an indispensable part to achieving quality education.

The purpose of this study is therefore to find out: 1. How performance assessment currently is conducted and how does this compare to international practice? 2. To what extent is the problem of performance assessment evident in secondary schools in Botswana? 3. What are the characteristics of an effective quality assurance system to ensure valid and reliable performance assessment marks? 4. How to develop quality processes and outcomes to produce valid and reliable performance assessment marks?

In developing quality assessment materials, a baseline study is conducted to define, identify and measure the extent of the problem in performance assessment. Based on the findings of the baseline study, materials of different kinds of assessment methods to reflect different kinds of achievement are iteratively developed by practitioners and other stakeholders at various stages by adopting a cyclic approach of design, evaluation and revision. The intervention is then implemented in the field and formatively evaluated. Its success is be measured by its practicality in real contexts.

Introduction

A number of countries, including developed ones, have been found to be either lagging behind in achieving quality or the quality of their education was declining, (UNESCO, 2004; Walker, 2006; Greaney & Kellaghan, 2001). Quality education as defined by UNESCO, (2004), and Ward & Dooney, (1999) is a set of processes and outcomes that are defined qualitatively. These processes and outcomes are a function of factors such as relevant curriculum, improvement in pupils' assessment methods, teaching materials, provision of physical resources, information systems, financing, management techniques, and above all teacher training (Clarke.2001; Grisay & Mählck, 2003).

Assessment is one of the crucial factors in the learning process that enhances quality (Stigggins, 2002), educationists have realized the importance of recognising not only the cognitive capability of the students, but also the manipulative and affective skills. Manipulative skills may be processes, activities, products or tasks such as debates, essays, discussions and problem solving carried out by students, (Nitko & Brookhart, 2007). Affective skills lead to the development of attitudes, collaboration, values, habits and skills, and interpersonal relations These are as vital in the future survival of the student as are the cognitive skills.

The introduction of performance assessment as a format of assessment triggered a paradigm shift in instructional process to facilitate cooperative learning and development of critical thinking skills whilst standardized achievement testing, which dominated assessment field for too long, overemphasized the low levels thinking (Linn & Gronlund, 1995). What constitutes performance assessment differs from one context to another, for the purposes of argument; performance assessment

involves a task to be done by students applying their knowledge and skills from several learning targets. A clearly defined criteria is then used to evaluate how well the student has achieved the objective (Nitko & Brookhart, 2007).

This definition implies that performance assessment can also be of paper-and-pencil provided that it has been structured in such a way that it places greater emphasis on the application of knowledge and skills in a simulated setting (Airasian, 2005; Gronlund, 2003). But Wiggins (1998) cautioned of a test that can be authentic-based but inauthentic in itself whereby the development of performance assessment tasks are perfect but the provision of the appropriate context of assessment and instruction to facilitate performance being different from that of a typical performance task.

Such problems are widespread (Tamakloe, Amedahe & Atta, 2001; Nitko, (1998) Spady, (1977) (Harlen, 1994 Torrance (1995). However, a number of authors have widely documented how performance assessment marks can be validated Tindal & Haladyna, 2002; McMillan, 2000 Airasian, 2005; Durrant, 2003). These include training teachers to assess performance tasks; approval of the schools to conduct assessment; internal and external monitoring; development of comprehensive assessment criteria; involvement of parents and students in assessment; small class sizes; using multiple assessors; reassessment of students, embedding quality in the system, development and registration of industry standards; the registration of private providers; accreditation of providers to assess standards; moderation of assessments; carrying out audit and collaborative development of exemplar assessment materials among others. Despite all these, school-based assessment is still weighted low because of the inherent problems of invalidity and unreliability.

This paper therefore aims to explicate/elucidate the process of developing performance assessment tasks in agriculture to assure quality in the assessment process leading to the production of dependable scores. In particular, it tries to address the following questions:

1.   How performance assessment currently is conducted and how does this compare to international practice?
2.   To what extent is the problem of performance assessment evident in secondary schools in Botswana?
3.   What are the characteristics of an effective quality assurance system to ensure valid and reliable performance assessment marks?
4.   How to develop quality processes and outcomes to produce valid and reliable performance assessment marks?

Conceptual Framework
Performance assessment in Botswana General Certificate of Secondary Education (BGCSE) Agriculture currently yields information of little value which is not reflective of students' capabilities. Scores generated throughout the duration of study are 'validated' through moderation by visiting schools to sample artifacts. Where there is doubt, marks are authenticated by interviewing the students. This shows the bias towards assessment of the product at the expense of processes. Such a system is embedded/characterised with quality control. However, the validity of performance assessment outcomes is dependent upon embedding quality assurance processes into

the assessment system. One such way is to develop exemplar assessment materials in collaboration with the practitioners using a Hybrid Model. The successful implementation of developed materials is a function of a quality embedded assessment system.

Conducting Performance Assessment in Botswana schools

Agriculture in Botswana Senior secondary schools is practically based as evidenced by two of its three aims enshrined in both the Teaching and Assessment syllabi (Ministry of Education, 2000; Ministry of Education 2001), namely:
1. Handling information, Application and Problem Solving
2. Practical and Investigative Skill

These contribute 70 % of the examination content. However, Paper 3, which is performance-based weighs only 20 %. Thus a student who performs well in paper in this paper is likely to obtain lower grade than a student who has high marks in other papers.

BGCSE Agriculture practical assessment comprises of practical tasks and a project just like New Zealand, whose assessment in addition covers oral performance in languages (REF). The practical tasks are assessed by the classroom teacher and his/her score is final. The assessment is guided by detailed marking criteria (Nitko & Brookhart, 2007; Smith, Smith & De Lisi, 2001) and is often done once. There is very little supervision and administrators tend to know very little or regard lowly performance assessment. The project involves problem investigation to find a solution to a real agricultural problem and then compiling a report on the findings. The score is then scaled down to 20 % of the final mark (Ministry of Education, 2000).

In Germany at primary level, the assessment procedures are largely the responsibility of teachers even for certification and selection purposes with minimal external intervention. Assessment is based on both cognitive and attitudes, behaviors and manipulative skills. Students are reassessed by a different teacher when they have not achieved and parents are closely involved. Teachers are professional assessors and mark their own students work and then scored by a second scorer. Verification is done by sampling some scripts for external scoring. France assessment is based on a combination of external examinations and continuous assessment. The 1983 Prost Report on the reform of the Lycee argued for a considerably increased measure of continuous assessment. However, Bonnet (2004).reported little progress being made in that regard.

The intention of using the classroom teacher to assess processes (Gronlund, 2003), is to ensure that those activities with ephemeral evidence (Rennert-Ariev, 2005), can also be assessed. The criteria for assessing practical tasks are delineated below:

Responsibility	the ability to resume responsibility for the task in hand, and to work from given instructions without detailed supervision and help.
Initiative	the ability to cope with problems arising on connection with the task, to see what needs to be done and take corrective action.
Technique	the ability to take practical tasks in a methodical, systematic way and to handle tools skillfully and to good effect.
Perseverance	the ability to see a task through to a successful conclusion with determination and sustained effort.
Quality	the ability to attend to detail so that the work done is well finished and well presented.

The report students produce is first scored by the classroom teacher and then externally moderated by a visiting moderator using detailed marking criteria. The moderator then discusses the students work with the teachers with the view to reconcile the marks (Hall and Harding 2002). Visiting moderation is preferred because the teacher discusses the students' work with the moderator hence this helps in validating the score given. Other examination bodies such as Mauritius (Ansun, 2000) and Pacific Senior Secondary Certificate (PSSC) also use visiting moderation PSSC applies visiting moderation to the Common Assessment Tasks (CATs) and Common Assessment Frames (CAFs) while Teacher Designed Tasks (TDTs) are statistical moderated.

Britain and South Pacific Board for Education Assessment (SPBEA), Radnor and Shaw (1995) and Broadfoot (1994), on the other hand, require schools to send samples of work after internal moderation by teachers for inspection, as a way of moderation. In Germany, there is little or no moderation on the continuous assessment awarded by the school which tends to be overgenerous in its marking while New Zealand and Sweden statistically adjusted the scores.

By either visiting the schools to interrogate the teacher and students, or getting the artifacts and teachers' marks and statistically moderate them in an attempt to verify them does not help in delivering quality education. Rather, structures should be in place to ensure that the processes leading to the production of the perfromance assessment marks are sound. The processes which led to the product are crucial to the success of an education system to produce credible human resource. This demands a paradigm shift in the production of materials used for assessment to help students' development

Characteristics of an effective QA assessment system

The characteristic of a sound assessment system have been described as being authentic and valid; encompassing the whole child; repeating observations of various patterns of behaviour; continuous over time; using a variety of methods for gathering evidence of students performance; providing an opportunity for joint observations and explanations between teachers and students, teachers and parents, and students and parents; being sensitive to individual differences, and providing means of systematic

feedback to be used in the improvement of instruction and student performance; use of detailed scoring rubrics and commitment by management (Nitko & Brookhart, 2007; Rennert-Ariev, 2005; Gronlund, 2003)

For performance criteria to be appropriately applied, to check how successful the students are in carrying out the task performance task is broken into its component parts. Performance criteria must be specific enough (Nitko & Brookhart, 2007) to focus the assessor on well defined characteristics of the performance. Once defined, the criteria permit consistent assessor's assessment of performance and consistent communication with pupils about their learning. It also enables students to accurately self-assess their work irrespective of their performance ability. These scoring rubrics could be in the form of checklists, rating scales, Likert scale or any other detailed scoring rubric specifying performance (Nitko & Brookhart, 2007; Smith, Smith & De Lisi, 2001)

A well crafted assessment should involve a variety of contexts and methods. Both quantitative and qualitative methods are used to gather evidence and evaluate quality of achievement. These include practicals, projects, portfolios, interviews, observations, and questionnaires among others. Observation of a few numbers of students is among the key methods of obtaining useful information in assessing students' skills (Harlen, 2006; Mindes, 2007).

The success of any comprehensive assessment programme depends on the quality of teachers trained to assess Studies have shown that many classroom teachers lack skills in assessment (Stiggins and Conklin, 1997; Kellagan & Greaney, 2003). Well trained teachers can easily set their own quality assessment tasks. Ravoice and Pongi, (2000) reported that in the South Pacific Board for Educational Assessment (SPBEA) system, teachers were provided with opportunities to develop some task frame, the tasks and the marking schedules for such tasks, (Teacher Designed Tasks (TDT))

Despite the rigour in applying these precautions in performance assessment, marks are still chartersed by high levels of unreliability and their validity is low. This is because we never address the core of performance assessment requirements, instead generalize the requirements. This calls for endeavours to continually find means of enhancing the dependability of performance assessment marks.

Developing an approach to producing Quality assessment processes

Approach to developing a system of producing valid and reliable marks

One way of achieving validity and reliability in performance assessment is through the use of quality assessment materials. This could be achieved by employing design-based approach whereby prototypes are iteratively developed in collaboration with practitioners. The development involves a number of phases all incorporating review/formative evaluation at the end of each phase. For the developed quality assessment materials to effectively function, they are to be implemented in a system embedded with quality assurance processes.

The approach to developing exemplar assessment materials is based on the Hybrid Model developed by blending the Generic Research Design Model (GRDM) developed by Wademan, with the Design For Six Sigma's (DFSS) model of

DMADDI. The Generic Design Research Model (GDRM) is a holistic approach, which does not emphasize isolated variables. It first, tries to put the problem in context, then iteratively designs, develops and implements the prototypes. Formative evaluation is infused in all stages of the model. The components of the Generic Design Research Model are depicted in Figure 1.
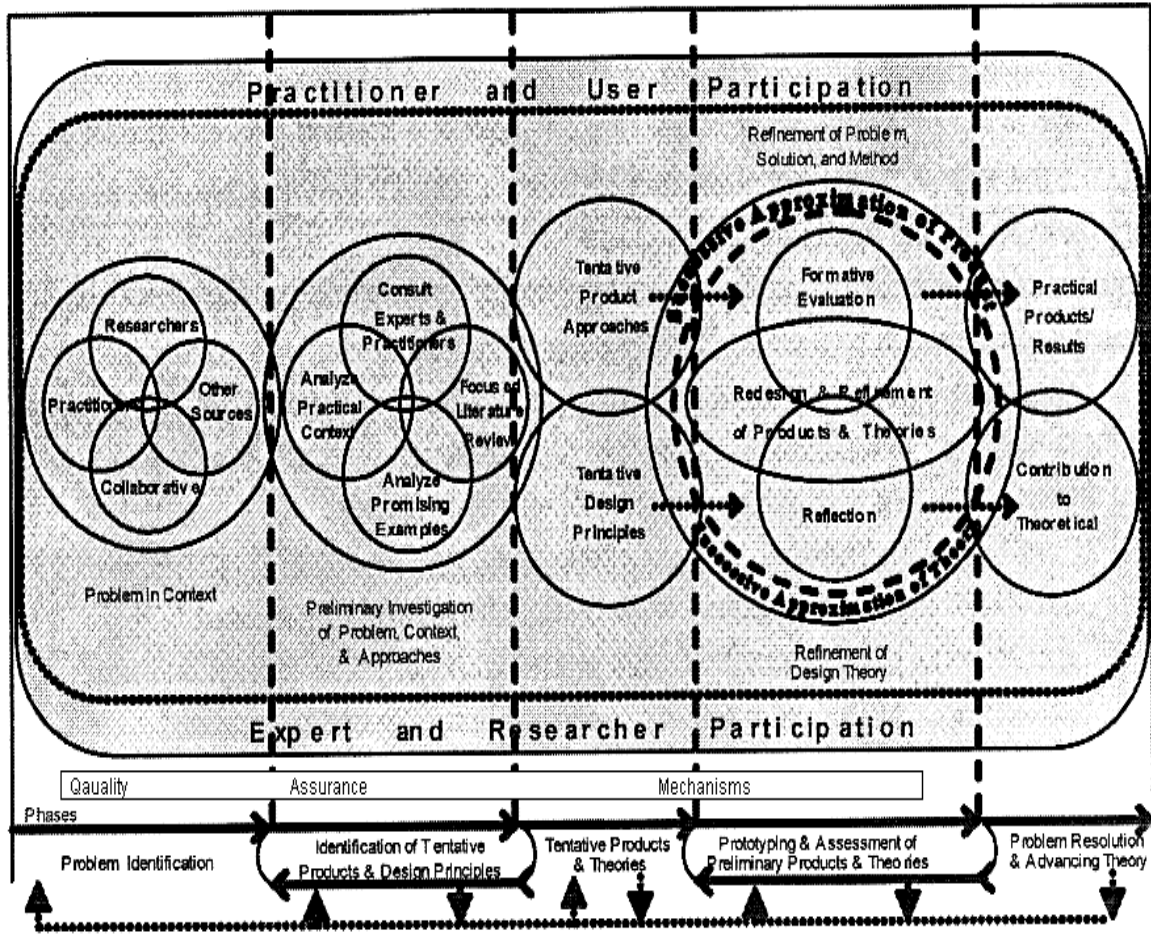

Figure 1: The Generic Design Research Model

On the other hand, Six Sigma, is a methodology that: (i) aligns core performance assessment processes with system requirements, (ii) systematically eliminates defects from existing processes, products and services and (iii) designs new processes, products and services that reliably and consistently meet system requirements (Islam, 2006). The DMADDI model has six phases of Define, Measure, Analyse, Design, Develop and Implement from which the acronym DMADDI came. These are the phases that are followed during the development of the assessment materials. The subsequent step depends on the successful execution of the preceding one, and quality control aspects infused within each step are robust and effective. It can diagrammatically be represented as shown in figure 2
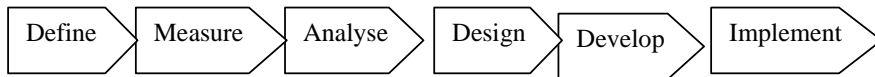
Fig 2: DMADDI mode

The GRDM is training inclined while DMADDI is business oriented. It is my strong conviction that schools should be run like business enterprises as such the development of assessment materials using some aspects of DMADDI will result in effective outcomes. The intention therefore is to borrow aspects of the DMADDI model (business inclined) and coalesce with GRDM (training inclined) to develop quality assessment materials.

Developing Quality Assessment Materials using the Hybrid Model

The first step in developing valid assessment materials based on the Hybrid Model is the problem identification. Problem identification involves conducting a baseline study which is achieved through problem definition, measurement and analysis. The baseline study determines the process capability of the system and the requirements of the practitioners, from their perspective so that interventions designed address practical problems.

After the problem has been fully defined, the next step is designing and development stage which is iteratively executed. Securing Management commitment is an essential component for the success of the design and development stages. The last stage is theory advancement whereby new theories of developing better products could be advanced. Diagrammatically, the Hybrid Model is as depicted in figure3.
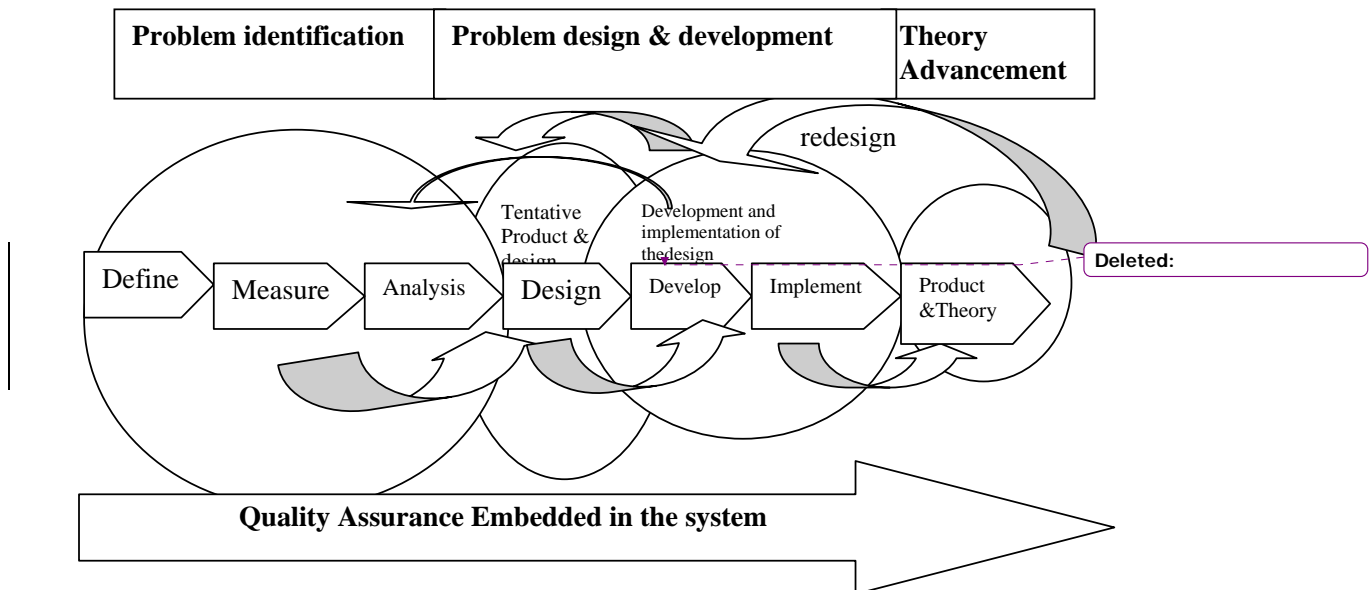
Figure 1: Research process for implementing quality assurance for valid and reliable marks

Problem identification

Problem identification is carried out through a baseline study. This constitutes defining, measuring and analyzing the problem.

Define

In order to successfully define the problem in totality, one of the first activities during this stage is to assemble a team of practitioners and stakeholders who are grounded in the subject. The purpose of define stage is to develop a clear understanding of the scope and the goal of the assessment project by all participants (Abramowich, 2005; Oakland, 2003). This stage attempts to define what is it that should change (Persse, 2006) in performance assessment. Dick, Carey & Carey (2001) and Islam (2006) pointed out that define is made possible by developing a charter which will narrate the purpose of; the project in measurable terms – how to measure success and specific parameters to be measured; and how this is going to be done; identify what is wrong and to what extent. Before moving on to measuring the problem, a review has to be carried out using a checklist to ensure that the problem had been defined comprehensively.

measure

Once the problem has been defined comprehensively, the next step is to measure the problem. The intention is to find out what targets to meet (Islam, 2006). This is achieved by identifying, prioritizing and quantifying the assessment targets. A number of Six Sigma tools could be employed to identify the assessment targets. According to Islam (2006), Goestch & David (1994), Doty (1996), and Eckes (2003), these tools include Affinity Diagrams, Pareto charts, Analytical Hierarchy Processes, among others.

On the other hand, Pairwise Comparison, Weighting, Analytical Hierarchy Process, and Multivoting (Islam, 2006) are used for prioritizing assessment requirements after identification. This, helps to identify issue(s) that hold the highest priority, and work on vital few rather than spending more time on trivial ones (Eckes, 2003). Once the ranking of the requirements is known, the next step is translating these requirements into clear targets that can be measured. Instruments are designed to translate requirements to measurable targets. These could be questionnaires, Quality Functional Deployment (QFD), interviews and assessment profile/observation checklist and rating scales. Then the review/evaluation is carried out before moving to analysis.

Analysis

Data analysis is performed using a variety of tools in order to determine the root causes of defects in performance assessment process (Islam, 2006), so as to establish an empirical basis for improving the processes (Persse, 2006). According to Abramowich, (2005) the analysis stage, is intended to identify value gaps and strengths and drill down to the root cause. The fundamental concern here is to identify what needs to be assessment. In essence, the work of constructing the

solution begins here. Before proceeding to the design stage, evaluation is conducted to ensure that all the necessary activities had been identified and analysed.

Problem design and development
Design
Based on the findings of the baseline study, the design team convenes to design and the artifact. Rainey (2005) defined product design as the process of defining and creating the product attributes fitting the intended purpose. The design stage specifies how the tasks identified under analysis stage will be assessed. A number of options are designed and the best addressing the problem is selected (Persse, 2006). The resources, budget or time constraint, as well as the conditions under which assessment will be implemented are the functions of the design (Oakland, 2003). The major factors affecting the delivery of the product are identified using a decision matrix.

One tool that is employed in designing an effective intervention is the QFD which links the needs of the end users with the design and development of the artifact (Abramowich, 2005). QFD is a practice for designing processes in response to identified needs (Goetsch & Davis, 1994). Rainey, (2005) argues that QFD ensures that artifacts produced are of the highest quality and reliability due to embedding end users' needs into the process and the iterative nature in development. Through QFD, the assessment requirements are translated into product specifications, product specifications are translated into design requirements. The design requirements are then translated into design specifications which culminated in the design criteria. The design phase culminates in the production of measurable objectives, assessment strategies, and prototype specifications. At the end of the design phase, an evaluation is instituted to determine that all steps have been accomplished before moving onto the next phase of development.

Development
Prototypes of the assessment materials are iteratively developed to match the design. Development is the actual process of transforming the inputs into an output. Persse, (2006) advises that the artifact developed has to be practically applicable with ease and compatible with the environment in which it will be implemented. The development team first identifies factors that prevent product from meeting its stated targets, for example, of yielding valid and reliable marks. Such factors are identified by the use of Ishikawa diagram/Fishbone diagram. The diagram is an analysis tool that provides a systematic way of looking at effects and the root causes that create or contribute to those effects (Basu & Wright, 2003; Eckes, 2003). Once the factors are identified, they are addressed to prevent them from causing development problems. Another tool that is used to identify possible failures or risks of a product and determine the frequency and impact of the failure is the Failure Mode Effects Analysis (FMEA). According to Abramowich (2005) and Islam, (2006), it ranks and prioritizes the possible causes of failures, as well as helps develop and implement preventative actions.

Implementation
During the implementation of the intervention performance is documented against project goals (The Design-Based Collective, 2003: Islam, 2006) such as defining the benefits realized from deployment of the intervention; describeingthe enhancements

and the performance metrics gained from the pilot; and outlining the kinds of deployment activities that could be required to implement the improvements in the existing system (Persse, 2006). Not only improvements are documented, even problems encountered, so that they are reworked immediately. At the end of implementation stage, both summative evaluation and formative evaluation are carried out.

Conclusion

The way performance assessment is conducted results in unreliable and invalid marks, consequently contributing minimal to the students' final grade. Theoretically, the conduct of performance assessment is should yield dependable marks. However, there vast problem associated with this format of assessment. One way believed to help validate performance assessment is to develop the assessment materials in collaboration with stakeholders. This should commence with the conduct of the baseline study to define the problem to be addressed, measure the problem and finally analyse to determine the root cause in performance assessment.

The design and development of the solution to the problem should then follow and the prototype produced implemented in a real classroom. This is evaluated and redesigned until the promising prototype is produced. However, the implementation of the produced assessment materials should be implemented in an environment where quality is embedded in the system.

6.0 REFERENCES

Abramowich, E. (2005). *Six Sigma for Growth: Driving Profitable Top-Line Results.* Clementi Loop: John Wiley & Sons Pty Ltd.

Airasian, P. W. (2005). *Classroom Assessment – concepts and Applications*. McGraw Hill. Boston. 5th Edition

Ansun, I., (2000). Introducing Competency-Based Assessment, in *Trends in Examinations and Perspectives in Assessment*. Red Unit. Mauritius.

Basu, R., & Wright, N. (2003). *Quality Beyond Six Sigma*. Oxford: Elsevier Ltd.

Clarke, S. A. (2001). Quality education for All: Living together, Democracy and Social Cohesion. In Byron, I. and Rozemeijer, S (eds). *Curriculum Development for Learning to live to live together.* The Caribbean Sub-Region. Final report of the Sub-regional seminar held in Havan, Cuba, 15-18 May 2001.. UNESCO. Ministry of Education of Cuba.

Broadfoot, P., (2004). Approaches to quality assurance and quality control in six Countries. In. *Enhancing Quality in Assessment*. London: Paul Chapman Publishing Ltd.

Dick, W., Carey, L., & Carey, J. O. (2001). *The Systematic Design of Instruction.* New York: Longman.

Doty, L. A., (1996). *Statistical Process Control*. 2nd Edition New York.. Industrial Press Inc.

Durant, E. (2003). Moderation Handbook. Wellington: JITO

Eckes, G (2003). *Six Sigma for Everyone*. New Jersey**:** John Wiley & Sons, Inc. Remember, people are rarely the root cause of a problem in a process.

Goetsch D. L., and Davis. S. B., (1994). *Introduction to total Quality: Quality management for Production Processing and Services.* 2nd edition. New Jersey: Prentince-Hall, Inc.

Grisay, A., & Mahlck, L. (1991). *The quality of education in developing countries: A review of some research studies and policy documents.* In Kellagan, T., & Greaney, V. (2003). Monitoring Performance: Assessment and Examinations in Africa. Paris: ADEA

Gronlund , N. E., (2003). *Assessment of student Achievement* , 7th Edition. Boston.

Harlen, W. (1994). *Enhancing Quality in Assessment.* Paul Chapman Publishing Ltd: London.

Harlen, W. (2006). Teaching, learning & assessing science 5-12. SAGE Publications Ltd: London.

Islam, K. A., (2006). *Designing and measuring: The 6 Sigma way. Pfeiffer, An Imprint of Willey: San Francisco*

Kellagan, T., and Greaney, V. (2001) *Using Assessment to improve the quality of education.* UNESCO. Paris.

Kellagan, T., and Greaney, V. (2003). *Monitoring Performance. Assessment and Examinations in Africa..*

Ministry of Education, (2000). *Botswana General Certificate of Secondary Education Teaching Syllabus: Agriculture.* Government Printer. Gaborone.

Ministry of Education, (2001). *Botswana General Certificate of Secondary Education Agriculture Assessment Syllabus.* Government Printer. Gaborone.

Mindes, G. (2007). Assessing Young Children. New Jersey: Merrill Prentice Hall.

Nitko, A. J., & Brookhart, S. M., (2007). Educational Assessment of Students. New Jersey: Merrill Prentice Hall.

Oakland, J. S. (2003). *Statistical Process Control.* Oxford: Butterworth Heinemann.

Persse, J. R. (2006). P*rocess Improvement Essentials.* Sebastopol: O'Reilley Media, Inc.

Rainey, D. (2005). *Product Innovation: Leading change through integrated product development.* New York: Cambridge University Press.

Radnor, H., & Shaw. K., (1995). 'Developing a Collaborative Approach to Moderation, in H. Torrance, (ed), *Evaluating Authentic Assessment.* Buckingham: Open University Press.

Rennert-Ariev, P., (2005). Theoretical model for the authentic assessment of teaching, Practical Assessment, Research and Evaluation. *A Peer-reviewed electronic journal.* 10 (2)

Smith, J. K., Smith, L. F., & De Lisi, R. (2001). Natural classroom Assessment: Designing Seamless Instruction and Assessment. California: Corwin Press Inc.

Spady, W. G., (1977). *Competency-based education*: A bandwagon in search of a definition. *Educational Research*, 6 (1) 9-14.

Stiggins, R. J. (1997). Edition. *Student-Centred Classroom Assessment.* 2nd. New Jersey. Prentice-Hall, Inc.

Stiggins, R. J. (2002). *Where is our assessment future and how can we get there from here?* In Lissitz and Schafer, Assessment in Educational Reform: both means and ends Allyn and Bacon:Boston

Tamakloe, E. K., Amedahe, F. K. & Atta, E. T. (2005). *Principles and methods of teaching. Accra,* Ghana Universities Press.

The Design-Based Collective, (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, 32(1), 5-8

Torrance. H., (1995). *Evaluating Authentic Assessment*. Buckingham: Open University Press.

UNESCO. (2004). EFA *Global Monitoring Report 2005. Education For All: The Quality Imperative*. Paris: UNESCO.

Ward, N & Dooney, T. (1999). *Irish Education.* Dublin: Oak Tree Press.

Wiggins, G.P., (1998). A true test: toward more authentic and equitable assessment" *Phi Delta Kappan*, 70 (9) 703-13.

Walker, D. (2006). *Toward Productive Design Studies.* In J. Van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (eds). *Educational Design Research* (pp 8 – 13). London: Routledge.