



Journal of Educational Assessment in Africa (Vol. 12, 2017)

A Publication of the
Association for Educational
Assessment in Africa
(AEAA)

Purpose of the Journal

The Journal of Educational Assessment in Africa is published yearly by the Association for Education Assessment in Africa (AEAA). The main objective of the journal is to provide information to enhance research and knowledge so as to improve educational assessment in Africa. The journal publishes manuscripts that contribute to improvement of educational assessment.

Copyright

Copyright © 2017 by the AEAA. The Journal of Educational Assessment in Africa (JEAA) is copyrighted. No part of this publication may be reproduced, stored, transmitted or disseminated in any form or by any means without prior written permission from the Association for educational Assessment in Africa (AEAA). Where information or data from the Journal is used, the Journal must be duly acknowledged.

Disclaimer

All papers in the Journal reflect views of the authors and not necessarily those of the Association for Educational Assessment in African (EAA).

Contributors to the Journal

1. **Moges Logaw (Dr.)**, mogeslog@yahoo.com, EFDR Meles Zenawi Leadership Academy, Ethiopia
2. **Prof. Kafui Etsey & Mr. Abraham Yeboah**, kafuietsey@yahoo.com University of Cape Coast, Ghana
3. **Amakire (Dr.) (Mrs.) & Hager Atisi Eremina**, amakiriatisi@yahoo.com, GNATIUS Ajuru University of Education, Rivers State, Nigeria
4. **Dennis Zami Atibuni (Dr.) & Prof. David Kani Olema (Ph.D)**, zamidennis79@gmail.com, Busitema University, Uganda
5. **Betty Nalukenge Habaasa**, Betty.Nalukenge@uneb.ac.ug/, bnnalukenge@gmail.com, Uganda National Examinations Board, Uganda
6. **Denis Ndasi Mofor, John Teneng Awa (PhD) & Kenneth Ngu Foncha (PhD)**, denisndasimofor@gmail.com, ndipingwi@vgmail.com, kfoncha@yahoo.com, GCE Board. Cameroon
7. **Dan Nokrach Odongo**, Dan.Odongo@uneb.ac.ug; danodongo22@yahoo.co.uk, Uganda National Examinations Board
8. **Felix Akuffo-Badoo**, FAKUFFO-BADOO@waecgh.org, The West African Examinations Council, Ghana
9. **Kagoro Thomas**, Thomas.Kagoro@uneb.ac.ug, Uganda National Examinations Board, Uganda
10. **Doreen Kawira Mbwiri** dkawira@knec.ac.ke, Kenya National Examinations Council, Kenya
11. **James Turyatamba**, James.Turyatamba@uneb.ac.ug, Uganda National Examinations Board, Uganda
12. **Ikechukwu E. Anyanwu (Dr.)**, **Felix I. W. Onwuakpa(Dr.) & Innocent Uche Ezenwanne (Dr.)**, aikmanuel@yahoo.com, felixonwuakpa@gmail.com & ucheezen@yahoo.com, National Examinations Council (NECO), Nigeria
13. **Florence Obong Capuc & David Weerhe**, obongfr@gmail.com, David.Weerhe@uneb.ac.ug, Uganda National Examinations Board, Uganda
14. **Helyn Kim(Ph.D) & Esther Care(Ph.D)**, (hkim@brookings.edu), (ecare@brookings.edu), The Brookings Institution, United States
15. **Boscow Mubita**, bmubita@exams-counci.org.zm/kumbonja63@gmail.com, Zambia
16. **Ramokoena Camilla' Matseko**, favour.ramokoena@gmail.com, Examinations Council of Lesotho, Lesotho
17. **Monono Ekema Humphrey (PhD)**, hemonono54@yahoo.com, **John Teneng Awa (PhD)**, ndipingwi@yahoo.com, **Kenneth NguFoncha (PhD)**, kfoncha@yahoo.com
18. **Olanrewaju Ojo Emupenne**, emupenne@gmail.com; **Skype:** lanre.emupenne, The West African Examinations Council, Yaba, Lagos, Nigeria.
19. **John Mary Vianney Mitana, Anthony Mugagga Muwagga, Mauro Giacomazzi & Omala Saint Kizito (PhD)**, Makerere University. Uganda
20. **Mwale Frederick**, frederick_mwale@rocketmail.com, Zimbabwe School Examinations Council
21. **Robert G. Dery (Rev.)** rgdery@waecgh.org, The West African Examinations Council, Accra-Ghana

22. **Prof. Henry O. Owolabi. Prof. Adams O. U. Onuka, Mayowa O. Ogunjimi (Dr.), Yusuf Mustapha (Dr.). Lukman Adaramaja(Dr.) & Dorcas Daramola (Dr.)** henrywolabi2000@yahoo.com. adamonuka@yahoo.com, ogunjimimayowa@gmail.com – bbmusty@yahoo.com – daramaja4real@yahoo.com – immaculatethabitha@yahoo.com, Association of Educational Researchers and Evaluators of Nigeria
23. **David Kani Olema (PhD)**, zamidennis79@gmail.com, Busitema University – Nagongera Campus, Uganda
24. **Uruemu Charity Ogbemor-Kigho& Adams. O.U. Onuka, uruemukigh@gmail.com.** ogbeborremu@yahoo.com, ada.otuoze@gmail.com/adamonuka@yahoo.com , Institute of Education, University of Ibadan, Ibadan Nigeria
25. **Prof. (Mrs.) Ifeoma Mercy Isiuogo-Abanihe, Ibidapo Olabode Ilesanmi & Irene Ovekairi Iuobe (Mrs.)**, Ifyabanihe@gmail.com, bodeibidapo@yahoo.com, ireneovekairi@gmail.com National Business and Technical Examinations Board (NABTEB), Benin City, Nigeria
26. **Banji Milumbe Shakubanza & Mike Hatembo**, bms Shakubanza@exams-council.org.zm; mhatembo@exams-council.org.zm, Examinations Council of Zambia, Zambia
27. Dr. Saverio Pido & Ruth Ndagire Ssenyonjo, Uganda National Examinations Board, Uganda
28. **Maria Atiamuga**, matiamuga@knec.ac.ke, The Kenya National Examinations Council, Nairobi. Kenya.

Foreword

The Association for Educational Assessment in Africa (AEAA) presents yet another issue of the *Journal of Educational Assessment in Africa (JEAA)*. This issue is the 12th volume and is a publication of papers that were presented at the 35th AEAA Conference which was held in Kampala Uganda, from the 7th – 11th August 2017. The theme of the conference: “Enhancing Efficiency and Effectiveness in Educational Assessment in an Era of Rapid Change” was informed by the need to carefully reconsider assessment practice in a period witnessing unprecedented challenges not only on how students learn, the environment they do so, but urgency to recast assessment priorities and methods that would yield cost effective and fast results given resource constraints. Issues related to these included how curriculum reforms impact assessment practice; novel methods for job placement, standard setting and special considerations for learners with learning impairment (Special Needs Education). All those facets require harnessing and deployment of ICTs in assessment not only to capture data and retrieve it, but use it in other novel situations such as computing complex assessment data and allowing expansive interpretation and use. Whether assessment of soft skills should be the role of the school or the assessment bodies was included in the debate and discussion thereof proposed how this might be done. Examination malpractice was discussed with special focus on innovative ways of combating it.

Articles published in this volume addressed the main theme under the following sub themes:

- Using technology to enhance efficiency and effectiveness in educational assessment in an era of rapid change.
- Relevance of aptitude tests in selection and placement
- Standard setting procedures in educational assessment
- Combating examination malpractice
- “Assessment of learning” or “Assessment for learning”
- Enhancing the quality of practical skills assessment
- Curriculum reform and educational assessment
- Integrating soft skills assessment in public examinations

Readers of this journal are welcome to yet another quality product, sharing assessment issues with scholars from Africa and the rest of the world.

Editorial Team

| Title | Name | Institution |
|--------------|-----------------------------|------------------------------------|
| Prof. | Joseph Oonyu | Makerere University, Kampala |
| Prof. | Betty Akullu Ezati | Makerere University, Kampala |
| Dr. | David Kani Olema | Busitema University |
| Dr. | Proscovia Ssentamu Namubiru | Uganda Management Institute |
| Dr. | Dennis Atibuni Zami | Busitema University |
| Mr. | James Turyatamba | Uganda National Examinations Board |
| Ms. | Florence Obong Capuc | Uganda National Examinations Board |
| Mr. | Thomas Kagoro | Uganda National Examinations Board |
| Mr. | Dan Kyagaba | Uganda National Examinations Board |
| Mr. | David Weerhe | Uganda National Examinations Board |

Table of Contents

Contents
Table of Contents vi

SUB THEME A: ASSESSMENT OF LEARNING OR ASSESSMENT FOR LEARNING 1

Practices of Student Assessment in Competency based Modular Instruction for Quality Education in Ethiopia: The Case of College of Education in Amhara Region Universities 1

Improving Assessment of Learning in Mathematics through Assessment as Learning 11

Subject interest and Biology Achievement of Senior Secondary Students in Rivers State: Effect of Assessment for Learning Strategies 21

Incorporating Assessment of Learning in Assessment for Learning: Shifting from Foundational Competences to Practical and Reflexive Competences in Learning Institutions 32

Teaching Approaches, Interest and Performance of Students in Physics at Uganda Certificate of Education: The Relevance of Constructivism Teaching Learning Theory 43

An Assessment of Inter-Rater Reliability Concepts: A Case of Examiners of Mathematics and English at the Uganda Certificate of Education 55

SUB THEME B: COMBATING EXAMINATION MALPRACTICE..... 65

Examination Malpractice as a Consequence of Media Publication of League Tables: The Case of the Cameroon General Certificate of Education Examinations 65

Impact of District Performance League Tables on Performance at the Basic Education Certificate Examination in Ghana 74

Effectiveness of Measures by Uganda National Examinations Board to Control Malpractice: Perspectives of Secondary School Students 84

Level of Effectiveness of Scout in Combating Examinations Malpractice: The Case for Uganda National Examinations Board..... 99

SUB THEME C: CURRICULUM REFORM AND EDUCATIONALASSESSMENT 114

Home based and school factors that influence pupil academic achievement at Class 3 level in Kenya. 114

Assessment Reforms at the Lower Secondary Level of Education in Uganda: From Norm Referenced to Criterion-Referenced Testing..... 123

| | |
|---|------------|
| SUB THEME D: ENHANCING THE QUALITY OF PRACTICAL SKILLS ASSESSMENT | 132 |
| Correlation between Practical and Theory Paper Performance in Home Science and Chemistry KCSE Subjects | 132 |
| Improving the Quality of Science Practical Skills Assessment in Secondary Schools in Nigeria..... | 145 |
| Benefits, Challenges and Procedures of Introducing Oral Skills Assessment in English Language at Uganda Certificate of Education..... | 154 |
| Considering Authenticity of Skills Assessments..... | 165 |
| Challenges of Setting and Conducting Practical Skills Assessment in Musical Arts Education in Zambia..... | 176 |
| Investigating the Appropriateness of the use of Oral Communication Competencies as a Practical Component in the Assessment of Language: The Case of Ordinary Level Special Bilingual Education (SBE) French in Cameroon. | 196 |
| Authentic Assessment of Candidates with Profound Hearing Impairment in Practical Subjects at Ordinary Level: Is Zimbabwe getting it Right? | 206 |
| SUB THEME E: INTEGRATING SOFT SKILLS ASSESSMENT IN PUBLIC EXAMINATIONS | 216 |
| Perceptions on Integration of Soft Skills Assessment in Public Examinations: The Case of Lesotho Basic Education | 216 |
| Assessment of Soft-Skills in Ugandan Secondary Schools: The Development and Use of Assessment Tools | 236 |
| The explicit assessment of skills for the 21 st century - a case study by Cambridge International Examinations..... | 247 |
| SUB THEME F: RELEVANCE OF APTITUDE TESTS IN SELECTION AND PLACEMENT ... | 257 |
| The Perceptions of Users of Aptitude Tests and its Suitability for Recruitment of Graduates in Ghana | 257 |
| Options functionality of English Language tests for Candidates' Selection and Placement into Tertiary Institutions in Nigeria: A pilot study | 265 |
| Baseline Survey Report on Aptitude Testing | 273 |
| SUB THEME G: STANDARD SETTING PROCEDURES IN EDUCATIONAL ASSESSMENT | 284 |
| Standard Setting Procedures: A Challenge for Modern Educational Assessment | 284 |

| | |
|---|------------|
| SUB THEME H: USING TECHNOLOGY TO ENHANCE EFFICIENCY AND EFFECTIVENESS IN EDUCATIONAL ASSESSMENT | 292 |
| Item Calibration for School-based Assessment in Delta State Nigeria | 292 |
| Parameterization of Test Using Item Response Theory (IRT): A Case Study of English Language Essay Test of the National Business and Technical Examinations Board (NABTEB) | 305 |
| Using an Online Query Tracking System to Track Examination Queries in Zambia | 319 |

SUB THEME A: ASSESSMENT OF LEARNING OR ASSESSMENT FOR LEARNING

Practices of Student Assessment in Competency based Modular Instruction for Quality Education in Ethiopia: The Case of College of Education in Amhara Region Universities

Moges Logaw (Dr.)

mogeslog@yahoo.com+251 914710293

Ethiopian Federal Democratic Republic MelesZenawi Leadership Academy, Addis Ababa (Ethiopia)

Abstract

This research investigated the practices of student assessment in competency based modular instruction in College of Education within Higher Education Institutes in Amhara region. The research looked into the training program conducted by Colleges of Education within the Universities. Relevant literature on practices of student assessment and competency based modular instruction were also reviewed. In the research different modular theoretical assumptions were examined and used in studying the practice. In Ethiopia, there are thirty-three public universities. Out of these figures by the use of purposive and stratified sampling techniques three colleges of education from the seven universities within the Amhara Region have been selected. Namely, Woldia University from the third generation, Wollo University from the second generation and Bahir Dar University from the first generation. After selecting Universities and issues that are related with student assessment, data was gathered from each College of education using questionnaire and an in-depth interview. Data analysis was made by using both chi-square and t-test for quantitative data and use of interpretational and reflective analysis for qualitative data. The findings showed that the practice of student assessment was not done as intended. Here, the practice of continuous assessment seemed to be weak in supporting the teaching learning process. Assessments were not made on time in according to the learner pace. The use of formative assessment was found less and it also highly relied on cognitive domains. Eventually, it was concluded that the Practices of student assessment in competency based modular instruction in the Colleges of Education were found weak and required consideration both from the government and institutions.

Key words: Quality Education, competency based modular instruction, assessment, and modularization

BACKGROUND

For instruction and learning to become effective, the teacher must be concerned with: the quality of education specifically quality of instruction, which means that instruction, must make sense to the students; the appropriate strategy to use; the incentive to the students for them to learn; and sufficient time for learning to occur (Bedaure , 2012). Teachers must adapt instruction to the students' level of knowledge and development, motivate them to learn, and manage their behavior.

One instructional strategy, which has recently gained popularity, is competence based modular instruction. Of all the various systems of individualized instruction,

modular approach is one of the recently used and combines many advantages of a number of separate instructional innovations, such as performance objectives, self-pacing, and frequent feedback (Hand et al., 2000).

Competence based modular instruction demands teachers to employ progressive assessment, use diversified teaching methods, carefully plan the lessons etc. (MOE, 2012). Here, teachers could have a chance to monitor the progress of students learning and the whole provision of the module and allows them to look in to the strength and pitfalls of the instructional process and take remedies for further improvement.

According to MOE (2012) assessment, competency is not only based on knowledge and attitude but primarily on the actual demonstration of the competency. Occupational standards or unit competency standards should be used as the basis for assessing achievement and students/trainees should be aware of them. Systems that embrace change through data generation, use and self-assessment are more likely to offer quality education to students (Glasser, 1990).

In Ethiopia, Higher Education Institutions have embarked on major reforms since the last two decades. Of all important tools of the reform, Business Process Reengineering (BPR) and Teacher Education System Over all (TESO) were the most known. In the reengineering of the teaching learning core process, modularization was proposed as a best way for the implementation of curricula and the production of competent graduates.

Similarly, techniques of assessment seem to be changing from the traditional way of students' evaluation to the competency based approach.

In this regard students are expected to be evaluated in all areas while they are practicing both in the institute and work place.

Despite the government's efforts to improve, the quality of teachers, and education in general continues to decline. Perhaps the most powerful indictment of the seriousness of declining quality in specified areas can be seen in the preparation of teachers in all levels.

Generally, due to the rapid expansion rate of HEIs, many assessment challenges are left untouched. As a result, many poorly prepared graduates were produced.

Based on the stated critical professional development limitations in enhancing quality education in HEIs in Ethiopia, this study was planned to examine the status of modularization with a special focus on practices of student assessment in competency based modular instruction in college of education within the HEIs in enhancing quality education

Objective of the Study

The general objective of the study was to examine the practices of student assessment in competency based modular instruction in colleges of education within the HEIs in enhancing quality education. The specific objectives of the study include:

1. To scan whether or not the practice of continuous assessment technique is actually on progress in measuring students competency.
2. To examine whether or not assessments were made on time in according to the learner pace.
3. To examine to what extent learning outcome and competencies based instructions are being considered in the assessment mechanism.
4. To evaluate the involvement of students in the process of assessment.
5. To examine the status of feedback given by the teacher.

REVIEW OF RELATED LITERATURE

Quality of Teaching and Learning

In improving quality of education, many countries increasingly focus on understanding complex interactions that take place at the school, classroom, and community levels as the primary engines of quality and as a way of engaging local actors to address the frequently weak link between policy and practice (Farrell, 2002). Of the factors that contribute to education quality at the local level, quality of teaching is recognized as the key, the factor without which other quality inputs are unlikely to be successful (USAID, 2006). Focusing on teachers' and principals' roles on quality is particularly important because they are the professionals primarily responsible for interpreting and implementing the constructivist, active-learning, and student-centered pedagogical approaches to improving education quality that underlie the reform policies of many countries. According to USAID (2006), quality education fell into three clear categories: input, process, and output factors. From the input side, it depends on teaching and learning techniques and assessment techniques such as active learning, student centered, competence based modular approach and continuous assessment techniques.

Virtually all countries, however, include two key elements as the basis of quality: students' cognitive learning (which is what achievement tests usually measure) and their social, creative, inter-personal, and emotional development.

One of the major indicators of quality education is cognitive learning that is the main explicit objective of most education systems, although there is a wide disagreement on what to measure as cognitive learning and how to measure it. The social, creative, and emotional development is rarely assessed in a significant way (Leu, 2005 and UNESCO, 2004).

Competence based modular instruction

The development of knowledge is taking place in more diverse context. Gibbons (1998) speaks about two modes of knowledge production. Mode 1 production refers to knowledge of the discipline –based type, typically produced in the classical universities. Mode 2 knowledge developments is the production of knowledge in the context of application, that is, it arises in the process of solving particular complex problems in collaborative trans-disciplinary teams and partnership, situated both within and outside higher education institutions.

Even though competence based instruction is more exercised in technical and vocational education, especially at secondary level, recently the competence-based approach has been found as well in higher education.

In developing countries many higher education institutions experience a growing gap between their curricula and the demands from society, business and industry for a more flexible workforce with high skills (competencies) in problem solving, team work and project management. They have mostly kept to the traditional functions and objectives of Western Universities (Maamouri & Wagner, 2001).

A way to conceptualize the relationship between education and the world of work is through competence –based modular instruction. Acquiring and developing competence is more than learning a set of skills. A

common term describing the acquisition and development of competence is competence –based modular instruction, where training is more associated with the mastering of skills (Kouwenhoven, 2009).

According to Kouwenhoven, a competency is conceptualized in the model as the capability to choose and use and apply an integrated combination of knowledge, skill and attitudes with the intention to realize a task. In a more elaborated way, the definition of competence is defined as follows:

Competence is the capability of a person or an organization to reach specific achievements. Personal competencies comprise: integrated performance oriented capabilities, which consist of the clusters of knowledge structure and also cognitive, interactive, affective and where necessary psychomotor capabilities, attitudes and values, which are conditional for carrying out tasks, solving problems and more generally, effectively functioning in a certain profession, organization, position or role. (Mulder, 2001:9)

Assessment in Competency Based Modular Instruction

Assessment is the process by which the University is able to confirm that a student has achieved the learning outcomes and academic standards for the module and/or award for the program for which he or she is registered (Boud and Falchikov, 2007).

Assessment is a generic term for a set of processes that measure the outcomes of students' learning, in terms of knowledge acquired, understanding developed, and skills gained. It serves many purposes. Assessment provides the means by which students are graded, passed or failed. It provides the basis for decisions on whether a student is ready to proceed, to qualify for an award or to demonstrate competence to practice. It enables students to obtain feedback on their learning and helps them improve their performance. It enables staff to evaluate the effectiveness of their teaching (Boud and Falchikov, 2007).

According to Meyer et al (2009) there are two distinct purposes of assessment: assessment of learning, and

assessment for learning. Assessment of learning involves measuring what and how much students have learned, tied to specific learning outcomes which are themselves derived from the graduate profile. Assessment for learning is focused on using assessments to help students improve and move forward in their learning. This kind of assessment is equally important in giving students the information they need to guide and promote their own learning so that they can meet the intended outcomes. Assessment for learning requires that academic staff assess in a manner that will allow them to identify what kinds of improvements are needed and communicate this information to students.

From the module design perspective, UCD (2011) has identified the following 6 principles that will assist in designing learning experience.

1. Allow students, where possible, have opportunity for regular, low stakes assessment with opportunity for feedback on their progress
2. Develop students' opportunities for in-class self and/or peer review of their learning against assessment criteria.
3. Allow students multiple opportunities for well-structured and supported collaborative learning and its assessment (peer and group-work, project work)
4. Consider the redesign of the learning sequence of module learning activities in an efficient and

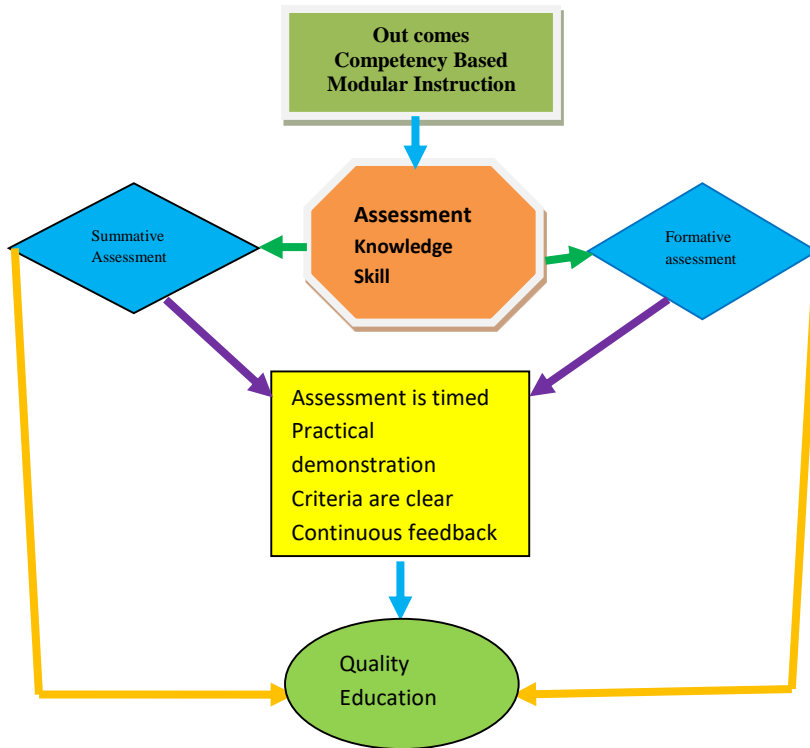
effective manner, including the related blended learning opportunities.

5. Introduce more active/task-based learning which uses more authentic assessments (i.e. subject/discipline identity)
6. Consider the student work-load demands within the module, as well as in parallel modules.

In conclusion, the need to improve the quality of education is the basic agenda of contemporary writings, political debate and media. In improving quality of education specifically in higher education, many countries mainly focus on three categories namely; input, process, and output factors. Among the most important quality education categories, an effective teaching and learning process with a competence based assessment techniques like active learning, student centered, competence based modular approach and continuous assessment are basic.

In competency based modular instruction, the mechanism of student assessment has to be aligned with the two distinct purposes: assessment of learning and assessment for learning. In doing so an assessment opportunity feedback, self and/or peer review of student learning against assessment criteria, collaborative learning and its assessment with blended learning opportunities, active/task-based learning which uses more authentic assessments and consideration of the student work-load are more important.

Conceptual frame work



From this map one can easily understand that in a competence based instruction both formative and summative evaluations are basic. Assessment done on knowledge, skill and attitude on time based on clear criteria and practice can bring together quality of education.

Design of the Study

The design which was preferred in this study, was both survey and descriptive study. This design is preferred to investigate issues that are related to the practices of student assessment in competency based modular instruction for Quality Education within the HEIs.

Research Method

This part includes a brief description of the research design, selection of respondents, the data collection instruments, the data collection procedures and the data analysis procedures of the study.

Relationship among theory and practice in student assessment was scanned through the use of survey method. On the other hand, a qualitative approach was been done as an important method in collecting the appropriate information. An in-depth investigation and empirical inquiry of events was employed to explore causation in order to identify the underlying impacts within its real-life context.

Selection of Respondents

In Ethiopia, there are thirty-three public universities. Out of these figures by the use of purposive and stratified sampling techniques three colleges of education from the seven universities (Gonder, Bahir Dar, Debreworkos, wollo, Debrebirhan, Woldia and Debre Tabor university) within the Amhara Region were selected. The main criterion for the strata was year of establishment. Here, organizational characteristics of the public universities made by ministry of education were used as a basis for selecting the sampled universities.

Therefore, from the seven public universities within the region, Woldia university from the new universities (third generation), and Wollo university from the young universities (second generation) and Bahir Dar university from the old university (first generation) were selected and included in the study.

There were two main reasons for selecting these universities. First, in Amhara region, each college within the universities was exercising the modular mode of delivery based on the module guide given from the Federal Ministry of Education. Secondly, this region has all types of Universities generation and is also easily accessible to the researcher as compared to the other regions in Ethiopia.

Once the departments from each college of education were identified, 40 students from year 2 who were in the pipe line of modular training and 15 teachers were selected from each college of education using random simple technique for the purpose of survey study. Thus, the total of teachers and students that participated in survey study were 45 and 120 respectively. Since the number of departments within each college is small, all departments were considered in the study.

Selection of informants in qualitative research is different from quantitative research. In qualitative research, the issue of representation is not the central point. Thus, sampling was not used for representativeness and generalization. Rather, selection was made purposefully. Therefore, in order to get the right information, I

contacted and discussed with different teachers and students who had a good knowledge about the issue. Thus, selection of respondents in this study was a process of actively seeking those information- rich persons.

Hence, for the purpose of interview, five students and three teachers from each college who involved in modular training were purposefully selected. The selection of teachers was based on their long time experience of consultancy and teaching in the Ethiopian higher education. Finally for the purpose of focus group discussion, six teachers from each college were selected using purposive sampling technique and discussion were made accordingly.

Data collection Instruments

For collecting the data, two questionnaires were developed: the teacher questionnaire and the student questionnaire. In order to determine questions and issues to be asked in questionnaires, a comprehensive literature review was made and in-depth interviews were conducted in a face-to-face situation with students and teachers. All the interviews were conducted in Amharic to make communication easier. All the transcribed materials were carefully translated from Amharic to English.

Moreover, the questions were asked to 18 participants in three focus groups. All members of the focus group were given background information about the study, the aim of the focus group and instructions about the procedures that were followed. Within the first part of the questionnaires questions on demographic information were included and then five point Likert type items concerning student learning assessment and feedback were added.

Method of data analysis and interpretation

In the study, in order to obtain reliable information on the Practices of student assessment in competency based modular instruction for quality education, both quantitative and qualitative analysis of the data was employed. Furthermore, chi-square and t-test were employed to examine the mean difference and level of significance among the respondents. This was supported

by the statistical package for social sciences (SPSS) computer program.

In analyzing qualitative data both interpretational and reflective analysis were used as the most important approach in order to create the necessary categories and to reflect my personal views on the bases of the information obtained.

Continuous assessment is a formative type of evaluation procedure concerned with finding out, in a systematic manner, the over-all gains that a student has made in terms of knowledge, attitudes and skills after a given set of learning experience. The results obtained from continuous assessment can be used to identify the students' weak areas so that teachers can give them special support in those areas. However, as many scholars agreed, the quality of those practices is deficient in many ways.

When we come to the type of assessment made in our studied Universities, it is not as such differed from the above suggestion. In a modular practice most teachers were not using regularly the practice of continuous assessment technique. Although, in teachers' response, 38.9% of them mentioned that the use of continuous assessment was in a right truck, majority of them (60.1%) did not agree with the application of continuous assessment. Similarly, except 3.5% of students' respondent, agreed with the use of continuous assessment while majority of them (96.5%) were shown their disagreement.

In terms of students' skill measurement, teachers and students had different perceptions. In this area, except 27.8% of teachers, great majority (72.2%) replied in the opposite direction saying that it was appropriate. On the contrary 92.9% of students' respondents were assured that assessment techniques that have been made at the time of modular instruction were not in apposition of measuring their skills.

Data Presentation, Interpretation and Discussion

Data Presentation and Interpretation

In this research, findings that are related with assessment techniques were clearly presented and explained in according to their priorities. This was done to answer the research objective posed in chapter one. The researcher tried to investigate and analyze in detail the basic issues of assessment practice from the two groups of respondents namely teachers and students as follows.

When we look at the assessments techniques made in terms of time and learner pace, different responses were obtained from the two groups of respondents. In this regard though, teachers and students do have significant mean differences ($p < 0.05$) on the extent of emphasis of assessment on time, both respondents disagreed (85.3%) on the issue of assessment being timed according to the pace of learning. Teachers who said assessment was not made with the expected time line with the students pace were 55.6%. While students who agreed with this idea were 94.8%.

Evaluation at the end of the module was not a significant problem in all Universities. Here, significant means difference was not also observed from the two groups of respondents. In this aspect except 14.7 % of the two groups of respondents, almost 85.3% of them agreed that evaluation had been done at the end of each module.

Similarly, in all our studied Universities, the implementation of competence based assessment was getting low. Almost 87% of the respondents from the two groups were agreed with this idea. Here, students were not seen while they assessed their practical demonstration. Except 38.9% of teachers' respondents, great majority of them (61.1%) disagreed with assessment made on practical demonstration. On the other side, results obtained from the students were much more different from the teachers. In their response almost 94% of them disagreed that assessment basically was made with practical demonstration of competence.

It is crucial for teachers to share assessment criteria with learners to promote the chances of learning taking place.

The assessment criteria should be clear and should not be added after learners have generated the work for a given task. It is therefore vital that all learners in a group understand what they are trying to achieve in a given task and why they are doing it. However, in our Universities, experience on the clarification of criteria for students on assessment has not yet matured. Students are assessed without a clear picture of what ought to be assessed. On such an issue, 93.3% of the two groups of respondents agreed that clarification of performance assessment criteria for learners was less.

As results indicated in the study, measuring all level of domains within the practice of modular instruction in all studied Universities was not done as needed. In assessing student's performance in all domains, except 15.8 % of students said "I am not sure" and 7% of respondents and "I agree", great majority of the two groups of respondents (76%) showed their disagreement. Moreover, as evidences has been obtained from the in-depth interview and focused group discussion, most teachers usually relay on assessment which is highly attached with cognitive domain.

The importance of feedback in improving the teaching learning process is undeniable. However, literature on feedback reveals that students are often dissatisfied with the feedback they receive, in terms of lacking specific advice to improve.

Similarly, in our Universities, the magnitude of the problem is high. In this case (92.9%) of students from the studied Universities reported that the experience of immediate feedback was weak and as a result students were not as such beneficiaries. Moreover, in the in-depth interview and focused group discussion both groups assured that immediate feedback was not done at regular bases.

Discussion

In a competence based modular instruction, varieties of assessment strategies are usually adopted as far as modularization is in place. Typically, assessment should

be integrated in to the instructional package to ensure the mastery of the outcomes. However, in our context in all studied Universities, assessments were not made on time in accordance to the learner's pace. Teachers were not in a position of looking at their students' progress at regular basis so as to check their strength and weakness. Similarly, trends in giving immediate feedback were found weak or too low either in the form of written or verbal dialogues.

The other important issue which was not properly treated in all studied Universities during modular instruction was student's skills measurement. The assessment techniques employed at all universities were not having the capacity to measure fully the performance of the student in all domains. The assessment techniques were fully attached with theories which did not go beyond the lower level of cognitive domain.

In competence based assessment, teachers were expected to be clearer in their techniques by inviting their students to join and reach at an agreement on the required procedure. However, in all studied Universities such type of trends seemed to be less. Here, students are expected to take all types of exam without being aware of the assessment criteria that are needed for modification.

Generally, the practice of assessment in competence based modular instruction within the studied Universities, lacks the two main distinct types of assessments: assessment of learning, and assessment for learning.

Conclusion and Recommendations

Conclusions

In our higher education institutes, teaching learning method is considering a change from traditional lecture and demonstration to self-paced modular approach. As a result, numerous public and private institutes have made this change in the past few years. Similarly, the three selected higher education institutes in Amhara region are contemplating this transition as well. Therefore, research was conducted to determine whether the change to modular instruction specifically the practices of student assessment in competency based modular instruction

could be qualitatively and quantitatively proven superior to traditional instruction.

In order to reach reliable results, a mixed type of research was conducted on three colleges of education from the selected HEIs. Data was gathered by the use of questionnaire, interview and focused group discussion both from the teachers and students to reach the following conclusion.

As has been indicated in the analysis and interpretation section, the practice of continuous assessment seems to be weak in supporting the teaching learning process. Similarly, assessments were not made on time in according to the learner pace. In most cases teachers preferred to assess their students at the end of each module which seems to be a summative type of assessment. On the other hand, the use of formative assessment was found less. Moreover, in all studied universities, assessment highly relied on cognitive domains. The rest; psychomotor and affective domains were either ignored or overlooked.

Moreover, recently the shift towards competency based education has reframed the design, implementation, assessment and evaluation of education. Within this context assessment should be focused either on the achievement of individual competences or conceptualized around the established milestones. From this perspective assessment that has been made in all sampled Universities seems to be weak. Great effort was not exerted to check students' improvement in the world of work. The assessment techniques were fully attached with theories which do not go beyond the lower level of cognitive domain.

Likewise, criteria that were used at the time of assessments were not shared for the owner of the subject (the learner). Students are expected to take all types of exam without being aware of the assessment criteria that are needed for modification. The need of clarification of performance assessment criteria for learners was totally ignored.

One of the main features of quality teaching is giving effective comments on students' work. From this

perspective as many scholars have agreed with, timely feedback is central to student learning. Timely feedback plays a decisive role in learning and development within and beyond the formal educational settings. Written annotations and comments on drafts or on finalized assignments in addition to verbal dialogues prior to or after submission are basic. However, in this research, results have shown us trends in giving immediate feedback was found weak or too low either in the form of written or verbal dialogues.

Recommendations

Overall, this study has tried to examine both internal and external enabling conditions of continuous assessment in HEIs within the selected samples. From the result it was clearly seen that significant limitation were found in all areas of assessment practices.

This calls for a closer attention of the practice of modularization across all HEIs. Therefore, in accordance with the finding, the following suggestions were forwarded for the purpose of improvement.

- In competence based modular instruction, the use of continuous assessment and timely feedback are fundamentals. Therefore, teachers who are involved in such type of mode of delivery should be trained and familiarized with the different techniques of assessment like self-assessment, peer assessment, formative and summative assessment. The training should focus on how teachers can carry out continuous assessment in the different teaching and learning stations with ease.
- Regular training and seminars/workshops should be constantly organized for teachers to update their knowledge of the process involved in the implementation of continuous assessment to further boost the realization of learning objectives as room still exists for improvement.
- Competency-based assessment has to be done using the following assessment strategies:
 - Direct observation
 - Multiple source feedback: which is a questionnaire based assessment strategy

that includes self-evaluation and feedback on observable behaviors from colleagues and co-workers.

- Audit and feedback: which is an assessment strategy that provides performance data (typically from records) with feedback
 - Portfolios and Reflective Learning Tools: Portfolios are technological tools that span the educational continuum and provide formative assessment of the proficiency of individual learning and improvement where scores and judgments are based on the individual data elements.
 - Finally a practicum guide with a performance assessment form should be developed and distributed to all stake holders.
- Assessment criteria should be clear so as to understand what students are trying to achieve in a given task and why they are doing it. Moreover, in order to know more about the overall change of the learner, the higher order of learning outcomes like psychomotor and affective domain should be checked with the help of the above stated strategies.

References

- Farrell, J.P. (2002). *The Aga Khan Foundation Experience Compared with Emerging Alternatives to Formal Schooling.* In Stephen E. Anderson (ed.).
- Glasser, W. (1990). *The Quality School, Managing Students Without Coercion.* New York: Harper Collins.
- Gibbons, M. (1998). *Higher Education Relevance in the 21st Century.* Paper prepared at the UNESCO World Conference on Higher Education. Paris, 5-9 October.
- Hand, L., Knowles, V., Pybus, L., Scivier, J. & Simpson, S. (2000). *What are we learning? about 'Developing Learning' modules?* Nottingham Business School Teaching & Learning Conference, Nottingham Trent University
- Kouwenhoven, W. (2009). *Competence-based Curriculum Development in Higher Education: a Globalised Concept?, Technology Education and Development,* AleksandarLazinic and Carlos Calafate (Ed.)
- Leu, Elizabeth. (2005). *The Role of Teachers, Schools, and Communities in Quality Education: A Review of the Literature.* Washington, DC: AED, Global Education Center. UNESCO. 2004. EFA Global Monitoring Report 2005: Education for All—the Quality Imperative. Paris: UNESCO.
- Maamouri, M., & Wagner, D. (2001). *University and Technology for Literacy/ Basic Education Partnership in Developing Countries (UTLP): Roundtable Overview.* Paris.
- Meyer, H.M., Rees, M., Fletcher, R.B., Johnston, P.M. and Anderson, H. (2009). *Tertiary Assessment & Higher Education Student Outcomes: Policy, Practice & Research.* Massey University, Wellington Campus, Wellington, New Zealand.
- MOE (2012). *A Guideline for Modularization to Ethiopian Higher Education Institutions.* Addis Ababa. Federal Democratic Republic of Ethiopia.
- Mulder, M. (2001) Competence development - some background thoughts, *Journal of Agricultural Education and Extension*, 7(4), 147-159.
- UCD (2011). *User's Guide to the General Regulations.* Retrieved on November 20, 2014, from <http://www.ucd.ie/registry/academicsecretariat/asug/modulesandcredits/>
- USAID (2006). *Issue Paper Perceptions of Ethiopian Teachers and Principals on Quality of Education.* Washington, DC: USAID/EQUIP1 Program

Improving Assessment of Learning in Mathematics through Assessment as Learning

Y. K. A. Etsey and Abraham Gyamfi

Department of Education and Psychology, University of Cape Coast, Cape Coast, Ghana.
kafuietsey@yahoo.com+233244148126 and ayebinegyamfi@yahoo.com +233541149484

Abstract

The purpose of the study was to determine the impact of assessment as learning on assessment of learning in Mathematics in public senior high schools in Ahanta West and Mpohor Districts of the Western Region of Ghana. The basic research design for the study was quasi-experimental post-test control group design. Multi-stage, stratified and purposive sampling techniques were used in selecting two public Senior High Schools. Two classes were randomly selected to constitute the experimental and control groups for the study. An intervention package based on the Senior High School Core Mathematics syllabus, using assessment as learning strategies and lasted for three weeks, was administered to the experimental group. The control group was taken through the traditional method of teaching. Post-test data was collected using a Mathematics Performance Test after the intervention. A Cronbach's alpha of 0.84 was obtained for the Performance Test. An independent sample t-test was used for the statistical analysis at a significance level of 0.05.

The results showed that students in the experimental class, who were taught with the assessment as learning strategies, performed significantly better than students in the control group that had the traditional method of teaching. No statistically significant gender differences were found in the experimental group. It was recommended that assessment as learning strategies should be used in teaching Mathematics as this would improve performance in assessment of learning.

Key words: Assessment as learning, Assessment of learning, Mathematics teaching

Background

Mathematics has been universally recognized as the fulcrum of the Sciences and many people in the world require its knowledge to overcome some life challenges. It is seen as an important subject in modern society and is useful in schools, workplaces, businesses and for personal decision-making. Mathematics is seen to be the language of everyday use whether in the market place, schools or even at home. It is the language of life and some individuals may not be able to survive without it because it is the tool needed to cope with the challenges of life (Omoegun, 2009). Okereke (2006) observed that Mathematics enhances students' thinking process to develop a mindset necessary for solving real life problems because it makes one examine a problem from

various perspectives through the application of logical patterns. Mathematics is fundamental to national prosperity in providing tools for understanding Science, Engineering, Technology and Economics (Kulbir, 2006).

The importance of Mathematics led the Ghanaian Government to make it a compulsory subject in Basic and the Senior High Schools as well as a prerequisite for admission to tertiary institutions. In Ghana, Mathematics is one of the core subjects taught at all levels and the criteria for selection of students into the tertiary institutions requires a mandatory pass in Mathematics in the West Africa Senior High School Certificate Examination (MOE, 2013). Any employment opportunity available for Senior High School graduates

requires a pass in Mathematics. This makes mathematics a – must – pass subject in the senior high schools. However, Mathematics teaching and learning at the

Senior High School level in Ghana have been plagued with poor results, giving teachers, parents, curriculum experts, counsellors and evaluators serious concern.

Table 1 shows the performance of Senior High School students in Core Mathematics in the West African Senior

Secondary Certificate Examination (WASSCE) over a five-year period in Ghana.

Table 1

Summary of May/June WASSCE Core Mathematics Results in Ghana from 2011-2015

| Year | Total Candidates | Number Passed (A1-C6) | Percentage Passed |
|------|------------------|-----------------------|-------------------|
| 2015 | 468812 | 118563 | 25.3 |
| 2014 | 409759 | 132762 | 32.4 |
| 2013 | 405356 | 148567 | 36.7 |
| 2012 | 174296 | 86677 | 49.7 |
| 2011 | 147234 | 64665 | 43.9 |

Source: West African Examination Council, (2015)

It could be observed from Table 1 that each year, the percentage of passes is less than 50%.

Table 2 shows the performance in Core Mathematics in the Schools chosen for the study.

Table 2

Summary of May/June WASSCE Core Mathematics Results in BaidooBonsoe and Mpohor Senior. High Schools from 2012-2015

| Year | Senior High School | Total Candidates | Number Passed (A1-C6) | Percentage Passed |
|------|--------------------|------------------|-----------------------|-------------------|
| 2015 | BaidooBonsoe | 655 | 97 | 14.8 |
| | Mpohor | 343 | 27 | 7.8 |
| 2014 | BaidooBonsoe | 510 | 35 | 6.9 |
| | Mpohor | 243 | 117 | 48.2 |
| 2013 | BaidooBonsoe | 916 | 73 | 8.0 |
| | Mpohor | 394 | 42 | 10.7 |
| 2012 | BaidooBonsoe | 433 | 148 | 34.2 |
| | Mpohor | 197 | 96 | 48.7 |

Source: West African Examination Council, (2015)

From Table 2, it could be observed that except for Mpohor in 2012 and 2014, performance in Core Mathematics in these schools was abysmal.

A number of reasons have been adduced for the poor performance in Mathematics. Owolabi (2003) explained that the failure rate may be attributed, among other reasons, to lack of interpreting, explaining, analysing and evaluating questions. Okereke, (2006) further noted that students seem to shy away from the subject for reasons which include phobia, teachers' attitude towards the teaching of Mathematics and students' negative attitude from the assumption that Mathematics is generally a difficult subject to study.

Studies (Etsey&Snetzler, 1998; Machin.&Pekkarinen, 2008; Asante, 2010; Stoet&Geary, 2013) have also shown that there is some gap in mathematics achievement between males and females with males performing better. Males are much more likely than females to pursue careers in mathematics-intensive fields, such as engineering and computer science.

Chukwuyenum (2013) in the study of the impacts of critical thinking skills and peer assessment on Secondary School students' performance in Mathematics in the Delta State of Nigeria, in the face of the low mathematics performance by senior high school students in the Delta state, found that there was a significant difference in performance between students who were exposed to critical thinking skills and peer assessment and those who did not receive it. Students who received the alternative assessment performed better than those who did not receive the critical thinking skills and the peer assessment. The study also revealed that there was no significant difference in performance due to gender.

Sparks-Langer, Simmons, Pasch, Colton and Starko (1990), opined that learning can be augmented and be more profound when students are encouraged to reflect on the learning event and exercise their judgment about the content and the processes of learning. The benefits of this "consciousness raising" are that students learn to chart their development, identify barriers, patterns and strategies, challenge their own beliefs and avoid repeating the same mistakes.

One of the ways suggested by which performance in Mathematics can be improved is by adopting assessment as learning strategies. Assessment as learning (AaL) is the process of developing and supporting learner metacognition (knowledge of one's own thought processes). Learners are actively engaged in the assessment process; that is, they monitor their own learning (Earl, 2003). Through this process learners are able to learn about themselves as learners and become aware of how they learn. They become metacognitive. WNCP (2006) stated that

Assessment as learning focuses on learners and emphasizes assessment as a process of metacognition (knowledge of one's own thought processes) for learners. Assessment as learning emerges from the idea that learning is not just a matter of transferring ideas from someone who is knowledgeable to someone who is not, but is an active process of cognitive restructuring that occurs when individuals interact with new ideas. Within this view of learning, learners are the critical connectors between assessment and learning. For learners to be actively engaged in creating their own understanding, they must learn to be critical assessors who make sense of information, relate it to prior knowledge, and use it for new learning. This is the regulatory process in metacognition; that is, learners become adept at personally monitoring what they are learning, and use what they discover from the monitoring to make adjustments, adaptations, and even major changes in their thinking. (p. 41)

Thus assessment as learning involves the learners to look at their learning and reflecting on their own abilities. With instructor guidance and through focused activities, learners are encouraged to think about and assess their learning. Teachers only assess learners' ability to assess themselves. AaL is a continuous process and involves self-monitoring and self-evaluation and assists learners to become aware of their learning process. It also requires teacher guidance and support.

As Etsey (2016) noted, AaL helps learners to take more responsibility for their own learning and monitoring future directions. It occurs when learners reflect on and monitor their progress to inform their future learning goals. When learners develop learning goals, think reflectively, self-monitor their learning it is assessment as learning. It develops learners' skills of metacognition, critical thinking and communication.

Through the process of assessment as learning, learners are able to learn about themselves as learners and become aware of how they learn. They reflect on their work on a regular basis, and decide what their next learning will be. Learners take more responsibility for their own learning and monitoring future directions.

Statement of the Problem

Mathematics sharpens the reasoning powers of human beings and increases mental alertness. It helps to produce breakthroughs that are of much benefit to nations and society. This is why the study of some Mathematics is compulsory up to the secondary level of all education systems. In Ghana it is a core requirement for admission to tertiary education.

However, performance in Mathematics is very poor in the West African Secondary School Examinations in Ghana. The effect is that several students who should gain admission to tertiary institutions in the first instance, but fail Mathematics, have to rewrite it again, thus wasting resources.

In response to the problem of poor performance in Mathematics, this study was carried out to find out if assessment as learning would lead to improvement in Mathematics performance in Senior High Schools in the Ahanta - West and Mpohor Districts of the Western region of Ghana. In addition, the study was to investigate whether assessment as learning strategies could bridge the gap in Mathematics performance between males and females.

Hypotheses

Specifically, this study was designed to test the following hypotheses:

1. There is no statistically significant difference in mathematics performance between the SHS students who are exposed to assessment as learning and colleagues who are not exposed to assessment as learning.
2. There is no statistically significant gender difference in performance within the experimental group who are exposed to assessment as learning.

Method

Participants

Eighty-one (81) students from two Senior High Schools in the Ahanta West and Mpohor Districts of the Western Region participated in the study. Thirty-two (32) students were in the experimental group and forty-nine (49) in the control group.

A multistage sampling procedure was adopted for the selection of the participants. In the first instance, purposive sampling procedure was used to select Baidoo Bonsoe Snr./Tech School because it is the only mixed school in the Ahanta West District and Mpohor Snr. High school had a higher population in terms of students and staff in the Mpohor District. A stratified sampling technique was then used to select a class from each course from both schools for the pre – intervention test. The purpose of the pre-intervention test was to find out intact classes from the two schools with no significant difference in mathematics achievement. Simple random sampling was also used to select one pair out of the pairs with no significant difference in the pre-intervention mean scores for the study. Finally simple random sampling technique was used to decide on which of the two schools would be the control group and which would be the experimental group.

Research Design

The research design for the study was quasi-experimental post -test control group design. According to Cook and Campbell (1979) and Adjei and Tagoe (2009), a quasi-

experimental design is one that looks a bit like an experimental design but lacks the key ingredient -- random assignment. A quasi-experimental study is a type of evaluation which aims to determine whether a program or intervention has the intended effect on study participants. Quasi-experimental studies take on many forms, but may best be defined as lacking key components of a true experiment. While a true experiment includes (1) pre-post-test design, (2) a treatment group and a control group, and (3) random assignment of study participants, quasi-experimental studies lack one or more of these design elements.

Since the most common form of a quasi-experimental study includes a pre-post-test design with both a treatment group and a control group, quasi-experimental studies are often an impact evaluation that assigns members to the treatment group and control group by a method other than random assignment. It consists of two groups, one training group and one control group. In this study, BaidooBonsoe Senior High School Form 2V Art (Visual Art) was the experimental group and Mpohor Senior High School Form 2H (Home Economics) was the control group.

Instrument

Two Mathematics Performance Tests were used for the study. One was developed for pre-intervention and one for post intervention. The pre-intervention test comprised of 20 multiple choice items of 4 options each and was based on "mensuration and fractions". The post intervention test comprised of 20 multiple choice items of 4 options each and was based on "fractions and number bases."

A pilot-testing of the test was carried out on 30 students each from two Senior High Schools; one in the Ellebelle District (BonzuKaku Snr. High) and the other in the Nzema East District (Nsien Snr. High). A table of specification was also used to ensure that the items cover all the relevant learning outcomes. Reliability coefficients of 0.78 and 0.84 were obtained for the pre and post intervention tests respectively.

Procedure

The consent of the headmasters for the schools was first sought and then support and collaboration from the students.

The study was carried out in three phases. During the first phase, a pre-intervention test was administered to 7 classes from BaidooBonsoe and 5 classes from Mpohor. After the pre-test data was analysed, a class each from both schools with no significant difference in mean scores were selected. At the later part of phase one, the purpose of the study was made known to the groups. All materials needed for the lessons such as jotters for both groups for note taking, exercise and assignment and a reflection guide and another jotter for the experimental group for reflective journaling were made available to the students.

In the second phase, an intervention package based on selected topics in the SHS Core Mathematics syllabus which lasted for three weeks was administered to the experimental group and assessed using assessment as learning strategies. One of the researchers assumed the role of a classroom teacher. He met the group thrice a week with each meeting lasting for 40 minutes. The topics for the study were "fractions and number bases". The same topics were treated with the control group also for three weeks and the same duration but with the use of the traditional forms of assessment.

During the intervention stage, the broad topics "fractions and number bases" were divided into subtopics. Lesson plans were prepared on each lesson to be taught for each group. Assessment as learning strategies were used in the entire lessons for the experimental group and traditional assessment methods for the control group.

For the experimental group, before every lesson began, the teacher refreshed students' mind on the reflection guide; before, during and after a lesson and its purpose. The reflection guide was a list of questions students were to answer as part of their reflection on the lesson learnt. The questions were categorized as monitoring reflection (before and during a lesson begins), after a lesson and

after a quiz or examination. Then the objectives of the lesson were written at a corner of the board to guide students to formulate their goals for the day. In the course of the lesson the teacher stopped and allowed students to do a reflective journal by answering specific questions during a lesson. With this, students were asked to write out what they would do at home about the lesson after highlighting where they understood better, had little understanding and where there was no understanding. Finally, students were asked to finish the journal after a lesson. Exercises and assignments were given to students for practice. The teacher marked answers to the first question. The second question was marked by the students themselves under the guidance of the teacher and discussing each student's error with them. Teacher asked students to exchange books to mark the answers to a third question under his guidance. Sometimes, students were made to write their questions based on what had been done in the class and solve them for marking. The students were made to answer reflection questions after a test or quiz such as 'how did I get this question wrong?'. This was done in every lesson with the experimental group.

For the control group, the teacher began by reviewing the Relevant Previous Knowledge (RPK) as done in normal classroom lessons. Questions and answers during and after the lesson were not different from the normal classroom setting. Evaluation of students' work was done solely by the teacher and the results were given to the students.

During the third phase, a post-test was administered to both groups under standard test administration conditions to determine the effect of the intervention. This was done a day after the end of the intervention. The test lasted for 30 minutes for both groups.

Data Analysis

Data was analysed by using independent sample t-test.

Results

Hypothesis 1

There is no statistically significant difference in mathematics performance between the SHS students who are exposed to assessment as learning and colleagues who are not exposed to assessment as learning.

To test this hypothesis, an independent sample t-test was used for the post intervention test data. The t-test was used because only two groups (experimental and control) were compared and these two groups were independent of each other. The Levene test for homogeneity of variances was significant indicating that the variances were not assumed equal. The analysis therefore took this information into consideration.

The descriptive statistics and the results of the comparison of the Control and Experimental Groups on the Post-Intervention Test are presented in Table 3.

Table 3

Descriptive Statistics and t test of the Control and Experimental Groups on the Post-Intervention Test

| Group | N | Mean | Std. Deviation | t | df* | Sig. (2-tailed) |
|--------------|----------|-------------|-----------------------|----------|------------|------------------------|
| Experimental | 32 | 13.50 | 3.13 | 8.77 | 46.94 | .000 |
| Control | 49 | 8.06 | 1.96 | | | |

* Degrees of freedom have been adjusted as Levene’s test shows that variances are not equal.

The results in Table 3 show that there is a statistically significant difference in performance between the experimental and control groups on the post intervention test at the .05 level of significance – $t(46.94) = 8.77, p = .000$ (two tailed). The experimental group ($M = 13.5, SD = 3.13$) performed better than the control group ($M = 8.06, SD = 1.96$).

Hypothesis 2

There is no statistically significant gender difference in performance within the experimental group who are exposed to assessment as learning.

To test this hypothesis, an independent sample t-test was used for the post intervention test data. The t-test was used because only two groups (male and female) were compared. The Levene test for homogeneity of variances was not significant indicating that the variances were assumed equal.

The descriptive statistics and the results of the comparison of the Male and Female Groups on the Post-Intervention Test are presented in Table 4.

Table 4

Descriptive Statistics and t test of the Male and Female Groups on the Post-Intervention Test

| Group | N | Mean | Std. Deviation | t | df* | Sig. (2-tailed) |
|--------------|----------|-------------|-----------------------|----------|------------|------------------------|
| Male | 14 | 13.07 | 3.69 | -.677 | 30 | .504 |
| Female | 18 | 13.83 | 2.68 | | | |

The results in Table 4 show that there is no statistically significant gender difference in performance between the male and female students on the post intervention test at the .05 level of significance - $t(30) = -.677, p = .504$ (two tailed). The males ($M = 13.07, SD = 3.69$) did not perform better than the females ($M = 13.83, SD = 2.68$).

The result implies that the intervention had an equal effect on both the male and female students. No gender had a greater advantage in the application of the intervention package in assessment as learning.

Discussion

The results showed that the experimental group that had the intervention performed better than the control group that underwent the

traditional method of teaching mathematics. (Experimental group, $M = 13.5, SD = 3.13$; Control group, $M = 8.06, SD = 1.96$). This, as noted by Etsey (2016), could have been a result of the students in the experimental group becoming more aware of the goals of instruction as the teacher wrote these on the board at the beginning of each lesson. The students then used the information provided as the foundation for assessment as learning. They could have consequently set their own goals for improvement in the learning of mathematics during each lesson. They also monitored their own progress determining where they were, where they reached and at what point they would end. During the monitoring process, they possibly identified their stumbling blocks and created ways for handling them.

As an aspect of assessment as learning, students possibly took ownership and responsibility of their own learning. They consequently did not depend solely on the teacher for “spoon-feeding”. They had home assignments related to each day’s activity to practice. This eventually helped to move their thinking forward (metacognition). Students most likely reflected on their work on a regular basis. A reflection guide, which was a list of questions students were to answer as part of their reflection on the lesson learnt, was provided. The teacher provided adequate time during the lesson for reflection.

Assessment as learning thus became a catalyst for improved performance in Mathematics. This study confirmed earlier studies (Onuka, 2007; Onuka & Oladipo, 2006) that the method of teaching and assessment has a profound effect on the performance of students. In Chukwuyenum (2013) study, there was a significant difference in performance between students who were exposed to critical thinking skills and peer assessment and those who were not. Students who received the alternative assessment performed better than those who did not receive the critical thinking skills and the peer assessment. In an Abledu (2000) study, pre-service teachers who were exposed to the alternative assessments during lessons did better than their counterparts who were not exposed to alternative assessment thus confirming that assessment as learning indeed could lead to higher performance in Mathematics. As Wilson and Wing (1998) summed it up, reflective learners are efficient and effective learners because it moves them from surface to deep learning.

The results also showed that there were no gender differences in performance in Mathematics in the experimental group (Males, $M = 13.07, SD = 3.69$; Females, $M = 13.83, SD = 2.68$). Several studies have shown that males generally did better in Mathematics than females (Etsey & Snetzler, 1998; Machin. & Pekkarinen, 2008; Asante, 2010; Stoet & Geary, 2013). Mathematics has been perceived as a male domain and a ‘no go area’ for females. This study has however shown that this situation cannot possibly prevail at all times. The teaching approach adopted could make a huge difference. As Bloom (1976) emphasized, instructional strategies where students actively participate in their own learning is critical for success. Assessment as learning was a strategy where students actively participated and monitored their own progress and this could have resulted in an improved performance for females.

Wilson (1996) noted that, in the literature, it is pointed out that for students to accomplish learning, teachers should provide meaningful and authentic learning activities to enable students to construct their

understanding and knowledge of this subject domain. In this study, by using assessment as learning, the teacher provided authentic learning activities. This possibly made the experimental group and the female students to construct their own understanding in Mathematics, leading to improved performance.

Conclusion

This study used three weeks for intervention in the teaching of Mathematics using assessment as learning as a strategy. It should be stated that the differences in performance after the three-week intervention could be explained by the assessment as learning strategies. This is because the groups were first measured to have no significant difference in mathematics performance before the intervention. Also the two groups were taught by the same teacher (researcher) for the same duration.

The students who received assessment as learning strategy were more active, showed more commitment in learning Mathematics and were very much involved in the assessment process. They involved themselves in the search for additional knowledge to advance their understanding of the topics taught. Students were involved in critical reflection. They reflected on the teaching and learning activities in the classroom and this possibly helped them to acquire critical thinking skills.

Recommendations

On the basis of this study, it is recommended that assessment as learning strategies should be used in the teaching of Mathematics as this would improve performance. It is also recommended that assessment as learning should be made a major topic in the teaching methods courses at the teacher training institutions, where this is not currently being done. This will help provide the knowledge and skills needed in using assessment as learning in the classroom. In-service training and workshops should be organised for teachers on the job by district and regional directorates on assessment as learning. This would help the teachers to acquire the

knowledge and skills needed in the use of assessment as learning as a teaching and learning strategy.

References

- Abledu, G. K. (2000). *The effects of alternative assessments on the attitudes and achievement in mathematics of female pre-service teacher*. Unpublished master's thesis, Department of Mathematics and Science Education, University of Cape Coast, Ghana
- Adjei, E. & Tagoe, M. (2009). *Research methods in information studies*. Accra: IAE (UG)
- Asante, O. K. (2010). Sex differences in mathematics performance among senior high students in Ghana. *Gender & Behaviour*, 8(2), 3279-3289.
- Bloom, B. (1976). *Human characteristics and school learning*. New York: McGraw Hill, Inc.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally College Publications Company.
- Earl, L. (2003). *Assessment as learning: Using classroom assessment to maximize students' learning*. Thousand Oaks, CA: Corwin
- Etsey, Y. K. A. (2016). *Assessment as learning: A metacognition approach to teaching, learning and assessment*. A paper presented at WAEC monthly seminar. WAEC: Accra.
- Etsey, Y. K. & Snetzler, S. (1998). *A meta-analysis of gender differences in student attitudes toward Mathematics*. ERIC Document Reproduction Service No. ED 435 543. Columbus, OH.
- Chukwuyenum, A. N. (2013). Impact of critical thinking on performance in Mathematics among senior secondary school students in Delta State, Nigeria. *Journal of Research & Method in Education* 3(5), 18-25.
- Kulbir, S. S. (2006). *The teachers of mathematics*. New Delhi: Sterling Publishers Private Ltd.
- Machin, S. & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science* 32, 1331-1332.

- MOE (2013). *Manual for selection of students into the tertiary institutions*. Accra: Author.
- Okereke, S. C (2006). Effect of prior knowledge and implication of mathematics tasks and concept to career types and gender on students' achievement, interest and retention. In U. Nzewi (Ed) *STAN procedures of the 47th Annual conference*, (pp. 253-259).
- Omogun, M. (2009). *Career guidance for Nigerian students*. Lagos: Bab Sheriff Ltd.
- Onuka, A. O. U. (2007). Teacher-initiated students- peer assessment: A means of improving learning assessment in large classes. *International Journal of Africa American Studies*, 6 (1), 18-24.
- Onuka, A. O. U & Oladipo, B. D. (2006). Systematic school based assessment for an improved cognitive performance. *Academic Journal of Research and Development 1* (1). 1-12.
- Owolabi, H. O. (2003). Antecedents of current procedures of evaluating learning outcomes in Nigeria educational system. *Nigeria Journal of Educational Foundations*, 6(1), 34-49.
- Sparks-Langer, G. M., Simmons, J. M., Pasch, M., Colton, A. B. & Starko, A. (1990). Promoting reflection in teacher preparation programs: A multi-level. *Journal of Teacher Education*, 41(5), 23-32.
- Stoet, G. & Geary, D. C. (2013). *Sex differences in mathematics and reading achievement are inversely related: within- and across-nation assessment of 10 years of PISA data*. Retrieved, 31/05/2017 from <https://doi.org/10.1371/journal.pone.0057988>
- Wilson, J. & Wing, L. (1998). *Self-assessment for students: Proformas and guidelines*. Armadale: Eleanor Curtin Publishing.
- Wilson, B. G. (Ed.). (1996). *Constructivist learning environments: Case studies in instructional design*. Englewood Cliffs, NJ: Educational Technology Publication.
- WNCP (2006). *Rethinking classroom assessment with purpose in mind*. Western and Northern Canadian Protocol for Collaboration in Education.

Subject interest and Biology Achievement of Senior Secondary Students in Rivers State: Effect of Assessment for Learning Strategies

Dr (Mrs.) Amakiri, Hager Atisi Eremina
Department of Educational Psychology, Guidance and Counselling,
Ignatius Ajuru University of Education, Rivers State, Nigeria.
amakiriatisi@yahoo.com

Abstract

The study examined subject interest and Biology achievement of senior secondary students in Rivers State Nigeria based on the effect of assessment for learning strategies. The Multistage sampling technique was used to compose a sample of 400 senior secondary two (SS2) students from public schools in five local government areas of Rivers State. A Non-randomized pretest-posttest control group, quasi-experimental design was adopted for the study. Students were assigned to four experimental groups and a control group. Two null hypotheses drawn from two research questions were tested at 0.05 level of significance. A 40 – item multiple choice objective test titled: “Assessment for Learning Biology Test” (ALBT) was used for collection of data. Copies of the instrument were validated by three subjects (biology) specialist and two experts in test and measurement. A trial-test was conducted on 40 non-participating students from the population. Rulon formula was used to establish the internal consistency coefficient of 0.71. Descriptive statistics, independent t-test and analysis of covariance (ANCOVA) was used for data analysis. The result revealed that there was a statistical significant difference among arts and science inclined students in the treatment groups. Interaction effect between subject interest and assessment for learning strategies was also significant. Assessment for learning strategies enhanced the achievement of arts and science inclined students. It was recommended that teacher trainees and professional teachers be adequately trained through workshops and seminars on appropriate utilization of assessment for learning strategies in classroom assessment.

Keywords: *Subject Interest; Biology Achievement; Assessment for Learning; Learning Strategies.*

Introduction

The low level of technological development in Nigeria for decades has been a source of concern to educators, educational institutions, organizations, stakeholders and the nation at large. The technological development of any nation depends largely on its natural and human resources. Science education is the foundation for technological development of any nation. It is based on this realization that the Federal Government of Nigeria, over the years, has invested huge resources in science education. Despite efforts made by the government to encourage and support individuals, institutions and agencies in the re-vitalization of science education in Nigeria, there is continuous decline in interest of students in science education especially at pre- higher education level. This is a dangerous trend because the enrolment of

students in science and science related courses in tertiary education is dependent on science students from senior secondary schools. Krapp (2002) stated that as students advance in secondary school, interest in science subjects (Physics and Chemistry) and Mathematics dwindles. The science subject most preferred, especially by arts inclined students as a non-vocational elective is Biology.

Nworgu (in Amakiri, 2008) observed that enrolment recorded in Biology is higher compared to other subjects. This is attributed to the assumption that Biology is less empirical and relevant to everyday life. The study of Sakiyo and Badau (2015) also affirmed that among science subjects, biology had the highest enrolment in the Senior Secondary Certificate Examination conducted by the West African Examination Council in Nigeria from

2008 - 2012. The study showed that students performed better in Physics, while the lowest achievement within this period of study (2008 - 2012) was recorded in Biology. Ahmed (2008) opined that despite the relevance of Biology, students' achievement in the subject in Senior Secondary Certificate Examination (SSCE) has degenerated. For Okoobah, Afolabi and Asikhia (2014), at the secondary school level, the unimpressive learning outcomes in science subjects are attributed to insufficient interest in science and technical education. Similarly, Bamidele (2004) observed that students' enrolment and performance in science subjects is affected by lack of interest in science based subjects because of the preconceived idea that science subjects are difficult.

Lia (2010) defined 'interest in learning' as a person's preference in a subject matter over other things. This means that an individual has a choice over another thing which may require a psychological positivity to deal with any factor that may arise during the period of study to enhance learning outcomes. In the context of this study, subject interest is defined as the preference and inclination of students towards a specific discipline (arts or science subjects). Students' preference for science subjects, according to Stables (in Nyamba & Mwajombe, 2012) refers to, students' interest as the desire, likes or willingly wanting to study science subjects. The study of Nyamba and Mwajombe revealed that there is a close relationship between subject preference and achievement. This is in line with the findings of Athanasou (1994), in a study carried out on technical and further education students in Australia, which revealed that there is a direct relationship between an individual's interest in a subject, his or her vocational choice and the educational achievement. In a classroom setting, the students' emotional and intellectual needs are based on interest. Hence, external forces cannot impose interest on anyone. However, the teacher can aid the learners' interest (Li, in Yu-Je, Chia-Hi, & Chin-Yaw, 2011). This implies that the students' interest and academic achievement in a course of study or subject could be influenced or directed by the teacher or instructor. A teacher's attitude, approach, method of teaching and assessment can

facilitate students' interest in a subject or course in which they had previously not indicated interest.

In the view of Kallaghan and Greaney (2001), a teachers' assessment deserves a second consideration in terms of improving the quality of education. The attention on assessment of students' achievement has highlighted assessment as an intrinsic aspect of instruction and learning. Hence, for a teacher to be effective during instruction and learning, formative assessment must be executed appropriately in the classroom. Academic accomplishment is a feedback system which reveals the limit to the course content attained (Horvath, 2011). This indicates that if appropriate feedback on a subject or course is readily available to learners, based on formative assessment, instruction and learning especially in science subjects would be achieved; the preconceived idea that science subjects are difficult would be jettisoned. It is imperative for teachers and learners to become stakeholders in the assessment process; it should be seen as a process for improving and motivating learning; boost students' interest in a course or subject and not a tool for threatening learners, generating scores, grading, comparison, classification and punishment.

Assessment for learning (AFL), is an aspect of formative assessment identified by Stiggins and Chappius (2005). Also, Black and Wiliam (1998a, 1998b, 1999), extensive research is based on assessment for learning. This comprehensive study by these researchers disclosed that assessment for learning strategies enhanced students' academic achievement.

Assessment for learning involves diverse techniques employed by instructors to elicit information on students' strengths and inadequacies which are ploughed back to the students for future lesson-planning (Black, Harrison, Marshall, Lee, & Wiliam; 2003). The following Assessment for Learning strategies were identified by Black, Marshall, Harrison, Lee & Wiliam (2003:3):

- Teacher's use of Questioning: it involves the use of questions by teachers to diagnose and extend students' ideas.

- Feedback through Marking: use of descriptive comments instead of grades to inform students on their academic achievement and direct learning.
- Peer and Self-Assessment: peer assessment refers to the process where learners rate their peers, based on clearly stated guidelines by the teacher or instructor. While self-assessment refers to one's appraisal of his/her own learning outcomes based on given criteria.
- Formative use of summative assessment: refers to the use of relevant items from previous examinations and test papers to assess learners during teaching/learning in the classroom.

Assessment for Learning is defined as: "the searching and involvement of proofs by teachers and learners to make input on where the learning is headed to, the direction it must go and the most appropriate steps required to achieving these goals". (Assessment Reform Group [ARG] 2002, p 44). The purpose of Assessment for Learning include:

- Checks learning to decide what to do next;
- Is designed to assist teachers and students;
- Employed in conversation about learning;
- Usually detailed and descriptive feedback in words (instead of numbers, scores and grades)
- Usually focused on improvement, compared with a student's previous best and progress toward a standard;
- Needs to involve the students (ARG, 2002, p 44).

Related literature also revealed that assessment for learning strategies are effective in promoting learning outcomes. For instance, Chin (2006) stated that the type of questions asked and how the questions are asked influences the cognitive process of students' to handle scientific knowledge. Similarly, Croom and Staire (2005) noted that appropriate questioning is positive to analytical reasoning which produce knowledge. For Vogler (2005), questions can monitor comprehension, help make connections to prior learning and can stimulate cognitive growth. On the effect of comment used by teachers, Black and Wiliam (1998) revealed that descriptive feedback

produces the highest achievement than the use of scores. Furthermore, Johnson (2004) opined that peer and self-assessment are utilized by educators to enhance learning; to increase students' involvement in the learning process; to increase social interactions and trust in others; to facilitate individual feedback and to focus students on the process rather than the product. Hogson and Pyle (2010) stated that self and peer assessment are a foundation of good practice of AFL and its importance, as it concerns the study of science, that the show-casing of misconception enhances students' personal study and gets them acquainted in science processing skills. Though literature on the effectiveness of assessment for learning on students' achievement in science subject abounds, the assessment strategies that prevailed among secondary school teachers at the time of this study were continuous testing and homework. The thrust of this study therefore was to determine if there was a difference in the Biology achievement of art and science- inclined students based on the effect of assessment for learning strategies.

Statement of the problem

The major goals of science education in Nigeria as stipulated in the National Policy of Education (2004) include the following; (a) producing scientists for national development, (b) service studies in technology and cause to technological development, and (c) providing knowledge and understanding of the complexity of the physical world, the forms and the conduct of life. To achieve these goals, in senior secondary school, biology among other science subjects (chemistry, physics and health science) is considered a core subject and a non-vocational elective. Based on the National Policy on Education, it is mandatory for all senior secondary students, irrespective of subject interest or inclination to enroll in one core subject. However, government educational policy to encourage the study of science subjects at pre-higher education though laudable has not yielded the expected result. Recent research has shown a steady decline in students' interest and high rate of failure in science subjects in pre-higher education (Okooah, Afolabi & Asikhia, 2014). This unimpressive trend is reflected in the senior secondary examination

results released by the West African Examination Council in Nigeria for the past eleven (2002-2012) years. These results reveal that the percentage of candidates who passed biology at credit level over these years were below 50% as shown: 31.52%, 44.15%, 24.69%, 35.04%, 48.60%, 33.37%, 33.94%, 33.87%, 33.70%, 38.50% and 38.82% respectively (Abimbola in Akanbi & Kolawole, 2014).

Biology is perceived especially by arts inclined students as the easiest among other science subjects. However, the steady decline in biology achievement among students could be attributed to the high enrolment of arts inclined students who opt for the subject even when they lack interest in it. In senior secondary schools in Nigeria, a student's career pathway is determined by his or her choice of subjects at the senior secondary two (SS2) level. Students are taught core arts and science subjects in senior secondary one (SS1), this is to enable them decide their subject preference in senior secondary two. If the foundation is not properly laid in science subjects in senior secondary one, most students opt for arts subjects based on the assumption that they can study on their own and pass their examination unlike the science subjects that are more empirical and require more guidance from teachers.

The implication of this drift from science subjects will negatively impact on the future of science, technology and human capital development in sciences in Nigeria. This may affect the achievement of major goals of science education and the technological advancement in Nigeria. There is therefore, an urgency to re-direct students' interest in science subjects through strategies that would foster learning, rather than imposing the subject on the learners.

This study investigates the effect of strategies of assessment for learning on biology achievement of arts and science inclined students.

Objective of the Study

The main objective of this study was to determine whether assessment for learning strategies could improve

arts and science inclined students' achievement in biology.

Specifically, the study sought to;

1. Determine if there is a significant effect of assessment for learning strategies on biology achievement of arts and science inclined students.
2. Establish if there is a significant difference on mean achievement scores of students based on subject interest.
3. Ascertain if there is a significant interaction effect of subject interest and AFL strategies on biology achievement of students.

Research questions

1. Is there a significant effect of assessment for learning strategies on the mean achievement scores of arts and science inclined students?
2. Is there a significant interaction effect of subject interest and AFL strategies on mean achievement scores of students?

Hypotheses

The following hypotheses were postulated to guide the study;

1. There is no significant effect of assessment for learning strategies on mean achievement scores of arts and science inclined students.
2. There is no significant effect of subject interest and AFL strategies on mean achievement scores of students.

Methodology

The procedure used to execute the research are discussed as follows:

Design

A pretest – posttest control group quasi-experimental design was adopted for the study.

Study population

The population of the study consisted of senior secondary Two (SS2) students duly registered in public secondary schools in the 23 Local Government Area of Rivers State. This study targeted SS 2 students because they have been exposed to the science subjects, specifically biology in

senior secondary one (SS1). Subject selection and preference are made by senior secondary students in SS2. This also indicated students' subject interest.

Sampling procedure

A sample of 400 senior secondary two students (arts and science inclined) was drawn via multistage sampling technique. Proportional stratified random sampling technique was used in the first stage. The area of the study was divided into upland and riverine region, while stratified random sampling technique was adopted in the second stage to select one secondary school from five Local Government Areas of Rivers State. The five schools represented the four treatment groups and control group respectively. Students were not randomly assigned to experimental and control groups because this would have disrupted the regular class time table in the schools used for the study. Hence intact classes made up of arts and science students were used.

Instrument

The instrument for data collection was "Assessment for Learning Biology Test" (AFLB). It consists of 40 multiple choice objective items developed by the researcher from the topics taught on aquatic habitat drawn from SS2 biology scheme of work. Each correct item was scored 1 mark with a maximum score of 40. To determine the content validity of the instrument, the researcher used a test blue print to ensure even distribution of the content and objectives of the lessons. Copies of the instrument were validated by three subject (biology) specialist and two experts in test and measurement. The instrument was further validated by administering it on 40 non-participating students from the population. Rulon formula was used to establish the internal consistency coefficient of 0.71.

Administration of instrument and analysis

After due consultation with the principals and biology teachers in the schools used for the study. The researcher obtained a time-table for teaching the students for a period of three weeks. The instrument was administered to participants as pretest to ascertain academic equivalence of the students before instruction and treatment. The pretest scores obtained was used as "the

covariate in the study. After three weeks of treatment, the instrument was reshuffled and administered as posttest to measure the biology achievement of the experimental and control groups. The researcher was assisted in the administration of the instrument by research assistants (biology teachers). Descriptive statistics (mean and standard deviation) was used to answer the research questions, while independent t-test and ANCOVA was used to test the null hypotheses at 0.05 level of significance.

Experimental Procedures

During the teaching/learning process which lasted for 90 minutes per lesson each day. Students in each of the experimental groups were treated with the following AFL strategies adapted from the research by Black and Williams, (1998): (Use of Questioning Comment only marking, self/peer assessment and formative use of summative assessment) while students in control group were assessed based on the traditional method of continuous testing and homework.

The following steps were employed by the researcher in classroom assessment of students.

Experiment Group 1(N= 70)

Treatment: "Use of Questioning" (UOQ)

- Students were engaged in classroom discussions, using incorrect answers drawn from classroom exercises and take-home assignment;
- Allotting more time for students to interface with their classmates on issues under consideration during the teaching and learning process.
- guiding students, to learn from the errors on responses to questions and collaborate with their classmates to reach a resolution; and
- Discouraging the use of hand signals to volunteer answers to questions but asking open-ended questions to ensure that students were able to contribute to the discussions. Questions that promote reflection and discussion were emphasized.

Experiment Group 2(N= 90)

Students in this group were assessed using “Comment only marking” (COM) this involved;

- Non-grading of class work and assignment in students books;
- using verbal, descriptive and written comments to guide and direct students classwork or assignment, not commenting on their personality or character;
- engaging students in scrutinizing their responses to straight forward exercises; and
- monitor the students’ responses to the comments made by sharing their assignment/classwork and notebook into two equal parts, using the left side for classwork and the right side to obtain feedback from students on comments made by the researcher.

Experimental Group 3 (N = 75)

The classroom assessment strategy used in this group was “self/peer assessment” (SPA).

- To facilitate effective interaction among the students’ during the teaching/learning process, the researcher identified students’ level of difficulty with on-going instruction by using a process known as “traffic lighting”. This process encouraged the students to indicate their level of understanding using a piece of paper with appropriate colour indicating their level of comprehension; green (good understanding, amber (partial understanding) and red (little or no understanding).
- Students who indicated good and partial understanding of instruction were paired to interact and facilitate learning while the researcher engaged students who indicated little or no understanding of the instruction and monitored the progress of the amber and green light bearing students.
- Students in these groups were encouraged to make inputs on responses to questions, classwork or assignment of their peers.

- Students were guided to assess themselves as well as their peers classwork and homework based on clearly stated mark schemes and guidelines, while the researcher supervised and cleared any misconception on the lesson or topics being discussed among the students.

Experimental Group 4: (N = 85)

The assessment strategy utilized in this group was “Formative Use of Summative Assessment” (FUST) it involved;

- The use of relevant items from previous Senior Secondary Certificate Examinations administered by the West African Examination Council (WAEC) and National Examination Council (NECO) for classwork assessment of topics taught during and at the end of every teaching unit.

Control Group (N=80)

The control group was assessed during the teaching/learning process based on the traditional method of continuous testing and homework.

Result

The results of the data analysis are as shown below:

Research question one:

Is there a significant effect of Assessment for Learning strategies on mean achievement scores of arts and science inclined students?

Hypothesis one:

There is no significant effect of Assessment for Learning Strategies on mean achievement scores of arts and science inclined students.

Table 1: Posttest mean and standard deviation of effect of AFL strategies on biology achievement of arts and science inclined students in each of the five groups.

| <i>Groups</i> | <i>Subject Interest</i> | <i>N</i> | \bar{X} | <i>SD</i> |
|---------------|---------------------------------|----------|-----------|-----------|
| UOQ | Arts inclined students (AIS) | 40 | 28.42 | 4.038 |
| | Science inclined students (SIS) | 30 | 25.90 | 3.575 |
| | Total | 70 | 27.34 | 4.021 |
| COM | Arts inclined students (AIS) | 50 | 28.96 | 3.833 |
| | Science inclined students (SIS) | 40 | 27.63 | 2.121 |
| SPA | Total | 90 | 28.37 | 3.241 |
| | Arts inclined students (AIS) | 38 | 25.24 | 2.880 |
| FUSA | Science inclined students (SIS) | 37 | 29.16 | 4.586 |
| | Total | 75 | 27.17 | 4.276 |
| | Arts inclined students (AIS) | 44 | 27.20 | 3.554 |
| CG | Science inclined students (SIS) | 41 | 23.76 | 2.737 |
| | Total | 85 | 25.54 | 3.611 |
| | Arts inclined students (AIS) | 40 | 19.90 | 1.817 |
| | Science inclined students (SIS) | 40 | 22.08 | 1.630 |
| | Total | 80 | 20.99 | 2.034 |

Table 1 reveals that the arts inclined students in the treatment groups UOQ, COM and FUSA had higher biology achievement scores (M = 28.42, SD = 4.03), (M = 28.98, SD = 3.83), and (M = 27.20, SD = 3.55) than did the science inclined students (M = 25.90, SD = 3.57), (M = 27.63, SD = 2.12) and (M = 23.76, SD = 2.73) respectively. While the science inclined students in the treatment group SPA had higher biology achievement

scores (M = 29.16, SD = 4.58) than did the arts inclined students (M = 25.24, SD = 2.88) respectively. Though in the control group, science inclined students had slightly higher biology achievement score (M = 22.08, SD = 1.81) than did the arts inclined students (M = 19.90, SD = 1.63), participants (arts and science inclined students) in the treatment groups had higher biology achievement scores than those in the control group.

Table 2: ANCOVA posttest mean achievement scores of arts and science inclined students in experimental and control groups

| <i>Source of Variation</i> | <i>Sum of Square</i> | <i>Df</i> | <i>Mean square</i> | <i>F</i> | <i>P</i> |
|------------------------------|----------------------|-----------|--------------------|----------|----------|
| Covariate(pretest) | 2106.420 | 1 | 2106.420 | 423.606 | .021 |
| Main Effect (AFL strategies) | 721.233 | 4 | 180.308 | 32.260 | .000 |
| Subject interest | 45.716 | 1 | 45.716 | 9.194 | .003 |
| Error | 1934.340 | 389 | 4.973 | | |
| Total | 275647.000 | 400 | | | |
| Corrected total | 7581.939 | 399 | | | |

Table 2 shows that $f(1,389) = 32.26, p.001$, for main effect (AFL strategies) is significant. This indicates a significant effect of AFL strategies on mean achievement scores (posttest) of arts and science inclined students when covariate effect (pretest) was partialled out. The null hypothesis that there is no significant effect of

assessment for learning strategies on mean achievement scores of arts and science inclined students is rejected while, the alternate is accepted. The table also reveals that $f(1,389) = 9.19, p.003$ for subject interest is significant. Independent t-test was employed to establish the direction of significant difference among the treatment and control groups as shown in table 3.

Table 3: t-test analysis of arts and science inclined students' mean achievement scores in experimental and control groups

| Group | Subject Interest | N | \bar{X} | SD | Df | t | P |
|-------|---------------------------------|----|-----------|-------|----|-------|------|
| UOQ | Arts inclined students (AIS) | 40 | 28.42 | 4.038 | 68 | 2.717 | .008 |
| | Science inclined students (SIS) | 30 | 25.90 | 3.575 | | | |
| | Total | 70 | 27.34 | 4.021 | | | |
| COM | Arts inclined students (AIS) | 50 | 28.96 | 3.833 | 79 | 2.094 | .039 |
| | Science inclined students (SIS) | 40 | 27.62 | 2.121 | | | |
| | Total | 90 | 28.37 | 3.241 | | | |
| SPA | Science inclined students (SIS) | 37 | 29.16 | 4.586 | 73 | 4.452 | .000 |
| | Arts inclined students (AIS) | 38 | 25.24 | 2.880 | | | |
| | Total | 75 | 27.17 | 4.276 | | | |
| FUSA | Arts inclined students (AIS) | 44 | 27.48 | 3.295 | 83 | 6.139 | .000 |
| | Science inclined students (SIS) | 41 | 23.46 | 2.675 | | | |
| | Total | 85 | 25.54 | 3.611 | | | |
| CG | Arts inclined students (AIS) | 40 | 19.90 | 1.630 | 78 | 5.635 | .016 |
| | Science inclined students (SIS) | 40 | 22.08 | 1.817 | | | |
| | Total | 80 | 20.99 | 2.034 | | | |

Table 3 shows that there is no significant effect of use of questioning and comment only marking strategies on mean achievement scores of arts and science inclined students, $t(68) = 2.72, p.008$, and $t(79) = 2.09, p.039$ respectively. However, there is a significant effect of self/peer assessment and formative use of summative assessment strategies on mean achievement scores of arts and science inclined students, $t(73) = 4.45, p.001$ and $t(83) = 6.14, p.001$. There is no significant difference in mean achievement scores between arts and science inclined students in the control group, $t(78) = 5.63, p.016$

Research question two

Is there a significant interaction effect of subject interest and AFL strategies on mean achievement scores of students?

Hypothesis two

There is no significant interaction effect of subject interest and AFL strategies on mean achievement scores of students.

Table 4: ANCOVA analysis on the interaction effect of subject interest and AFL strategies on mean achievement scores of students.

| Source of Variation | Sum of Square | df | Mean square | f | P |
|-----------------------------|---------------|-----|-------------|---------|------|
| Covariate (pretest) | 2106.420 | 1 | 2106.420 | 423.606 | .021 |
| Treatment (AFL strategies) | 721.233 | 4 | 180.308 | 32.260 | .000 |
| Subject interest | 45.716 | 1 | 45.716 | 9.194 | .003 |
| Treatment* Subject interest | 251.223 | 4 | 62.806 | 12.630 | .000 |
| Error | 1934.340 | 389 | 4.973 | | |
| Total | 275647.000 | 400 | | | |
| Corrected total | 7581.939 | 399 | | | |

The result presented in table 4 reveals that the interaction effect of subject interest and AFL strategies is statistically significant, $f(1, 389) = 12.63, p .001$. The null hypothesis that there is no significant interaction effect of subject interest and AFL strategies on mean achievement scores

of students is rejected. The alternate hypothesis is accepted. Figure 1 below shows the profile plot of the interaction effect of subject interest and assessment for learning strategies.

Figure 1: Profile Plot of Estimated Marginal Means of subject interest (arts and science inclination) and AFL strategies.

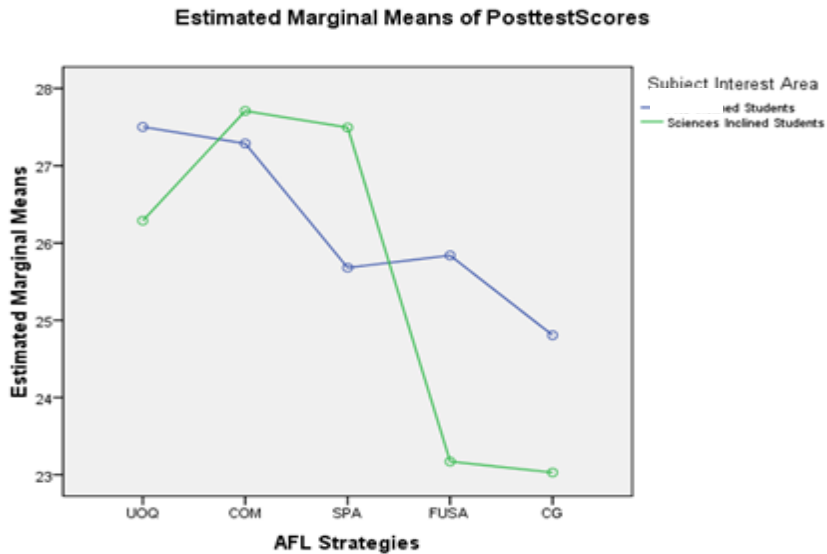


Figure 1 reveals the interaction effect of subject interest and AFL strategies on biology achievement of students.

Discussion of findings

The result on research question one (Table 1) shows that the Assessment for Learning strategies that enhanced the biology achievement of arts inclined students were: use of questioning, comment only marking and formative use of summative assessment; while biology achievement of science inclined students was enhanced by self/peer assessment strategy. This result reveals that assessment for learning strategies that enhanced biology achievement of students differ based on subject interest.

This is in tandem with the findings of Nyamba and Nwajomba (2012) who stated that there is a correlation between subject preference and achievement.

The result of hypothesis one (table 2) reveals that there is a significant effect of assessment for learning strategies on biology achievement of arts and science inclined students in the treatment groups. It could be inferred from the result that assessment for learning strategies enhanced the biology achievement of arts and science inclined students. This agrees with the findings of Black and Williams (1998) that assessment for learning strategies enhance students' academic achievement. Specifically, the independent t-test analysis (table 3) revealed that there is a significant difference in biology achievement between arts and science inclined students subjected to self/peer assessment and formative use of summative assessment based on their posttest scores.

The findings on research question two and the corresponding hypothesis (table 4 & Fig 1) revealed a significant interaction effect between subject interest and AFL strategies.

Conclusion

The findings have established that assessment for learning strategies enhanced biology achievement of arts and science inclined students. Science inclined students achievement is enhanced by self/peer assessment strategy, while arts inclined students achievement is facilitated by use of questioning, comment only marking and formative use of summative assessment strategies. Science instruction and learning involves interaction with peers, critical thinking and use of experiments. Therefore

students' interest in science subjects could be aroused and sustained through interaction with their peers. This process could be facilitated through appropriate utilization of assessment for learning strategies.

Recommendations

The following recommendations were made based on the findings:

- In order to encourage and motivate students to learn and improve their achievement in Biology, teachers should take into consideration students subject interests; hence, science inclined students should be subjected to self/peer assessment strategy to enhance their achievement in Biology, while arts inclined students should be subjected to use of questioning, comment only marking and formative use of summative assessment strategies to improve their biology achievement.
- Teacher trainees and professional teachers should be adequately trained through workshops and seminars on appropriate utilization of assessment for learning strategies in classroom assessment. This would enable them effectively implement the strategies to enhance teaching/learning and students' interest.

References

- Ahmed, M.A. (2008). Influence of personality factors on biology lecturers' Assessment of difficulty levels of genetics concepts in Nigerian Colleges of Education. (Unpublished PhD Thesis), University of Ilorin, Nigeria.
- Amakiri, H.A.E. (2008). Effect of use of continuous assessment on senior secondary students' performance in ecology in Port Harcourt Local Government Area. (Unpublished M.Ed. Thesis), Rivers State University of Science and Technology, Port Harcourt, Rivers State.
- Assessment Reform Group (2002). Assessment for learning: 10 Principles. Retrieved January 26, 2013 from http://www.qca.org.uk/libraryassets/media/4031_aflprinciples.pdf.

- Akanbi, A.A. & Kolawole, C.B. (2014). Effects of guided-discovery and self-learning strategies on senior secondary school students' achievement in biology of education and leadership development, 6(1), 19-39.
- Black, P. & Wiliam, D. (1998a). Inside the Black box. London: Kings College, London School of Education. Retrieved January 20, 2013, from blog.discoveryeducation.com/assessment/files/..blackbox_article.pdf.
- Black, P. Harrison, C., Lee, C. Marshall, B. & Wiliam, D. (2003). Assessment for Learning: putting it into practice. Maidenhead, Open University Press. Retrieved February 18, 2013 from www.amazon.co.uk>books?science of Natures>Education.
- Chin, C. (2006b). Teacher questioning in science classrooms: What approaches stimulate productive thinking? Retrieved January 18, 2012 from <http://repository.nie.edu.sg/jspi/bitst>.
- Croom, B; & Staire, K. (2005). Getting from q to a: effective questioning. *The Agricultural Education Magazine*, 78, 12-14.
- Hogson, C. & Pyle, K. (2010). A literature review of assessment for learning in science. Retrieved January 6, 2013 from <http://www.nferac.uk/publication/ AASO2>.
- Horvath, K. (2011). *Effects of peer tutoring on student achievement*, Ohio: Ronard Publishing Ltd.
- Johnson, R. (2004). Peer assessment in physical education. *Journal of Physical education recreation and Dance*. 75(8), 33-41.
- Kellaghan, T. & Greaney, V. (2001). Using assessment to improve the quality of education, UNESCO: International Institute for Educational Planning, Retrieved February 20, 2013 from <http://www.unesco.Org/iiep>.
- Okooboh, M.A; Afolabi, M.A. & Asikhia, O.A. (2014). Science and technical education in a democratic government for sustainable National development. *The Conference: Journal of Contemporary Educational Thoughts*, 191, 44-48.
- Krapp, A. (2000). Structural and dynamic aspect of interest development: theoretical considerations from an autogenetic perspective. *Learning and Instruction*, 12, 383-409.
- Nyamba, S.Y. & Nwajombe, K.K. (2012). Students' preferences on science subjects: Does this affect their performance? A case of Udzungwa secondary school, Kilolo, Iringa, Tanzania. *International Journal of Science and Technology*. Retrieved March 20, 2013 from <http://www.ejournals.ofsciences.org>.
- Sakiyo, J. & Badau, K.M. (2015). Assessment of the trend of secondary school students' academic performance in the sciences, mathematics and English: Implication for Attainment of the Millennium Development Goals in Nigeria. *Science Research Journal*, 2(2), 31-38.
- Stiggins, R. & Chappius, S. (2005). Putting Testing in Perspective: It's for learning. Retrieved february20, 2012 from www.Mc3edsupport.org/.../putting-testing-in-perspective-its-for-learning
- Vogler, K. (2005). Improving your verbal questioning. Clearing House. 79, 98-103.
- Yu-Je, L; Chia-Hui, C. & Ching-Yaw, C. (2011). The influences of interest in learning and learning Hours on learning outcomes of vocational College Students' in Taiwan: Using a teachers' instructional attitude as the moderator. *Global Journal of Engineering Education*. Retrieved February 18, 2013 from www.wiete.com.au/journal/GJEE/Publish/vol13no3/01-lee-Y-Jpdf

Incorporating Assessment of Learning in Assessment for Learning: Shifting from Foundational Competences to Practical and Reflexive Competences in Learning Institutions

Dennis Zami Atibuni (PhD)*¹ and David Kani Olema (PhD)¹

zamidennis79@gmail.com, Mobile +256 782 305 430.

Busitema University – Nagongera Campus, Department of Education, P.O. Box 236, Tororo.

Abstract

All efforts invested in teaching are futile without ascertaining that learning has taken place. The only proof that learning has taken place is an assessment result or outcome. In considering this single most important proof of teaching and learning, we have provided in this paper a desk-based investigation of three types of assessment outcomes: practical competence, foundational competence, and reflexive competence. The paper aimed to achieve two objectives: (a) to explore the principles of constructive assessment and (b) to suggest the approaches to shifting from summative assessment of the foundational competences to formative, constructive assessment of practical and reflexive competences. Literature reveals that in Uganda, mainly foundational (lower order cognitive domain) as opposed to practical (affective, psychomotor, and higher order cognitive domain) competence is assessed, with little, if any, reflexive competence ingrained in the process. In this case, the learner's reflexive competence of decision making with understanding, and with an ability to adapt appropriately and responsibly to change in unforeseen circumstances and to explain the reasons behind these adaptations cannot be ascertained. We propose that effective assessments for learning practices should have the potential to greatly increase both student achievement and motivation through students' and teachers' knowledge of learning outcomes, existing gaps in learning outcomes, and how to close the gap. This paper thus provides strategies of incorporating assessment of learning in assessment for learning so as to promote practical and reflexive competences in the learners. An assessment that achieves these objectives needs to result in accurate information, provide descriptive rather than evaluative feedback to students, and incorporate student engagement involving the 21st century skills.

Key words: assessment of learning, summative assessment, assessment for learning, formative assessment, practical competence, foundational competence, reflexive competence

Introduction

The term assessment takes various meanings. According to Cizek (1997), assessment could mean new formats for gathering information about learners' achievements, for example portfolios; a new attitude towards gathering information on learner progress, an attitude that is kinder and gentler than the approach of many standardized tests; a new ethos of empowerment in which assessments are planned and conducted to serve the information needs of learners; or a new process that gathers and synthesizes information about the learner.

Most popular reference to assessment is about the process of assessment. In this respect, the South African

Qualifications Authority (SAQA, 2001) defines assessment as the process of collecting evidence and making judgments about learners' achievements or non-achievements in order to arrive at decisions about them. The term *assessment* is sometimes synonymously interchanged with the term *evaluation*. Evaluation is regarded as a part of assessment, and refers to the making of a value judgment about the learner's competence against the assessment criteria (Scriven, 1991).

Apart from diagnostic, formative, and summative assessment, there are other types of assessment. These include the following:

- **Continuous assessment:** This involves assessing learners regularly in a manner that integrates teaching and learning.
- **Continual assessment:** In this learners are assessed repeatedly using the same or a similar technique often for summative purposes. Luckett and Sutherland (2000) aver that in continual assessment, each assessment is treated independently of other assessments and there is limited feedback to learners.
- **Integrated assessment:** This refers to assessing a number of outcomes together, assessing criteria together, and unit standards or learning units together. It also refers to using a number of different assessment methods to assess the achievement of outcomes. Integrated assessment may also mean routine collecting of evidence from a given setting, and also acquiring of evidence from other sources such as supervisors' reports, testimonials, logbooks, and so on.
- **Authentic assessment:** This concerns the assessment of complex performances and higher-order skills in real-life contexts. According to Baron and Boschee (1995), authentic assessment is contextualized, involves complex intellectual challenges, and does not involve fragmented and static bits or tasks. The learner is required to perform real-life tasks.
- **Quantitative assessment:** In quantitative assessment, learning performances are unitized.
- **Qualitative assessment:** This focusses on whether the final mark tells us how well the performance matches the outcome. It does not generalize but rather describes the performance for each individual learner.
- **Extended assessment:** This does not place a time limit on the collection of evidence, depending on the nature of the activity.

With this background of assessment, student learning should be construed as a process by which learners

actively construct their own knowledge and skills rather than being a process of teacher transmission of knowledge to learners (Nicol & Macfarlane-Dick, 2006). This entails active learner engagement in interacting with subject content, teachers (who at this level are facilitators), fellow learners, institutional and other environmental resources and personnel, and in personal reflection. This means that learning outcomes are multi-faceted ranging from the dispositional outcomes in knowledge, values, and attitudes which cannot be easily observed in the learner to the more explicit, easily observable skills that the learner can be seen to apply in interacting with others. Therefore, a one-time assessment of learning outcomes grossly undermines the efficacy of knowledge, values, attitudes, and skills that are seasonally prevalent in the learner, for instance, in the course of cooperating in a group assignment. Moreover, learning is not an end in itself; acquisition of a piece of a certain knowledge, skill, value, or attitude stages the learner for a higher ascent on the rungs of the learning or education ladder. In other words, there is need to shift from assessment of learning (which is commonly referred to as summative assessment of learning outcomes at one point in time, usually at the end of a period of instruction; Sadler, 1998) to assessment for learning (which refers to formative assessment done as a strategy to inform and guide the learner to progress in a desired direction; Sadler, 1998).

However, as noted by Nicol and Macfarlane-Dick (2006), there is unfortunately a slow shift from summative assessment to formative assessment in establishing learning outcomes. Cizek (1997) notes that traditional assessments target a learner's ability to demonstrate the acquisition of knowledge. More often than not, any attempt at formative assessment is tailored towards summative assessment of knowledge rather than specifically intending it to generate feedback on performance to improve and accelerate learning (Sadler, 1998). In other words, the so-called formative assessment strategies are a precursor for a summative measure of knowledge mainly at lower order cognitive domain demanded by a prescribed syllabus. The danger with such assessment is that it is mainly manned by 'teachers' as

opposed to facilitators or educational mentors, which makes it difficult to see how students can become empowered and develop self-regulation skills needed to prepare them for life-long learning (Boud, 2000).

Traditionally, assessment has been almost entirely summative in nature, with a final explanation and the teacher as the sole and unconditional judge. Key players who offer enormous support in the learning process such as the self, peers, and other personnel who are not ‘teachers’ are grossly sidelined in the assessment process and often eliminated from giving feedback to the learner. In addition, the learner is denied the opportunity to actively construct and understand assessment feedback messages through constructive, active, and cooperative learning.

Summative assessment reduces learning to a simple cognitive process involving a didactic transfer of information, which ignores the way feedback interacts with motivation and beliefs (Nicol & Macfarlane-Dick, 2006). According to Dweck (1999), feedback of summative assessment is generally externally provided and so influences how learners feel about themselves, and what and how they learn. On many occasions, learners are pushed to dread the ‘failure’ side of summative assessment and end up indulging in acts such as examination malpractice that promote success in examination scores but subvert learning. Furthermore, summative assessment greatly tasks teachers to set, moderate, administer, and mark tests, and process and disseminate test scores. This, in the face of ever growing number of learners, poses a challenge that predisposes teachers to negative assessment practices that distil examination malpractice. Such practices include setting of items that lack in reliability and validity, defective grading, and aiding learners in examination malpractice.

Given that assessment practices influence whether learners adopt a deep or surface approach to learning, the routine assessment techniques the learner is subjected to need to be those that foster deep learning. Surface learning is easy to achieve, for example, through relying

heavily on an examination-oriented curriculum that rewards learners for ability to reproduce knowledge (Sutherland & Peckham, 1998). Deep learning, on the other hand, is achieved by constructive or alternative assessment practices (Geyser as cited in Fourie-Malherbe & Strydom, 2016). In other words, assessment should be *for* rather than *of* learning only, such that it provides an experience for learners to learn *from*, that is, promotes deep learning. Put another way, learning experiences should be structured in such a way that they are integrated with assessment strategies that enhance learning *from* assessment as opposed to the traditional learning *for* assessment. The traditional and alternative assessment practices should be conflated in curricula, programmes, and units of learning. Whereas traditional assessment imbues the end of a teaching-learning sequence, constructive assessment should instead provide the beginning of learning.

If assessment is used as a tool for learning, it can impact positively on a learner. However, summative assessment is associated with negative backwash effect in which learners are guided in their learning by prompting them to “rely on old examination papers or when they learn to pass a test” (Biggs, 2000, p.141 cited in Fourie-Malherbe & Strydom, 2016). On the other hand, formative assessment has positive backwash effect in which “assessment tasks are deliberately and firmly linked to the learning standards in the curriculum – when assessment is aligned with the curriculum. ... Learners will then focus on knowledge and skills that will help them achieve learning and other outcomes and meet the assessment criteria.”

Borrowing from South Africa, educators need to take into consideration the following socio-economic implications of assessment (Lockett and Sutherland, 2000, p. 99) when designing and implementing curricular:

- Show greater accountability to communities, learners and the government.
- Cater for a diverse student body which bring different life experiences to their studies.

- Assess relevant knowledge and generic skills in unfamiliar contexts so that high quality graduates who can apply their knowledge and skills in a changing world of work are produced.
- Assess the learner's ability to integrate knowledge and skills from a range of disciplines.
- Ensure that students become life-long learners.
- Shift from a content-based to an outcomes-based approach.
- Involve a wide range of assessment methods in order to achieve the above.

The above considerations point to the fact that governments and examining bodies should take assessment as the most powerful lever for educators to influence the way students learn. The educators need to emphasize constructive assessment in which the focus is on assessment for learning. Constructive assessment practice based on assessment for learning should be guided by sound assessment principles. This paper therefore set out to achieve two objectives: (a) to explore the principles of constructive assessment and (b) to suggest the approaches to shifting from summative assessment of the foundational competences to formative, constructive assessment of practical and reflexive competences.

Principles of Constructive Assessment

All stakeholders of education – learners, parents, employers, learning institutions, and the general public – want the assurance that assessment results are authentic. Credibility in assessment is assured through assessment procedures and practices that are guided by certain principles. Smith (1995) suggests the following principles on which constructive assessment should be based at institutional/curricular level, faculty level, and learning units:

1. Assessment should be an integral part of learning that should focus on deep, active learning, and involve high order cognitive skills.

A plethora of studies (*e.g.*, Beattie IV, Collins, &McInnes, 1997; Chin & Brown, 2000; Hay, 2007;

LeCun, Bengio, & Hinton, 2015; Offir, Lev, & Bezalel, 2008; Warburton, 2003) shows that learners will adopt either deep learning or surface learning strategies, which respectively result in high and low quality learning outcomes. According to Fourie-Malherbe and Strydom (2016), a deep approach to learning is characterised by making connections, an active search for meaning (understanding), and development of critical thinking, whereas a surface approach to learning is characterised by an intention to complete the requirements of externally imposed tasks such as tests and examinations and the memorization of discrete items of information. Learners are highly adaptive basing on the assessment practices in force. Therefore, strategies that impart deep learning should be used.

2. Assessment should be an integral part of programme and module design.

All those involved in delivering programmes and modules should agree on assessment strategies and decisions. Assessment tasks should be designed to match the learning outcomes. When planning programmes and modules, assessment strategies should be part and parcel of the decisions.

3. Assessment purpose should be based on determining the assessment methods and techniques.

Diagnostic, formative, and summative assessments have different purposes. Therefore, assessment methods and techniques should be adapted to serve the different diagnostic, formative and summative assessment purposes.

4. Relevant assessment criteria should be identified and applied in the course of assessment.

Assessment criteria are statements that describe the standard to which learners must perform the actions, roles, knowledge, skills, understanding, values, and attitudes in the outcomes (the South African Qualifications Authority [SAQA], 2001). They should be clear, precise, and transparent statements against which successful or unsuccessful performance is assessed. The

criteria need to be linked to the learning outcomes; understandable to all people who use them; clearly expressed, using language that is accessible; explicit about what is expected from the learners; at the right level; and clearly communicated to the learners. Furthermore, the criteria should describe practical, foundational, and reflexive competences which should be assessed as a unit, and not as separate entities. Norm-referenced assessment should be discouraged in an outcomes-based approach to constructive assessment.

5. Assessment processes should be reliable.

Reliability is about consistency. Assessor bias based on learners' demographic characteristics should not be tolerated. Different assessors should assess learners' work to the same standard and consistency should be observed throughout the assessment process. Learners should be thoroughly briefed about the criteria and outcomes. Test items should be moderated, double-marked, sampled, or cross-moderated to ensure reliability. Administration of tests should be thoroughly proctored to minimize malpractices.

6. Assessment tasks should be valid.

Rather than assessing what is easy to assess, assessment methods and processes must assess what they set out to assess. In order to achieve validity, assessors should state clearly what outcomes are being assessed and use an appropriate approach to and method of assessment.

7. Assessment should be transparent.

In order to have confidence in the assessment system, all parties concerned need to understand the system and be assured that it is well planned, that it works in practice, and is regulated properly. This involves telling learners well in advance what is expected of them, providing a clear brief with indicators of what does and does not constitute satisfactory performance, providing clear guidelines for handing work in, offering equivalent levels of support to learners prior to submission, indicating how

failure is communicated and can be redeemed, and providing access to assessment regulations and information.

8. Assessment should be fair.

Fairness in assessment means that learners at all levels are accorded an unbiased opportunity to be treated equally and fairly. This implies that assessors should adhere to set criteria; materials on which assessment is based do not privilege particular learners; there is clarity in terms of what is being assessed; learners should not be identified by their written work; there is no comparison of one learner's work with that of others; there are systems in place to regulate and moderate assessment, as well as reassessment and appeal mechanisms; the facilities and practices should enable all learners to complete tasks on the same basis; assessment methods should be varied; and equal support should be given to learners by staff.

9. Assessment tasks should be practicable.

It is important to ensure that the tasks set can be achieved in the time available and within the existing constraints. These might include the facilities available to learners, the number of learners using the same equipment, books, the number of support staff available, and the like.

10. Assessment workload should be realistic.

Over-assessment, that is, too many tests or assignments is an abuse of assessment. The number of assessments should be regulated by ensuring the following:

- Staggered deadlines to spread the workload evenly.
- Coordinated assessment timetables.
- Assessment schedules published in advance.
- Administrative assistants to support assessment.
- Recognition by academic managers through pay and other benefits that assessment is an integral part of an academic's workload.
- Streamlining assessment by using numbered lists, ICT, etc.

11. Assessment should include a wide range of approaches and methods.

Assessment methods must be appropriate. A diversity of assessment methods should be applied because some methods may advantage certain learners while disadvantaging others. Different methods are required to assess different outcomes, and different kinds of competencies. Different methods are also required to assess process and product. Diversity promotes effective learning in different contexts and prevents boredom. Different approaches to consider in assessment include self-assessment, peer assessment, group-based assessment, computer assisted assessment (CAA), and workplace-based assessment. Both formative and summative assessment techniques should be applied within a given unit.

12. Assessment should provide feedback to support the learning process.

While summative assessment of learners' progress has to be made, feedback should be used to enhance learning. Assessment for learning which is mainly formative assessment requires constructive and developmental feedback to guide and direct the learners to improve.

13. Assessment should be integral to quality assurance procedures.

Assessment has a major influence on what learners learn, how effectively they learn, and consequently on the quality of their entire learning experience. Therefore, national and institutional quality assurance bodies should provide the policies and procedures relating to the conduct and management of assessment. Issues to address should include using a sufficient diversity of assessment methods; encouraging deep, active, and reflective learning through assessment; and providing opportunities for learners to apply their knowledge and understanding to different problems and contexts.

Approaches to Constructive Assessment: Harnessing the Desired Shift from Assessment of Learning to Assessment for Learning

In shifting from assessment of foundational competences to that of practical and reflexive competences, constructive rather than the traditional summative assessment should be undertaken. The process of constructive or alternative assessment provides answers to the questions of why assess, what to assess, who assesses, when to assess, and how to assess. This paper looks at this process by considering each of the questions in turns. It further provides insights into grading and providing feedback to the stakeholders.

Why assess?

The main purpose of assessment is to generate information on which decision making can be based. Decisions are often classified as institutional or individual decisions. In education, institutional decisions are those made by the management of the institution concerning the learners. Individual decisions are those the individual makes about himself or herself (Mehrens & Lehmann, 1991). Mehrens and Lehmann provide suggests a further classification of educational decisions as instructional, guidance, administrative, or research/programme evaluation decisions.

Instructional decisions aid the teacher to provide knowledge concerning the learner's entry behaviour; setting, refining, and clarifying realistic goals for each learner; evaluating the degree to which the objectives have been achieved; and determining, evaluating, and refining the instructional techniques. Meanwhile instructional decisions aid the learner by communicating the teacher's goals, increasing motivation, encouraging good study habits, and providing feedback that identifies strengths and weaknesses.

Guidance decisions provide learners with information on vocational planning and personal growth. Constructive assessment diversified as aptitude tests, achievement tests, personality inventories provide the learners with

data about the significant characteristics and help them develop realistic self-concepts.

Administrative decisions include selection, classification, and placement decisions. In selection decisions one decides whether to accept or reject a person for a particular programme or treatment. In classification one decides the type of programme or treatment. In placement decisions one decides the level of treatment. Administrative decisions are also involved in areas such as curriculum planning, hiring and firing of teachers, and career ladder or merit pay decisions.

Research and programme evaluation decisions cut across the three preceding types of decisions. Instructional, guidance, and administrative decisions may all be based on research. Research decisions are made whenever information is gathered as a prelude to the decision making. Knowledge in 'why assess' affirms that assessment should be conducted for practical and reflexive rather than sheer foundational reasons.

What to assess?

Traditional assessment in outcomes-based education and training emphasizes the assessment of outputs and end products or outcomes. The learner's applied competence is neglected. Applied competence refers to the union of practical competence, foundational competence, and reflexive competence. Practical competence is the demonstrated ability, in an authentic context, to consider a range of possibilities for action, make considered decisions about which to follow, and to perform the chosen action. Practical competence is grounded in foundational competence where the learner demonstrates an understanding of the knowledge and thinking that underpins the action taken. Practical and foundational competence are integrated through reflexive competence, in which the learner demonstrates the ability to integrate or connect performance and decision making with understanding, and with an ability to adapt appropriately and responsibly to change in unforeseen circumstances and to explain the reasons behind these adaptations (SAQA, 2001).

Learners have to demonstrate that they have acquired and mastered the competences as described in the outcomes and assessment criteria. In other words, evidence of acquired competence is required. Evidence can come from a variety of sources. It is the responsibility of the assessor to ensure that enough and appropriate evidence has been collected to make an accurate judgment about the learner's competence. The following questions on the quality of evidence can be used as guidelines when assessment is planned:

Does the evidence relate to the outcomes and criteria to be assessed?

Can the evidence be attributed to the learner? Is the evidence the learner's own work?

Is there enough evidence to meet all the criteria needed to certify the learner as competent?

Is the assessor confident that the performance can be repeated?

Is the evidence related to current competence?

The best kind of evidence is *direct evidence* which can be observed directly by the assessor. This is the most valid type of evidence. Examples of direct evidence include direct observation of assessment tasks, oral and written questioning. Indirect evidence is evidence produced about the learner from other sources. These are usually third party sources such as testimonials, training records, medals, prizes and trophies. It is often necessary to corroborate these forms of evidence given that they could be easily faked. Historical evidence tells the assessor what the learner was capable of doing in the past. It is the least valid type of evidence because it does not prove current competence. Examples of historical evidence include projects and portfolios, training records, and testimonials.

Who assesses?

Constructive assessment should go beyond the educator assessing the learner. It should include self-assessment by the learner and peer assessment where a group of learners assess each other. Self-assessment may be used for many activities and purposes ranging from informal activities to promote reflective learning (self-reflection) to formal strategies of self-assessment (Fourie-Malherbe &

Strydom, 2016). Learners should be enabled to develop self-assessment skills so that they are able to judge the effectiveness of their own performance accurately.

Self-reflection also known as metacognition or reflection serves different purposes. It can help the learners to plan a task, monitor their progress, and evaluate their accomplishments. It requires active learner engagement. It is triggered by an unusual or perplexing situation or experience; involves one's responses, beliefs, and

premises in the light of the situation at hand; and results in the integration of a new understanding of one's experience.

Peer assessment refers to assessment of the learner by his or her peers. It provides a valuable means of involving learners in their own and others' learning. It is productive in terms of improved learning; learners can learn a great deal from each other. It is particularly helpful with increasing number of learners as the availability of assessor feedback decreases (Brown, 2005).

When to assess?

| When to assess? | Why assess? | What kind of assessment? | Who assesses? |
|---|---|-------------------------------------|--|
| At the beginning of a new section of work | To find out what the learners know about a topic in order to plan learning activities. | Diagnostic (or baseline) assessment | Teacher |
| Continuously (during learning activities) | To monitor and support learning; both teachers and learners receive immediate feedback. | Formative assessment | Teacher, learner (self- and peer assessment) |
| When learners appear to be having difficulty with some of the work. | To establish the cause of the learning difficulties. | Diagnostic assessment | Teacher |
| At the end of a section of work or a learning unit. | To establish and record overall progress of learners towards desired outcomes. | Summative assessment | Teacher, assessor, external assessor. |

Source: Fourie-Malherbe &Strydom (2016)

How to assess?

The process of constructive assessment should include the following activities:

Collecting evidence on learners' performance by means of different assessment methods and techniques.

Comparing the evidence to the assessment criteria.

Judging the learners' performance as pass or fail, competent or not yet competent in summative assessment.

Making instructional and/or learning decisions based on the learner's progress, addressing and the like, to enable the learner to try again and to improve on the previous performance.

Constructive assessment can be done formally or informally. Informal assessment is done as part of learning during interaction with learners. Unfortunately, formative assessment is often informal. It usually takes a few minutes and the assessor can respond almost immediately without quantifying the results. However, if formative assessment is to be constructively used to enhance learning, it requires careful planning to integrate it with learning and teaching. Formal assessment, on the other hand, is planned in terms of methods, as well as in terms of when it will take place. The learners are informed ahead of time that they will be assessed. Summative assessment is often regarded as structured formal assessment. It is usually more time-consuming and requires planning and careful organization.

It is clear that traditional methods of assessment which consider the teacher or assessor to be the only legitimate are no longer adequate to meet the demands and expectations of the 21st century job market. These methods include the usual one to three hour examinations, essays, and classroom-based pen-and-paper tests. These traditional approaches need to be creatively adapted to tap the 21st century skills so as to achieve the sustainable development goals among the learners. Variations of the old approaches include the following:

Computer assisted assessment (CAA)

Computers can be used to provide detailed, individualized feedback to large numbers of learners. CAA is usually based on a mastery learning model from which learners receive immediate feedback, and can

repeat or progress at their own pace. This approach is most suited for multiple choice questions, but it can also be adapted creatively with a range of other approaches and methods (Luckett & Sutherland, 2000).

Workplace- and community-based or studentship assessment

As opposed to the learning and testing of knowledge and theory in isolation from contexts of application, in workplace assessment learners immerse themselves in the real work world. It is beneficial for the development of professional skills and competences because it gives the learners an opportunity to apply their knowledge in the field as supervisors or mentors offer support supervision face-to-face to the learners (Luckett & Sutherland, 2000).

Group-based assessment (group assessment)

This approach recognizes that all learning takes place in a social context and that professional identity is best developed through interaction with a community of professionals. In this approach, learners are required to work in teams and they may be assessed as a group or individually. A combination of both methods of assessment is preferred as team spirit should be enforced, yet learners will graduate as individuals. This strategy of assessment requires careful planning and management of the different roles and contributions of the learners in a group. The planning should take care of social loafing and the difficulties involved in fairly sorting out the different contributions learners may have made to the product of the group.

Learning contracts

This is a way of ensuring that learners take responsibility for and manage their own learning through developing and maintaining their own achievement records and portfolios. These contracts can serve as detailed, qualitative records of learner achievement for prospective employers. Learning contracts focus on the development of generic skills or critical outcomes across the programme.

How to Grade Learners

Grades are the currency of the classroom. As students learn, they earn grades as their basic extrinsic reward. These grades have important consequences for the judgment of learners by others (e.g., parents, employers, and assessors) and themselves (e.g., as they develop their self-concepts). Grades are symbols that carry meaning

and are used in communicating with others. Therefore, assessment needs to cater for how to combine the separate performance scores from different topics judged at various levels of understanding into one final grade.

How to Provide Constructive Feedback

Constructive feedback to learners includes seeing their marked scripts, assignments, and other tasks, and should not be limited to a grade on a list. Learners should have access to their assessment scripts and the assessor should be available to address learner queries, adjust marks, and address appeals during this period. Improper disclosure of the learners' assessment results (e.g., on a notice board) should be avoided.

Good feedback that contributes to learning should link the learner's goals and intentions rather than the assessor's. Criticism should be set in a context of appreciation of and respect for the learner and his or her need for self-confidence and self-esteem. The following are some of the principles of constructive feedback to individual, subgroup, or whole group learners:

Feedback should be timely to be meaningful. Late feedback is a waste of everybody's time.

Feedback should be quick and swiftly delivered, for example, through e-mails and CAA.

Feedback should include more than ticks. It should be detailed and descriptive.

Feedback should be given before work is assessed, in other words, before grades or scores are allocated.

A score or a grade alone without any qualitative justification does not support learning.

Feedback should be honest, but also motivating.

It should be realistic around issues that the learner can grasp and act upon for improvement.

Negative feedback can damage a learner's motivation to learn. Therefore, the assessor should consider the affective factors involved in giving and receiving feedback.

Unhelpful, judgmental feedback that focusses on the learner as a person should be avoided. The assessor should distinguish between the person who produced the work and the work itself, and comment on the work.

Judgments that leave no room for improvement should be avoided (e.g., *You can do better than this*). Rather the assessor should write constructive comments that will tell the learner how to improve.

Feedback should give positive comments before negative comments.

Conclusion

A paradigm shift from traditional, summative assessment of learning to constructive assessment is inevitable if assessment is to enhance learning. Therefore, the assessment practices in institutions of learning should be periodically reviewed to focus on assessment for learning so as to maximize the positive effects of assessment in the face of ever-increasing resources. Constructive assessment principles should be followed to enhance assessment for learning. Assessment for learning should be premised on a rationale of shifting from foundational to practical and reflexive competences of the learners. The approach taken should answer questions related to why assess, what to assess, who assesses, when to assess, how to assess, how to grade, and how to give feedback to the stakeholders.

Recommendations

Institutions should have an assessment policy that sets out clear and effective procedures for its implementation. The policy and its procedures should ensure academic and professional standards in the design, approval, implementation, and review of assessment strategies for programmes and courses. In this case the teaching staff should be appropriately trained, experienced, and competent to assess. This should be catered for both at pre-service and in-service levels. Therefore, teacher

training institutions should adopt interactive constructivist pedagogies in a bid to graduate teachers with appropriate constructive assessment approaches.

Constructive assessment should be conducted within a framework of rules and regulations governing it. Assessment practice manuals should specify all the vital assessment criteria including formative assessment, provision of feedback to learners, weighting of class

marks, disciplinary and appeals procedures, regulations for marking, and grading.

References

- Authority, S. A. Q. (2001). *Criteria and guidelines for assessment of NQF registered unit standards and qualifications*. SAQA.
- Baron, M. A., & Boschee, F. (1995). *Authentic Assessment: The Key to Unlocking Student Success*. Order Department, Technomic Publishing Company, Inc., 851 New Holland Avenue, Box 3535, Lancaster, PA 17604; toll-free tele.
- Beattie IV, V., Collins, B., & McInnes, B. (1997). Deep and surface learning: a simple or simplistic dichotomy? *Accounting Education*, 6(1), 1-12.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167.
- Brown, S. (2005). Assessment for learning. *Learning and teaching in higher education*, (1), 81-89.
- Chin, C., & Brown, D. E. (2000). Learning in science: A comparison of deep and surface approaches. *Journal of Research in Science Teaching*, 37(2), 109-138.
- Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroads. *Handbook of classroom assessment: Learning, adjustment, and achievement*, 1-32.
- Dweck, C. (1999). *Self-theories: their role in motivation, personality and development*. Philadelphia, PA: Psychology Press.
- Fourie-Malherbe, M., & Strydom, S. (2016, September 12-16). *Becoming a lecturer: Teaching and learning in higher education*. African Doctoral Academy Joint Doctoral School of Stellenbosch University and Makerere University.
- Hay, D. B. (2007). Using concept maps to measure deep, surface and non-learning outcomes. *Studies in Higher Education*, 32(1), 39-57.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- Luckett, K., & Sutherland, L. (2000). Assessment practices that improve teaching and learning.
- Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology*. Chicago, IL: Holt, Rinehart, and Winston.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2), 199-218.
- Offir, B., Lev, Y., & Bezalel, R. (2008). Surface and deep learning processes in distance education: Synchronous versus asynchronous systems. *Computers & Education*, 51(3), 1172-1183.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in education: principles, policy & practice*, 5(1), 77-84.
- Scriven, M. (1991). *Evaluation thesaurus*. Sage.
- Smith, B. (1995). *The assessment of student learning*. Paper presented at a conference in Hong Kong on improving university teaching.
- Sutherland, L., & Peckham, G. (1998). A Re-Appraisal of Assessment Practices in the Light of the South African Qualifications Authority (SAQA) Act. *South African journal of higher education*, 12(2), 98-103.
- Warburton, K. (2003). Deep learning and education for sustainability. *International Journal of Sustainability in Higher Education*, 4(1), 44-56.

Teaching Approaches, Interest and Performance of Students in Physics at Uganda Certificate of Education: The Relevance of Constructivism Teaching Learning Theory

Dr. Peter Kagaba and Dr. Grace Milly Kibanja

kagabapeter558@gmail.com

Uganda

Institution of Affiliation: Uganda National Examinations Board

Email: kagabapeter558@gmail.com and kibanja@gmail.com

Telephone Contact: +256772450059, +256701114198 and +256776811177, +256701811177.

Abstract

The poor performance of candidates in science subjects has raised great concern from stakeholders over a longtime. Some of the causes of this problem may be attributed to the teaching approaches of science subjects in schools and the students' interest in the subjects among other factors. The purpose of the study was to investigate the teaching approaches of Physics in schools and examine their influence on the interest of students at Uganda Certificate of Education (UCE). A quasi- experimental survey design was used. The scope included six schools selected from three districts in mid-Western Uganda. Sampling of schools was purposive to cater for homogeneity and capture the average and low performing schools in those districts basing on the 2010 UCE results. Questionnaires were administered to students and teachers and classroom observations were done to capture the teaching/learning processes. A census sampling approach for 600 students was used of which 420 students fully responded to all the instruments. Other respondents included; six (6) teachers, twelve (12) subject experts of Physics (SEs) who were sampled purposively. A two-factor split-plot (one within-subject factor and one between-subject factor) ANOVA was conducted. Findings revealed that teaching approaches significantly influence interest, learning, and performance of students in Physics at UCE. It was therefore recommended that teachers of Physics should embrace the principles of constructivism teaching/learning theory and enlighten students on their importance in learning. All teachers of Physics at UCE should be graduate teachers trained to teach Physics. Teachers of Physics at UCE should be adequately motivated by Government and other stake holders to enable them teach effectively. Further research may be undertaken to investigate the influence of constructivism teaching approaches with regard to gender and school type.

Key words

Teaching Approaches: Skills and methods used to impart knowledge and skills to learners.

Interest: A feeling of wanting to or a quality that attracts one's attention and makes him or her want to be involved in something.

Introduction

Background

This presentation is part of a thesis presented for the award of the author's PhD degree of Mbarara University of Science and Technology on 21st December 2016.

Among the objectives dwelt on was teaching approaches and interest of students in Physics at Uganda Certificate of Education (UCE).

The current trends in education put great emphasis on science development and accountability in schools. To achieve this, the approaches and quality of teaching, and students' interest are potential factors that cannot be under estimated (Angella, David, & Marvis, 2001). The role of science in this modern era of technology is wide and profound (Awodun, Aladejana, & Oni, 2014). Science is the basis of general development and current digital revolution (UNAS, 2008).

Science subjects in this respect include; Physics, Chemistry, Biology and Agriculture but the study dwelt only on Physics because it cuts across most science subjects and it is the most utilized basic science subject in most technology and technology-related professions (Olarinmoye, 2000). Therefore, there is increasingly high demand for teachers to adopt models and methodologies which yield high learning outcomes especially in Physics.

In several countries, research studies and experience have recorded a number of teaching approaches of science subjects including Physics with their advantages and disadvantages. The objectives of secondary education in Uganda, as proposed by the Education Policy Review Commission (1989), are to enable individuals to develop basic scientific, technological, technical, agricultural and commercial skills required for self-employment. Ministry of Education and Sports (2014) therefore put in place several policies and programs and interventions to fulfil the recommendations of the Commission. These interventions include; Policy of compulsory science at lower secondary (UCE), affirmative action of government sponsorship at University and tertiary institutional level, construction and equipment of laboratories at secondary and primary teachers' colleges and providing science kits to secondary schools, recruiting secondary science teachers, organizing in-service training of science and Mathematics teachers, and training of teachers in ICT skills (UNAS, 2008).

Despite government support in uplifting the teaching and learning of Physics, performance in it has been poorer than in other science subjects over a number of years. The trend of candidates' performance in Physics for five years (2008-2012) may be seen in Table1.

Table 1: *Students' Percentage Failure Rates per Paper in Physics at UCE 2008-2012*

| Years | 2012 | | | 2011 | | | 2010 | | | 2009 | | | 2008 | | | |
|--------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|---------|
| Papers | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | Average |
| % | 46.9 | 63.6 | 44.1 | 41.7 | 50.9 | 38.7 | 44.5 | 61.1 | 54.4 | 56.7 | 67.2 | 54.4 | 45.4 | 55.4 | 70.6 | 55.6 |

Source: UNEB records 2008-2012

Table 1 indicates that the performance of students per paper in Physics at UCE for the five years was not satisfactory. A combined average failure rate of 55.6% in a national examination raises a lot of concern as to what should be done to alleviate it. Previous studies have established several factors that contribute to students' performance which include teaching approaches and interest of students in the subject taught (Ainley, 2008; Leong, 2005; & Worse, 2007). Teaching approaches may enhance interest in learning and thus influence other factors like performance (Ainley, 2008). It was for these arguments that this research on teaching approaches and interest of students in Physics was carried out.

Problem

Whereas there is a lot of emphasis on teaching and learning science subjects and Physics in particular by Government of Uganda, the performance of candidates in Physics at Uganda Certificate of Education has continued to be low. The danger is that, if this scenario persists, the country may not have competent scientists to sustain its technological development (Zavuga, 2008). Effort must be taken to arrest the threat. Hiilya (2010) articulates on the unprecedented interest in reforming pedagogical practices in Sub Sahara Africa including Uganda in the last two decades. The reform efforts are characterized by a move from teacher-centered instruction to child-centered pedagogy (CCP). Medina, James, and Lawrence (2010) also observed that teachers in schools hardly employ child-centered approaches to teaching, which may explain the poor quality of education in Universal Primary Education (UPE) and Universal Secondary Education (USE) schools. They recommended for the need to focus more on teacher supervision to compel teachers to attend to their duties and use child-centered methods of teaching. Much as several studies have been done on the topics elsewhere, no specific study of the kind in Uganda has been done. Even then, the content gap of the two variables put together has not been covered. There was need, therefore, to investigate the influence of teaching approaches to students' interest in Physics.

Purpose

The purpose of the study was to investigate the influence of teaching approaches of Physics on students' interest with a view to recommending suitable teaching approaches that can improve on students' interest in Physics.

Objectives

The objectives of the study were to:

1. Find out the most prevalent teaching approaches of Physics at UCE, challenges and coping strategies while applying them.
2. Assess the extent to which teaching approaches of Physics at UCE influence students' interest in Physics.

Hypothesis

The research tested the following hypothesis:

There is no significant difference between the mean interest scores of students of Physics taught using interactive approach (Experimental Group E) and those taught using other approaches (Control Group C).

Scope

The study covered the science subject of Physics at UCE. It focused on the teaching approaches being used by teachers of Physics during the teaching process and their influence on interest of students. Students of senior three sampled across the country participated. Teaching approaches meant; the procedure taken by teachers to impart knowledge and skills to learners. Interest meant; the state of wanting or the curiosity or attention one puts to learning about a subject. For purposes of proximity, and homogeneity, a sample of low and average performing secondary schools in one area of mid-western Uganda participated. The research was carried out within one year, 2014.

Significance

The findings of the study may help teachers to employ suitable teaching approaches while teaching Physics at UCE and beyond. Students of science subjects may benefit from the recommendations given as it may enhance their interest in learning Physics. It is expected that the Directorate of Education Standards (DES), will

strengthen the monitoring of the teaching-learning process in the class room to establish challenges of teachers and enhance their teaching efficiency. The Ministry of Education and Sports through its respective institutions may accordingly cater for adequate teaching of Physics. The results of the research are also an additional aspect of literature to the body of knowledge on teaching approaches and interest of students in Physics at UCE

Conceptual Framework

In this research, teaching approaches was an independent variable against a dependent variable i.e. interest. However, there are several other extraneous variables like class size, school type, gender etc. that did not appear in the conceptual framework and could compromise the reliability of the results of the study. These were controlled by use of Quasi-experiment (Donna, 1993).

Students' interest depends on how teachers articulate themselves in delivery of subject matter, on the motivation they get out of the teaching-learning process and the opportunity given to them to participate in the learning activities (Leong, 2005). As they all participate in learning activities, they share knowledge which in turn may create interest in the subject taught.

Theoretical Framework

There are theories that relate teaching, interest and learning variables to students' performance. The Constructivism (teaching /learning) theory is a theory that explains how knowledge is constructed in the human being. The theory bases teaching and learning on applications of discovery learning, hands-on, experimental, collaborative, project- based and task-based (Hein, 1991). Biggs, 2002; Bodner, 1986; and Appleton, 2008; mention that recent trends in science education research have attempted to marry constructivist views of learning with actual practice of teaching science subjects, Physics inclusive. The relevant principles in constructivism learning theory to this research were:

1. The crucial action of constructing meaning is mental:

The learning happens in the mind. Physical actions and hands-on experience are necessary for learning. There is need to provide activities that engage the mind and hands.

2. Motivation is a crucial component in learning:

Motivation takes several forms, one of which is interest. Even with the best teaching approaches, the learner may not effectively perceive anything taught unless he or she has interest in it. Do the teaching approaches enhance any interest in students of Physics?

3. Learning is a social activity:

Education should not isolate learners from social interaction. Progressive education recognizes the social aspect of learning and uses conversation, interaction with peer.

The constructivism teaching- learning theory (CLT) was used to guide this research. CLT is described as a theory of knowledge with roots in philosophy and psychology (Von, 1995). The CLT refers to the idea that learners construct knowledge for themselves where each learner, individually and socially constructs meaning as he or she learns (De Vries, 2002).

Methodology

For the research paradigm, the study adopted a quantitative approach which uses Post-Positivist claims of knowledge. Some of the Post-Positivist features include cause and effect thinking and the test of a theory among others (Crestwell, 2009; Crestwell & Plano, 2007). The study used a quantitative approach based on the quasi-experimental study design triangulated with survey design. Quasi-experimental design was used because of handling intact groups of students and survey was used to reach a large size of the population (Amin, 2005). It was carried out using six schools, six teachers of Physics, 12 subject experts (SEs) and 600 students. Table 2 shows the sample selections and their characteristics.

Table 2: Population size, Sample Size, Sample Selection and Sample Characteristics

| Category | Study Population | Sample | Sampling Technique | Gender | | Instrument |
|---------------------|------------------|--------|--------------------|--------|--------|----------------|
| | | | | Male | Female | |
| Subject Specialists | 24 | 12 | Purposive | 10 | 2 | Questionnaire |
| Teachers | 30 | 6 | Purposive | 5 | 1 | Questionnaire |
| Students | 600 | 600 | Census | 342 | 258 | Questionnaires |
| Schools | 30 | 6 | Random | Mixed | Mixed | - |

In Table 2, all the schools were of mixed gender. Four of them had single streams of 50-60 students. The remaining two schools were bigger with three streams each having 50-60 students. The objectives of the study, however, did not take into consideration of many sample characteristics.

The instruments used for data collection included an open-ended questionnaire for SEs. This questionnaire asked SEs to give their opinion on teaching approaches and on students’ interest in Physics, then to propose 3 suitable teaching approaches for teachers of Physics. A 20-item questionnaire was designed for students to estimate their interest in Physics and a 10-item questionnaire for teachers seeking their opinion on teaching approaches, students’ interest in Physics and challenges while teaching Physics. All the questionnaires were validated by research experts. The students’ questionnaire was tested for content validity whose content validity index was 0.83 and acceptable (Amin, 2004). The reliability was tested using the test-retest reliability index. The results were analyzed using the Pearson Product –moment Correlation Index. The Correlation Index 0.78 was acceptable (Amin, 2004).

The sampled schools were randomly given codes to conceal their identity. Schools S₁-S₃ belonged to experimental group E and S₄-S₆ belonged to control group C. The recommended teaching approaches by SEs were given to teachers in group E to strictly apply while teaching during the research as an intervention. Teachers in group C were not given any specific teaching approaches but were told to teach normally during the

research. All teachers in group E and C taught for a period of one year using the same syllabus for senior 3 provided by the National Curriculum Development Centre (NCDC).

At the beginning of data collection before any teaching began in the first school term, the students’ questionnaire on their interest in Physics was administered to them all. The objective was to establish their initial respective levels of interest in Physics. In the middle of each term, classroom observation visits were carried out to establish whether teachers in group E were appropriately applying the recommended intervention. At the end of term 3, the questionnaire on students’ interest was re-administered to all groups E and C. The objective was particularly to establish the impact of the intervention on group E.

Ethical issues were noted but taken care of accordingly. Questions on personality and identity were avoided. Schools were coded to provide anonymity and conceal places of origin. Respondents were neither asked to sign on any document nor coerced to participate in the research. Quoted authors were also recognized by citation.

Data analysis for objective 1 was done qualitatively while that for objective 2 was done quantitatively. This quantitative data analysis was done using the SPSS 18.0, verified at 0.05 level of significance. Statistical verification of the hypothesis was by the two –factor split plot ANOVA. Ethical considerations were technically handled and limitations were controlled for.

Results

.Objective One: Find out the most prevalent teaching approaches of Physics at UCE, challenges and coping strategies while applying them.

Objective one was to find out the most prevalent teaching approaches of Physics at UCE. Data for this first part of

objective 1 was provided by SEs and teachers from their respective questionnaires. The responses were qualitatively handled and ranked according to frequency of responses. The six most common teaching approaches mentioned by the teachers and subject experts are found in Table 3.

Table 3: Responses from SEs and teachers of Physics most common Teaching Approaches

| S/N | Identified Teaching Approaches | No. of Teachers and SEs (18) | % of cases |
|-----|---|------------------------------|------------|
| 1 | Group work for both practical and theory assignments | 18 | 100.0 |
| 2 | Hands-on-activities | 18 | 100.0 |
| 3 | Constructivism teaching approach which provides activities that engage the mind and hands | 17 | 94.4 |
| 4 | Problem solving where details and apparatus are given | 16 | 88.9 |
| 5 | Guided individual experimentation and students performing the experiment in groups | 15 | 83.3 |
| 6 | Lecture teaching approach | 4 | 22.2 |

From Table 3, all respondents 18 (100.0%) identified group work for both practical and theory assignments and hands-on activities as commonest teaching approaches. Constructivism teaching approaches which provide activities that engage the mind and hands was identified by 17 (94.4%) of the respondents. This approach is the one the research applied for the experimental group E to derive its influence on interest in Physics. Problem solving where details and apparatus are given was mentioned by 16 (88.7%) of the respondents. Guided individual experimentation and students performing experiment in groups was given by 15 (83.3%). Lecture teaching approach was least mentioned by 4 (22.2%) of the respondents who were all teachers.

While applying these teaching approaches, teachers of Physics indicated that they experienced several challenges show with their respective frequencies: large

numbers of students in class 6 (100.0%), inadequate teaching and learning materials 6 (100.0%), students' negative attitude towards and no interest in Physics 5 (83.3%), inadequate classroom space (83.3%), some concepts seem abstract and hard for students to understand 5 (83.3), wide syllabus 4 (66.7%), weak students 3(50.0%), and less hands-on activities 2 (30.7%).

Despite the above challenges, the teachers provided specific coping strategies to facilitate their teaching process. They are shown with their respective response frequencies as: demonstration approach 6 (100.0%), use of many shifts during practical lessons 6 (100.0%), instilling hope and interest in Physics 4 (66.7%), remedial lessons 4 (66.7%), persuade school authority to stock the library 4 (66.7%), improvising with relevant available materials 3 (50.0%).

Objective two: Assess the extent to which teaching approaches of Physics at UCE influence students' interest in Physics.

This objective was translated into the hypothesis below due to the quantitative nature of analysis of its data.

Hypothesis One: There is no significant difference between the mean interest scores of students of Physics taught using inquiry-based or interactive teaching approaches and those taught using any other teaching approaches.

Data for this hypothesis was collected from the students' questionnaire on their interest in Physics. The same

questionnaire was administered twice, at the beginning and end of the research for both students in group E and C. The mean interest scores of students in groups E and C were comparatively presented in table 4.

The hypothesis was tested using the two factor split-pilot Analysis of Variance (ANOVA) due to repeated measurement of variables. To get the difference between the mean interest scores of students in experimental group E and control group C, further analysis was carried out by testing the homogeneity of variance and multiple comparisons. Table 4 gives the estimated marginal means, the descriptive statistics and Levine's Test of Equality of Error of Variances.

Table 4: Descriptive statistics and Levine's Test of Equality of Error of Variance

| Test | Group | Mean | SD | F | Sg |
|------|--------------------|-------|-------|-------|-------|
| 1 | Experimental (E) | 73.17 | 11.21 | 10.33 | 0.001 |
| | Control (C) | 71.72 | 15.13 | | |
| | Total | 72.62 | 12.84 | | |
| 2 | Experimental (E) | 79.00 | 8.49 | 10.69 | 0.001 |
| | Control (C) | 74.17 | 11.49 | | |
| | Total | 77.17 | 10.01 | | |

Table 4 shows that the mean interest scores in the first test, T₁, and the second test, T₂, for experimental groups E, are higher than that of control group C both at the beginning and end of the study i.e. (73.17,79.00) against (71.72,74.17) respectively. It is noteworthy that the

experimental group E mean interest scores both at the beginning and end of the study are higher than the total mean scores for both groups E and C i.e. 73.17 against 72.62 and 79.00 against 77.17 respectively.

A Two Factor Split-Pilot ANOVA was then carried out due to repeated measures as shown in Table 5.

Table 5: Two Factor Split-Pilot ANOVA for experimental and control groups

| Test 3 | Type | 111 sum of squares | Mean square | df | F | Sig |
|-----------------|------|--------------------|-------------|----|-------|-------|
| Within subject | | 569.41 | 569.40 | 1 | 11.81 | 0.001 |
| Contrasts | | | | | | |
| Between subject | | 1961.49 | 1961.49 | 1 | 9.27 | 0.002 |
| Effects | | | | | | |

Table 5 indicates that there is a significant difference between group means because the p-values 0.001 and 0.002 are less than the set significant level of 0.05. Thus the null hypothesis is rejected implying there is a significant difference between the mean interest scores of students taught using inquiry-based or interactive teaching (Experimental Group E) and those taught using any other teaching approaches (Control group C).

Discussion

The study sought to; establish the influence of teaching approaches on interest of students in Physics at UCE. The discussion of the findings was presented objective by objective.

Objective 1: To find out the most prevalent teaching approaches of Physics at UCE, Challenges and coping strategies while applying them

The major sources used to establish the most prevalent teaching approaches of Physics were the SEs and teachers. A number of them were relatively similar but SEs specifically brought out teaching approaches of *constructivism, guided individual experimentation, problem solving* and *group work* type for both practical and theory tasks. The teachers mentioned *hands-on-activities* and *lecture teaching approach*. Experimentation and hands-on-activities have similar tasks and may be taken to be the same. Besides the lecture teaching approach mentioned by teachers, all the SEs' recommended teaching approaches of Physics seemed to have been the correctly chosen effective intervention for this research. The students in experimental group E whose teachers were guided to adopt them in the teaching process, relatively improved in interest in Physics. The research was developed on the principles of the Constructivism Teaching/Learning Theory which impresses on the need for self-construction of meaning, motivating the learner during the learning process, sharing of knowledge and evidence of practical applications (Appleton, 2008). In general, all the teaching approaches mentioned by SEs and teachers positively influence interest of students in Physics.

The teaching of science may not have a specific approach; it is situational depending on the surrounding factors (FAWE, 2007). The factors may include inadequate teaching/ learning materials, huge numbers of students in class, inadequate space for learning but whatever approach is adopted will most likely fall among those established by this research.

The teachers, for example, mentioned *lecture teaching approach*, due to huge numbers of students handled from inadequate space in classrooms. Worse (2007), and other researchers recommend constructivist teaching and differential instruction whose requirements are far-fetched for several upcountry USE and private secondary schools in Uganda. However, by all standards and irrespective of the hardships schools meet, the appropriate teaching approaches should be learner-centered.

The second part of objective 1 sought to establish challenges teachers of Physics meet while teaching plus coping strategies they apply. Among the major challenges the teachers gave were: large numbers of students in class, inadequate teaching and learning materials, students' negative attitude towards Physics, inadequate space, concepts that seem abstract, wide syllabus, weak students, and less hands-on activities. These challenges are normal in any developing country but they may rightly be mitigated by the coping strategies the teachers mentioned. Major among the coping strategies were; demonstration, shifts, instilling hope and interest, remedial lessons, and improvisation. The teachers of Physics, therefore, need to be adequately trained and qualified to have appropriate coping strategies whenever challenges arise in the classroom.

Objective 2: Assess the extent to which teaching approaches of Physics at UCE influence students' interest in Physics.

Due to the quantitative nature of handling the data in this objective, it was translated into a hypothesis as follows;

Hypothesis 1: There is no significant difference between the mean interest scores of students of Physics taught using inquiry-based or interactive teaching approaches (Experimental Group E) and those taught using any other teaching approaches (Control Group C)

The results obtained after testing the hypothesis indicated that there was a significant difference between the two groups' interest score means, thus rejecting the hypothesis.

For students in experimental group E before and after the intervention, their mean interest scores improved (Tables 9). Teaching approaches that provide learners in group E with opportunity of practical manipulation of skills and practical exposition, seem to have enhanced their interest in the subject they studied (Ajaja, 2009). It is also in agreement with University of Minnesota (2008) which supports practices of students' group work, which was among the teaching approaches employed by the teachers in group E, to be good initiators of interest of students in what they study. The improvement in interest levels in Physics by students in experimental group E may highly be attributed to the intervention of teaching approaches applied by their teachers. Teachers in control group C may or may not have applied the approaches used by teachers in experimental group E. In any case, it did not have significant impact on group C's students' interest as their teachers were not accountable for any approaches they used.

Much as the mean interest scores for students in the control group also rose (Tables 4), the mean interest scores difference between group E and C before the intervention was significant, implying that before the intervention, the two groups had similar interest levels. The mean interest scores difference at the end of the study could only be attributed to the intervention. Like SEs did, Leong (2005) also intimated in his study that specific teaching approaches improve on the interest of students in any subject as long as the teacher or instructor uses them adequately.

Conclusion

In objective 1, the most prevalent teaching approaches of Physics at UCE mentioned by SEs and teachers include; Constructivism teaching based on students' discovery learning; Individual experimentation where apparatus, purpose and instruction are given and students perform, analyze and make reports; Group work where tasks are assigned to groups and each group presents results; Problem solving where only basic information is given and students solve the tasks; and hands-on-activities. These teaching approaches were referred to as inquiry-based/ interactive teaching approaches. Teachers in Uganda also apply Lecture teaching approach due to circumstantial challenges.

In objective 2, the interactive teaching approaches have a significant impact on interest of students in Physics at UCE. The research established that students specifically taught using constructivism teaching approaches had their interest in Physics increase higher than those students taught using any other teaching approach.

Recommendations

Basing on the two objectives of the research and following the research findings, discussion and conclusions, the following recommendations were made:

1. Teachers of Physics need to embrace the principles of constructivism teaching/ learning theory and then enlighten students on their importance in learning. The teaching approach is preferred because it bases teaching and learning on the applications of discovery learning, hands-on, and project-based.
2. All teachers of Physics at UCE in upper secondary should be graduate teachers trained to teach Physics. Graduate teachers are deemed to be more grounded in the subject matter and more confident during the teaching process.
4. Teachers of Physics at UCE need to be adequately motivated (financially, morally, materially or otherwise)

by government, private entrepreneurs and other Institutional leaders to enable them teach effectively to elicit students' interest in what is taught.

7. The Ministry of Education should adopt curricula reforms that encourage scientific skills acquisition and test the mastery of those skills through continuous assessment. The processes that solely adopt summative assessment produce low scores that demoralize learners and they lose interest in science subjects like Physics.

9. Teachers of Physics should use friendly language while teaching and provide local examples so that the concepts are applied in everyday situations thus give value to the subject.

10. The use of ICT or e-teaching/learning and (digital teaching/learning) should be adopted for learners to easily understand complex concepts in all education institutions.

Recommendation for Further Research

A replication of similar researches may be carried out at other levels of education and for different science subjects. Further study may specifically be undertaken to investigate the influence of interactive or constructivism teaching approaches with regard to gender and school type for these could have been the other extraneous variables responsible for the shortfall in the significance of some results.

Study Contributions

The study has added new knowledge concerning application of interactive/inquiry-based teaching approaches of Physics at UCE. Identified research gaps have been closed and solutions to effective teaching through recommendations that have been provided. For future Quasi-experimental studies, it can be stated that the bigger the *effect size* the greater the influence of an intervention on a dependent variable.

References

- Ainely, M. D. (2008), *Some perspective on interest in learning and classroom interaction*. University of Melbourne, Australia. <http://www.ure.edu.au/98pap/ain9804.htm>. Downloaded 23/06/2010.
- Ajaja, O. P. (2009), *Evaluation of science teaching in secondary schools in Delta State*. Delta State University, Abraka, Nigeria-Kalma-Raj.
- Amin, M. (2004). *Foundations of statistics inference for social science research*. Makerere University Printery, Kampala, Uganda.
- Amin, M. (2005). *Social Science Research, Conception, Methodology and Analysis*. Makerere University Printery, Kampala Uganda.
- Angela, H., David, W., & Mavis, K. (2001). The conceptual change approach to improving teaching and learning: An evaluation of a Hong Kong staff development program. *Higher Education Journal*, 42(2), 143-169.
- Appleton, K. (2008). Developing Science Pedagogical Content Knowledge Through Mentoring Elementary Teachers. *Journal of Science Teacher Education*, 19, 523-544.
- Awodun, A. O., Aladejana, A., & Oni, S. A. (2014). Students' Variables as Predictor of Secondary School Students' Performance in Physics. *International Journal of Scientific and Research Publication*. 4(8), 2250-3153
- Barbara, O., & Chng-ho, Y. (n.d). Threats to Validity of Research Design. Best, J. W., & Kahn, J. V. (1989). *Research in Education*. Pergamon Press, New Delhi.
- Biggs, J. B. (2002). *Aligning Teaching for Constructing Learning*. The Higher Education Academy.
- Bodner, G. (1986). Constructivism: A Theory of Knowledge. *Journal of Chemical Education* 63(10), 873- 877.

- Creswell, J. W., & Plario, C. V. I. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: SAGE Publications.
- Creswell, J. (2009). *Research Design. Qualitative, Quantitative, Mixed Methods Approach*. Sage Publications, Inc.
- De Vries, R. (2002). *Developing Constructivist Early Childhood Curriculum: Practical Principles and Activities*. Teachers College Press: New York.
- Donna, L. S. (1993). Controlling for Extraneous Variables in Experimental Research. A Research Note. *Accounting Education*, 4(2), **169-188**. DOI: 10.10.80/096339289500000020.
- Education Policy Review Commission (1989). *The White Paper on Education*. Kampala, Uganda.
- FAWE (2007). *Difficulties Faced by Girls in the Study of Science, Mathematics, and Technology Subjects. Findings of the Female Educationists in Mathematics and Science in Africa*. FEMSA Nairobi Kenya. A Simple Guide and Reference. 11.0 Update (4th Edition). Boston: Ally&Bacon.
- Hein, E. G. (1991). *Constructivism Learning Theory. The Museum and the Needs of People*. CECA International Committee of Museum Educators Conference.
- Hiilya, K. A. (2010). Pedagogical Renewal in Sub-Sahara Africa: The Case of Uganda. *Comparative Education*, **46(2)**.
- Kakooza, T. (2002). *Research: An Introduction to Research Methodology*. Kampala.
- Ken, A. (2010). *Using Theory to Guide Practice: Teaching Science From a Constructivist Perspective*. Faculty of Education, University of Central Queensland, Rock Hampton, Queensland 4702, Australia.
- Leong, L. (2005). Improving Students' Interest in Learning. Some Positive Techniques. *Journal of Information Systems Education*, 16(2), **1-2**
- <http://findarticles.com/p/articles/mi-qa404/is-200507/ai-nl4824687>.
- Medina, M. G., James, W. & Lawrence, B.(2010). Does Teaching Methods and Availability Of Teaching Resources Influence Pupils' Performance? Evidence From Four Uganda. Economic Policy Research Center. Kampala- Uganda.
- Melissa, H. (nd). Missing Data and How to Deal: An Over View of Missing Data. Population Research Center.
- Ministry of Education & Sports (2014). The Education Sector Strategic Plan (ESSP). Kampala, Uganda.
- Ogunleye, B. O., & Adepoju, O. F.,(2011). Everyday Phenomenon in Physics Education. Impact on Male and Female Students' Achievement Attitude and Practical Skills in Urban and Peri-urban Settings in Nigeria. *Pakistan Journal of Social Science* 8(6), **316-324**.
- Olarinmoye, R. D. (2000). Strategies for Effective Teaching of Modern Physics. A Keynote Address Presented at the STAN Physics Panel Workshop Held at Government College, Katsina State 24th-29th April 2000.
- Onah, D. U., & Ugwu, E. I. (2010). *Factors Which Predict Performance in Secondary School Physics in Ebonyi North Educational Zone of Ebonyi State , Nigeria*. Department Of Industrial Physics, Ebonyi State University, Abakaliki, Nigeria. *Adv. Applied Scientific Research*, 1(3), **255-256**.
- Robert, A. K., & John , R. O. (2014). Uganda SESEMAT Program: Impact and Challenges In its Implimentation. A Case Study of Jinja District. *International Journal of Education and Research*, 2(6).

- Sara, G., Sylvia, R., & Nicky, B. (n.d). Adjusting for Selection Bias in Retrospective, Case-Control Studies. *Oxford Journals, Science and Mathematics, Biostatistics 10(1)*, 17-31.
- Trigwell, K., Prosser, M., & Waterhouse, F. (1999). Relations Between Teachers' Approaches to Teaching and Students' Approaches to Learning. *Higher Education, 37(1)*, 57-70. Kluwer Academic Publishers. Sydney. NSW 2007. Australia.
- <http://link.springer.com/article/10.1023/A:100354831394>. Accessed 06/07/2013.
- Uganda National Academy of Sciences (2008). *Promoting Good Practices in Science and Technical Education: Science Prosperity*. ISSN: 1919-7078. Kampala, Uganda.
- UNEB (2000). *The report on the work of candidates of Physics Paper 1 at UACE*. Unpublished Report of UNEB. Kampala, Uganda.
- UNEB (2008, 2009, 2010, 2011, 2012). Performance of Candidates by Subject. Unpublished Report of UNEB. Kampala, Uganda.
- University of Minnesota (2008). *What is Active Learning?* Center for Teaching and Learning . University office Plaza, Suite 400, 221. University Ave. S. E. Minnesota MN55414. <http://www.umn.edu/ohr.teaclearntutorialsactivewhat/index.html>. Downloaded 12/12/2015.
- Von, G. E. (1995). *Radical Constructivism. A way of Knowing and Learning*. Studies in Mathematics Education Series: 6. Falmer Press, Taylor & Francis Inc, 1900 Frost Road, Suite 101, Bristol, PA 19007.
- William, M., K., T. (2006). *Research Methods Knowledge Base*. Quasi-Experimental Design.
- Worse, B. S. (2007). *Teaching and Learning. Teaching Approaches*. <https://www.bing.com/video/search?q=worse+2007>.
- Zavuga, R. (2008. August 04). Revised method of teaching science subjects. *The New Vision, Kampala p.21*. Retrieved from <http://www.opinion.Allafrica.org>.

An Assessment of Inter-Rater Reliability Concepts: A Case of Examiners of Mathematics and English at the Uganda Certificate of Education

By

Betty Nalukenge Habaasa

Email: Betty.Nalukenge@uneb.ac.ug/bnnalukenge@yahoo.com

Uganda National Examinations Board

Abstract

The Uganda Certificate of Education (UCE) annual examination is used to place students at the advanced secondary or tertiary level of schooling. Examiners of UCE converge at marking centres to score candidates scripts in various subjects. Basing on the three concepts: consensus, consistency and measurement estimates, the study sought to estimate the inter-rater reliability of marking Mathematics and English Language at UCE with an aim of improving the assessment process. Cluster random sampling was adopted in determining examiners for the study. Six marking teams from each subject paper were randomly selected. Sampled examiners re-marked scripts of their subject-papers during the normal marking procedure. Estimates were established using Kappa coefficients and intra-class correlation coefficients while Generalisability theory analysis determined the variations in scores. Sources of variations were determined using logistic regression. The consistence estimates of English Language and Mathematics were very high with low standard errors implying highly similar ordering of scores by examiners. However, a weak agreement (0.1) of scores between Chief Examiners and Examiners was associated with English Language. This was attributed to extraneous variables related to learners that contributed 80% variation in scores. A moderate agreement (0.6) was associated with examiners of Mathematics Paper 1 where Section A attained perfect agreement and section B had moderate agreement with magnitudes of 78% and 22% for extraneous variables for each section respectively. Sources of variation included section marked in Mathematics Paper 1 which had a significant effect on the examiners' deviation in scores. Hence consensus and consistent estimates provide information on assessment of learning while the Generalisation theorem that give information on teachers and learners throws light on assessment for learning. Therefore consensus and consistent estimates should be combined with other analysis like Generalisation and Rasch analysis to provide meaningful feedback to the teaching and assessment of learners.

Key words: *inter-rater reliability, consensus, consistent and measurement estimates, Generalizability*

Introduction

It is every student's dream that examination results represent the true level of their achievement. This feedback provides evidence of achievement of learners to the stakeholders and reflects assessment for learning. There is need to ensure that high-stake examinations are reliable. There are three sources of unreliability: within-candidate variation; between-examiner variation (inter-rater reliability) and within-examiner variation (intra-rater reliability) (Murphy, 2011; Gwet, 2008; Stemler & Tsai, 2008; Pantzare, 2015). This study focused on inter-

rater reliability which is defined as: the degree of agreement (consensus) among raters, Gwet (2008); the degree of consistency among raters (Stemler & Tsai, 2008) or a measure of the variation in scores between different markers evaluating the same student performance (Pantzare, 2015). Some studies have determined inter-rater reliability using either consensus estimates (Jonsson and Svingby 2007, Stemler & Tsai, 2008) or consistent estimates (Pantzare, 2015; Bandede and Adewale, 2013; Capuc, 2011). However, Stemler (2004) notes that describing inter-rater reliability as a

single concept is imprecise and potentially misleading especially when dealing with high-stake assessments. Countries and examination bodies are in a continuous process of ensuring that their assessments are reliable (Ofqual, 2010; Murphy 2011; Bandeale & Adewale, 2013).

Uganda National Examination Board (UNEB) conducts annual summative examinations mainly for certification purposes and placement of students to other levels of education. The Uganda Certificate of Education (UCE) is one of the examinations used to place students at the advanced secondary or tertiary level of schooling. At UCE, Mathematics and English Language are considered key subjects in determining the overall grade of a candidate hence the reason for choosing them in this study. Each of the two subjects consists of two examination papers. However, this study focused on Mathematics Paper 1 and English Language Paper 1.

UNEB has put in place a series of procedures/measures to ensure quality marking of scripts. These include among others: standardization meetings to co-ordinate the marking scheme with all examiners; Conveyor Belt Marking System (team marking, where examiners mark specific questions in a script), team leaders' re-marking of at least 10% of randomly sampled scripts in each marked script envelope and chief examiners monitoring team leaders' work (Odongo, 2010). However, these procedures are not standardized to inform assessment for learning. With all these measures put in place, variations in marking is still observed in several scripts in some subject papers. It was therefore necessary to establish the extent of these inconsistencies in reference to consensus, consistence and errors associated with raters.

In another dimension, it has been argued that scoring of particular subjects like Mathematics is more reliable than other subjects (Pantzare, 2015). On the other hand scoring of extended-answer items like in English Language are most difficult to judge consistently compared to objective assessment items (Brown, Glasswell & Harland (2008). On the other hand consistent of scores has been associated with a marker's highest education (Suto and Nádas, 2008)

and well trained examiners coupled with marking experience (Fink, 2005). At UNEB, teachers who are engaged in marking the UCE examination are of varying experience, age and education. The study further investigated the effect of these characteristics on the examiners' reliability.

Theoretical Framework

Three theories have been advanced to explain Inter-rater reliability studies: Generalizability Theory, Item Response Theory and Multi-Level Modelling. The use of the conveyor belt system in marking UNEB examinations, where examiners are nested in teams without multiple memberships made the study less ideal for Rasch modelling in item response theory. Yet the assumption that all variables should be normally distributed in multi-level modelling was not applicable to the data compiled in the study. The study therefore adopted the Generalizability theory (G-theory) with no distribution assumptions.

The G-theory is an extension of the classical reliability theory. Classical Test theory expresses the observed score (X) as a linear combination of the true score (T) and measurement error (E) i.e $X = T + E$ (Webb & Sheverson, 2005). However, Classical Test Theory does not differentiate among the various sources of errors i.e, $Var(X) = Var(T) + Var(E)$, hence the common term "undifferentiated random error". On the other hand, G-theory pinpoints the sources of measurement error, disentangles them, and estimates each one ($\sigma^2_E = \sigma^2_x + \sigma^2_T$) (Smith & Kulikowich, 2004). This therefore enables the estimates of variation in scores. This study adopted the generalizability theory to establish the impact of score deviations in marking Mathematics Paper 1 and English Language Paper 1. In particular the study was based on three facets: questions, teams, examiners and their interactions. In addition, other examiners' factors investigated in the study included examiners' qualification and section marked.

Statement of the problem.

Undertaking of Inter-reliability studies should be part and parcel of any examination procedure. However what has been done so far is specific and does not fully inform learning. In addition, inconsistencies in marking of some scripts of Mathematics Paper I and English Language Paper I have been occasionally observed during scrutiny of scripts before the award of grades. The irregularity in marking may affect the integrity of an award/or qualification by inflating or deflating the marks and hence affecting the grades of students. It was, therefore, important that the levels of unreliability of examiners of Mathematics and English Language papers are investigated, interpreted and corrected to increase the reliability of assessment outputs and enhance assessment for learning.

Purpose of the study

The purpose of this study was to estimate the Inter-rater reliability of marking UCE examinations in Mathematics Paper I and English Language Paper I with a view to improving the marking process at UCE. The specific objectives were to:

1. Estimate the degree of inter-rater agreement of examiners of English Language Paper I and Mathematics Paper I examinations at UCE.
2. Estimate the degree of inter-rater consistency of English Language Paper I (112/1) and Mathematics Paper I examinations at UCE.
3. Establish the sources of score variations in marking English Language Paper I and Mathematics Paper I.

Methodology

A cross-sectional design was adopted in the study. A total of 40 and 36 groups of examiners that marked English Language Paper I and Mathematics Paper I respectively in 2015 constituted the population. The choice of examiners was important because they had already been grounded in the implementation and use of the marking guides which is crucial in inter-rater reliability studies (Fink, 2005).

Cluster sampling of examiners was done in such a way that it did not disrupt the normal marking process. Six teams/groups from Mathematics Paper I and six teams from English Language Paper I participated in the study. Systematic sampling was used in identifying the ten scripts for English Language Paper I and Mathematics Paper I used in the study.

Research Instruments

Data compilation forms were used to capture the data: one was used to record original score of marks on the scripts during the normal marking and the other was used to capture data generated from the study.

Data Collection Procedure

The data collection procedure described below was done for each subject paper i.e Mathematics Paper I and English Language Paper I.

As the scoring of UCE 2015 progressed, ten marked script envelopes were randomly picked. Each envelope picked was assigned a unique range of 10 marks. A script corresponding to the range of the envelope was picked. Each of the ten script was re-written eight times by different writers and packed in returning envelopes similar to those used in final exams. The study scripts were mixed with live scripts. Examiners whose Group leaders randomly picked the study script envelopes from the strong room automatically participated in the study. The marked scripts went through the normal process of checking after which they were withdrawn from the live scripts.

The Chief examiners marks formed the reference marks which were used to compare with scores awarded during the study to determine consensus estimates. Examiners profiles were later corresponded with data bank of examiners in order to get other characteristics of examiners used in the study (qualification, experience). Data was later recorded on computer forms and entered in the computer for analysis.

Data Analysis Procedures

Data analysis was done in stages using STATA 3.0 statistical software package in the following aspects:

consensus estimates, consistency estimates and variation of scores.

In order to obtain consensus estimates (inter-rater agreement), examiners scores were transformed for each item in each script based on the reference marks: 0 – represented no deviation and 1 - represented deviation from the chief examiners’ mark. Percentage pairwise agreement for each item in each subject and Cohen Kappa estimates were established for the examiners. The averages of all the estimates gave the overall mean consensus estimates of the examination paper. The study adopted bench mark statistics for kappa as described by Landis & Kech, (1997) (below 0: poor, 0 to 0.2 - slight agreement, 0.21 to 0.40 - fair agreement, 0.41 to 0.60 - moderate agreement, 0.61 to 0.80-substantial agreement and 0.81 to 1.00- almost perfect agreement) to describe the nature of agreement.

To establish consistency estimates, intra class correlation coefficients (ICC) were established. Chinn (1991) and Multon (2012), recommend that any measure should have an intra-class correlation coefficient of at least 0.6 to be useful. Standard Errors of Measurements

(SEM) were determined to compliment the intra class correlation coefficients obtained on each item.

Finally to identify variations in scores firstly, variance component estimates were obtained to disentangle the error within the inter-rater agreement obtained (consensus estimates) basing on the generalizability theory. Secondly, the effect of examiners’ characteristics on marking reliability was attained. The rate of deviation of each examiner was computed as a percentage of the number of items deviated and the total number of items marked. Deviation was the dependent variable with “0” to indicate no deviation and “1” to represent a deviation in the scores. Relationships with examiners number of items deviated and their characteristics were established using Kruskal-Wallis test. Further analysis was done using logistic regression to explain the effect of the significant variables in the previous analysis. This enabled simultaneous analysis of both categorical and non-categorical variables.

The background characteristics of raters assessed in the study comprised of: highest qualification attained and section of the paper marked. Table 1 presents the distribution of these characteristics.

Table 1
Examiners qualifications and section marked in English Language and Mathematics

| Characteristics | Subjects | | | |
|-----------------|------------------|------------|-------------|------------|
| | English Language | | Mathematics | |
| Qualification | Frequency | Percentage | Frequency | Percentage |
| Diploma | 10 | 40.0 | 10 | 26.3 |
| Degree | 15 | 60.0 | 28 | 73.7 |
| Section marked | | | | |
| Section A | 9 | 36.0 | 8 | 21.1 |
| Section B | 16 | 64.0 | 30 | 78.9 |
| Both sections | | | | |

The examiners/raters were mainly teachers with a degree qualification for both Mathematics (73.7%) and English Language (60.0%) as shown in Table 1. More than half of the raters marked section B of the English Language Paper 1 (64.0%) and Mathematics Paper 1 (78.9%). Since

there were more questions in Mathematics to be attempted than English Language, this led to more raters of Mathematics (38) participating in the study compared to English Language (25).

The findings are further presented according to the objectives specified in the study.

Objective One: To estimate the degree of inter-rater agreement of scores

The analysis was done separately for both English Language Paper 1 and Mathematics Paper I. The examination paper of English Language Paper 1

comprised of two sections. Section A consisted of one compulsory item, while the rest of the questions were contained in section B. Candidates were expected to identify one question in section B to make a total of two questions required for the paper. The consensus estimates (percent agreement and Cohen Kappa (κ) estimates of examiners are summarized in Table 2.

Table 2

Mean Value of pairwise percent agreement and Kappa estimates among pairs of examiners

| Section | Mean percent agreement | κ |
|--------------------------------------|------------------------|----------|
| English Language Paper I | | |
| A | 56.0 | 0.12 |
| B | 50.0 | -0.02 |
| Average for English Language Paper I | 53.0 | 0.05 |
| Mathematics Paper I | | |
| A | 93.5 | 0.89 |
| B | 75.7 | 0.27 |
| Average for Mathematics Paper I | 84.6 | 0.58 |

Table 2 indicates that, the mean agreement of English Language Paper 1 section “A” (56.0%) is slightly greater than that of section “B” (50.0%). The overall results for English Language Paper 1 reveal a pairwise percentage agreement of 53.0 and an average pairwise Cohen Kappa estimate which approximates 0.1 implying slight agreement of examiners of English Language with Chief examiners as regards the scores awarded in the study.

Results further shows the overall percentage agreement of items in section A and B as 93.5% and 75.7 % respectively for Mathematics Paper 1. While Kappa estimates are 0.89 and 0.27 in section A and B respectively. The overall percentage agreement in section B is 84.6% and $\kappa = 0.6$ which indicate moderate agreement of scores between examiners and Chief examiners.

Objective Two: To estimate the degree of inter-rater consistency.

The rates of consistency of examiners’ scores in the various teams during the study were assessed by the Intra-Class Consistency (ICC) and summarized in Table 3. This was done for English Language Paper 1 and Mathematics Paper 1.

Table 3*Intra-class correlation coefficients of English Language Paper 1 and Mathematics Paper 1*

| Question | Intraclass Correlation | P-value | S.E |
|---------------------------------|---------------------------|---------|------|
| English Language Paper 1 | | | |
| 1 | 0.996 | .000 | 0.18 |
| 2 | 0.945 | .000 | 0.07 |
| Mathematics Paper I | | | |
| 1 | 1.000 | . | 0.00 |
| 2 | .996 | .000 | 0.02 |
| 3 | .971 | .000 | 0.06 |
| 4 | .979 | .000 | 0.08 |
| 5 | 1.000 | . | |
| 6 | 1.000 | . | |
| 7 | 0.985 | .000 | 0.11 |
| 8 | 0.982 | .000 | 0.06 |
| 11 | 0.991 | .000 | 0.12 |
| 13 | 0.989 | .000 | 0.19 |
| 14 | 1.000 | .000 | 0.32 |
| 17 | 0.995 | .000 | 0.21 |

In Table 3 results show that, the ICC values for English Language Paper 1 in Questions 1 and 2 are all greater than 0.9, which indicates excellent agreement between the multiple raters of English Language. This implies that examiners of English Language Paper 1 were able to implement the scoring rubric in a consistent manner since intra-rater consistency is greater than 0.7. The standard errors are all less than 0.2.

Table 3 shows that the ICC values for all items in Mathematics Paper 1 are greater than 0.9. This implies excellent consistency between the multiple examiners in the different marking teams of Mathematics. The standard error of measurements for items in section A

(Questions 3, 4, 7 and 8) approximate to 0.1. This is low and implies high reliability of examiners. The standard errors of items in section B slightly increases to 0.2 on average, with question 14 having the highest measure of 0.3.

Objective Three: To obtain sources of measurement errors

The Measurement estimates as stipulated in the generalizability analysis was based on variance component estimates. The factors/facets considered in the study included the question and teams and their interactions. The findings are summarized in Table 4 and 5.

Table 4*Score Variance Components breakdown for English Language Paper 1*

| Facets | Components Estimates | % Contributions |
|-------------------------------|------------------------|-----------------|
| Teams | 6.35×10^{-12} | - |
| Team – questions interactions | 1.30421 | 20% |
| Residual | 5.1921 | 80% |

Results in Table 4 show that there is no evidence of the effect of marking group on the total score variance. The teams-questions interaction effect is 20% of the total

score variances. The remaining percentage of 80 reflects the residual variance which is attributed to confounding factors within the candidates. Similarly, results of Mathematics Paper 1 are summarized in Table 5.

Table 5

Score Variance component breakdown for Mathematics Paper 1

| Facet/Factor | Section A | | Section B | |
|-----------------------------|-----------------------|-----------------|------------------------|-----------------|
| | Component Estimates | % Contributions | Component Estimates | % Contributions |
| Teams | 3.07×10^{-5} | - | 4.87×10^{-15} | - |
| Teams-questions interaction | 0.3503 | 22 | 5.8579 | 46 |
| Residual | 1.2677 | 78 | 6.8714 | 54 |

The results in Table 5 indicate that the effect of teams on the score variability of scripts is negligible in both sections A and B of Mathematics Paper 1. However, there is evidence of team-questions interactions effect with a contribution of 22% and 46% in sections A and B respectively on the total score variability of scripts. The effect is greater in section B compared to section A which correlates with the results of inter-rater agreement of 0.58 and 0.89 respectively. It also follows that residual variances are higher in section A (78%) than section B (54%) which represent confounding factors associated with the candidates.

Objective four: To identify possible causes of examiner's variations in scores.

The probable factors that may influence examiners' variation in scores were based on examiners' qualifications and section marked. Rate of deviations in this study referred to the ratio of the examiners' number of items with deviations to the total number of items marked. The results of Examiners of English Language based on Kruskal-Wallis test are shown in Table 6.

Table 6

Score deviations of English Language Paper 1 by examiners' qualification and section marked

| Characteristics | Frequency (n = 27) | Rank Sum | Avg. | p-value |
|----------------------|--------------------|----------|------|---------|
| Qualification | | | | |
| Diploma | 10 | 139.50 | 13.9 | 0.9747 |
| Degree | 17 | 238.50 | 14.0 | |
| Section | | | | |
| Section A | 11 | 171.00 | 15.5 | 0.2892 |
| Section B | 16 | 207.0 | 12.9 | |

Section: Either Marked Section A or B.

Results from Table 6 reveal no significant variations in discrepancy by examiners' qualification and section marked ($p > 0.05$) in English Language Paper 1.

Results of score deviations based on examiners' characteristics of Mathematics Paper 1 are illustrated in Table 7.

Table 7*Score deviations by examiners' qualifications and section marked in Mathematics Paper 1*

| Characteristics | Frequency(n = 38) | Rank Sum | Avg. | p-value |
|----------------------|-------------------|----------|------|---------|
| Qualification | | | | |
| Diploma | 11 | 208.50 | 18.9 | 0.8389 |
| Degree | 25 | 457.00 | 18.3 | |
| Section | | | | |
| Section A | 6 | 63.00 | 10.5 | 0.05 |
| Section B | 6 | 117.00 | 19.5 | |

Section: Either Marked Section A or B.

Results from Tables 7 reveal no significant variations in deviations by teacher's qualification ($p > 0.05$). In other words, the deviation in marks allocated did not depend on

the examiners' qualifications. However, the characteristic of the section marked gave a border line probability of 0.05 implying further analysis are shown in Table 8.

Table 8*Summary of Logistic Regression Results.*

| Deviation | Odds Ratio | Standard Error | P |
|----------------|------------|----------------|-------|
| Age | 0.9308 | 0.0939 | 0.478 |
| Experience | 1.1840 | 0.1727 | 0.247 |
| Section marked | 3.0687 | 1.6181 | 0.033 |
| Qualification | 0.9116 | 0.7872 | 0.915 |
| Cons | 1.5662 | 5.2301 | 0.893 |

The results in Table 9 show that, the "section marked" as a significant variable in determining the deviations of marks in Mathematics Paper 1 ($p < 0.05$). The results show an odds ratio of 3.068. This implies that for every unit increase in the section marked (moving from A to Section B) the odds of making an error increase by 3 compared to marking section A.

Discussion

The consensus estimates were below the recommended value in both English and Mathematics. They were much lower in English Language Paper 1 compared to Mathematics Paper 1 indicating slightly weak agreement and moderate agreements with respect to the reference marks respectively. This corresponds to the findings of Jonsson & Svingby (2007) who reported that most studies had consensus estimates below the recommended value. This suggests that interpretation of the marking guides differ greatly within examiners of English Language compared to examiners of

Mathematics. The findings therefore, are in agreement with Pantzare (2015), who indicated that Mathematics rubrics are easier to interpret and score than any other subjects because of their objectivity. English language appear more complex to mark than other subjects. In fact at UCE marking centres examiners of English Language are usually among the last groups to complete marking. The lower levels of inter-rater agreement in English Language could be attributed to extraneous factors that contributed 80% of the variation in scores as revealed in the generalizability analysis. This implies that there are factors associated with candidates (content covered, skills acquired, candidate attitude, environment etc) which may not easily be disaggregated. However, in Mathematics Section A had agreeable consensus results compared to section B. The generalizability analysis in Mathematics further demonstrated that the effect of examiner-question interaction was outstanding in both sections A and B. The variance component estimates corresponding to this

effect were 22% and 46% of the total estimates in section A and B respectively. This therefore implies that the interpretation of candidates' responses and score distribution by examiners produces more error in section B than section A of the Mathematics examination paper section. These findings are relevant in estimating assessment for learning in both English Language and Mathematics Paper 1. The results therefore correspond with the degrees of inter-rater agreement which is higher in section A rather than B. This therefore suggests that consensus estimates provide information that enables examiners improve their standardization methods so that marks are closer in agreement with each other. However, a combination of the generalizability analysis provides more information on learners and teachers that can enhance assessment for learning.

The findings further indicate that examiners of English Language Paper 1 and Mathematics Paper 1 have a high level of consistency. All the items in the examination papers have an intra-class correlation coefficient of at least 0.9 which is quite high and above the acceptable levels of 0.7 according to Bland (2007) with small standard error of measurements. Consistent estimates only provide information on the ordering of scores, which alone can't contribute to assessment for learning. It was noted that section A had a slightly higher ICC rate compared to section B of the English Language examination paper. Since Section A is compulsory to all candidates this may have enabled examiners to master the marking guide because of high frequency of similar candidates' responses. The findings are in agreement with Capuc (2011) who reported a higher internal reliability of section A compared to section B of the English Language Paper 1 examinations. For Mathematics Paper 1, the findings revealed that half of the questions in section A had a consistency of 1.00 which indicates perfect agreement as regards the order of score of marks.

As regards possible sources of variations in scores, the section marked in Mathematics Paper I had an influence on the examiners' deviations. Deviation of examiners increased with those who marked section B. This

corresponds with the generalizability theory analysis findings which indicated a bigger percentage variation effect in section B (46%) while a smaller percentage (16%) was associated with section A. This further suggests that there are variations in marking sections A and B in Mathematics Paper 1. This finding is in agreement with Sutos & Nadas (2008b) who mentioned that items with more complex strategies were marked less accurately compared to those with simpler strategies.

The outcomes of the study point out that deviation in examiners marks in English Language Paper 1 and Mathematics Paper I do not depend on examiners' qualifications. The results correspond to findings resulting from the variance component breakdown (generalizability analysis) that indicated that the effect of examiners had no contribution to the final error. This implies that the deviations could have been caused by other characteristics of examiners or extraneous variables attributed in the students' responses as earlier indicated. Therefore, an integration of the concepts of inter-rater reliability with inter-rater reliability theoretical analysis provides more meaningful information that can enhance assessment for learning.

Conclusions

Basing on the findings of the study, the researcher made the following conclusions:

1. The inter-rater agreement estimate of examiners of English Language Paper 1 can be described as slightly weak agreement while that of Mathematics Paper 1 as a moderate agreement between chief examiners and examiners. This also implies that the degree to which examiners of English Language interpret the marking guide differs between examiners and their chief Examiners. This is caused by extraneous factors within the candidates taking the English Language Paper 1 Examination that accounted for 80% of the total error.
2. The high intra-class correlation coefficient of English Language and Mathematics Paper 1 revealed excellent consistency of examiners of English Language Paper 1 in scoring

3. candidates' scripts. This therefore implies that although examiners differ in the interpretation of the marking guide they do apply it systematically during marking.
4. Errors were further identified with examiners of Mathematics who marked sections A and B, followed by those marking section B only. No major errors were identified with examiners marking section A only.

Recommendations

Basing on the foregoing discussion and conclusions, the following recommendations are made:

1. Since 80% of the total variance components in English Language Paper 1 are due to the effect of the candidates' extraneous variables, soft skills assessment should be implemented in the teaching and learning of English Language. In addition the teaching of composition writing should be enhanced to ensure that students get the necessary skills required for the examinations.
2. More time should be spent on co-ordination of scoring rubrics of English Language Paper 1 and section B of Mathematics Paper 1
3. The common aspects of estimations of inter-rater reliability should be combined with theoretic models (Generalizability analysis, item response theory and Rasch) of inter rater reliability to provide feedback that can enhance assessment for learning.

References

- Bande, S. O., & Adewale, A. E. (2013). Comparative analysis of the reliability and validity coefficients of WAEC, NECO and NABTEB constructed mathematics examination. *Journal of Education and Social Research*, 3 (2), 397-402.
- Capuc (2011), A Comparative Analysis of the strength, validity and reliability of English Language Paper 1 sections A and B of Uganda Certificate of Education Examination. *UNEB Research Report*.
- Fink, A. (2005). Evaluation fundamentals: Insights into Outcomes, Effectiveness and Quality of Health Programs. *Thousand Oaks: SAGE Publications, Inc.*
- Jonsson, A & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *School of Teacher Education, Malmo University, SE-205 06 Malmo, Sweden*
- Murphy, R.J.L (2011). Reliability of Marking in eight GCE examinations. DOI:10.111/212.
- Odongo, D.N (2010). Implications of the 2010 theme on the role of the senior examiner at the Uganda National Examination Board. *A report presented at the 2010 UACE Senior Examiners' Meeting, December, 2010.*
- Ofqual (2010), Public perceptions of unreliability in examination results in England. *A paper presented at the 36th International Association of Education Assessment (IAEA) annual Conference in Bangkok, Thailand, 22-27, 2010.*
- Okelowange, P.J. (2004). Marking Reliability in Uganda Advanced Certificate of Education (UCE) examination with reference to History paper 3. *A research report published by Uganda National Examination Board, Kampala.*
- Pantzare A.L (2015). Inter-rater reliability in large –scale assessments. *Practical Assessment, Research & Evaluation preonline.net.*
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating inter-rater reliability. *Practical Assessment, Research, and Evaluation vol 9*
- Stemler, S. E., & Tsai, J. (2008). Best practices in inter-rater reliability: Three common approaches. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 2949). Thousand Oaks, CA: Sage

SUB THEME B: COMBATING EXAMINATION MALPRACTICE

Examination Malpractice as a Consequence of Media Publication of League Tables: The Case of the Cameroon General Certificate of Education Examinations

Denis Ndasi Mofor (MSc.)

denisndasimofor@gmail.com

John Teneng Awa (PhD)

tndipingwi@gmail.com

Kenneth NguFoncha (PhD)

kfoncha@yahoo.com

Cameroon GCE Board

Abstract

Institutional accountability, outcomes-based education and the use of results to make choices between schools are justifications for the ranking of schools based on performances of the Cameroon General Certificate of Education (GCE) examinations. This paper explores the statistical vagueness of this practice in the context of functionality, relevance and comprehensibility and posits that league tables are misleading as they encourage interpretations that are damaging to students and schools as they focus only on the overall percentage pass, at the expense of other outcomes (value-added) that are equally important but not easy to measure. The consequence of such a practice, we argue, is the inclination to cheat and the unorthodox practice by school administrators who often screen candidates before registration into internal and external candidates. This is in order to maintain only the academically strong ones as internal candidates in pursuit of high percentages. The survey research design was adopted for the study. A closed and open-ended questionnaire was administered to 520 students who have written the GCE examinations, purposively drawn from the fifth, lower-sixth and upper-sixth forms in both Government, confessionnal and lay-private schools in the South West Region of Cameroon. The closed-ended items were analyzed quantitatively while the open-ended items were qualitatively analyzed. Results showed that 81.7% of the candidates asserted that announcing individual results over the radio and seeing them in newspapers causes anxiety among candidates. They equally indicated that they would prefer their results forwarded confidentially rather than published in the media. The study posits that student achievement is dependent on a variety of influences other than those brought to bear on the school and recommends that the media publication of GCE results according to league tables should be scrapped in an effort to curb examination malpractice and other unorthodox practices.

Keywords: *Examination Malpractice, League Tables, Statistical Vagueness*

Introduction

The Cameroonisation of the University of London General Certificate of Education (GCE) Examinations in 1977 was done with all related practices. Since then, media publication of results in the form of league tables has been a regular practice at the Cameroon General Certificate of Education (GCE) Board with no public or

academic debate as to the worthiness of this practice. Justification for the practice assumes institutional accountability, outcomes-based education and the use of results to make choices between schools. While the government of the UK in 2002 and 2006 integrated the value added method and the contextual value-added (CVA) respectively, with arguments that schools vary in

the quality of their intake and family background of the child, the practice in Cameroon is the provision of basic information on the general percentage. This practice has never been modified since its inception.

Background to the Study

Managing pressure from the public that usually comes along with the publication of results is a difficulty the Cameroon General Certificate of Education (GCE) Board

is yet to resolve. While for students it is all about the anxiety to get their results, the publication of General Certificate of Education (GCE) examination results provides the opportunity for brisk business to the media in Cameroon. Cases are many where the police have been invited to protect the premises of the Board from the public as they impatiently await results.





Figure 1: Picture of anxious public hunting for GCE examination results at GCEB Head Office and News Papers

League tables at the Cameroon GCE Board are produced in two categories. The first category represents the overall percentage pass of schools. In this grouping, the schools are ranked according to the percentage of candidates who are successful in four or more papers at the Ordinary Level and two or more at the Advanced Level. In the arithmetic calculation, if school A has 20 candidates who all are declared successful, she scores 100% ahead of school B which may have 500 candidates and 480 are successful. In the present arrangement, school B scores 96% which is considered by all stakeholders as an inferior performance to that of school A, irrespective of their contribution to national development numerically. In the second category, candidates are classified in honourroll according to points/grades obtained. The schools of such candidates are highlighted. These rankings are extensively published in the national and local newspapers and are used by schools in their promotional literature. The supposition is that the position in the rankings can be treated as a measure of the quality of their educational provision. There have been a number of incidents over the years where the GCE Board has been petitioned to correct the percentage pass for some schools

and broadcast same over the national media or face litigation in the law courts. The schools threatening such action usually claim they deserved a 100% pass but were given a lower percentage pass which they feel will be injurious to their future admissions.

After the publication of GCE examination results, especially around September when schools are competing for new enrolments, Cameroonian towns are inundated with flyers, posters and billboards trying to lure parents and students see their schools as high performing schools. Unfortunately, most of these glamorous percentages are sometimes fabrications of the schools.

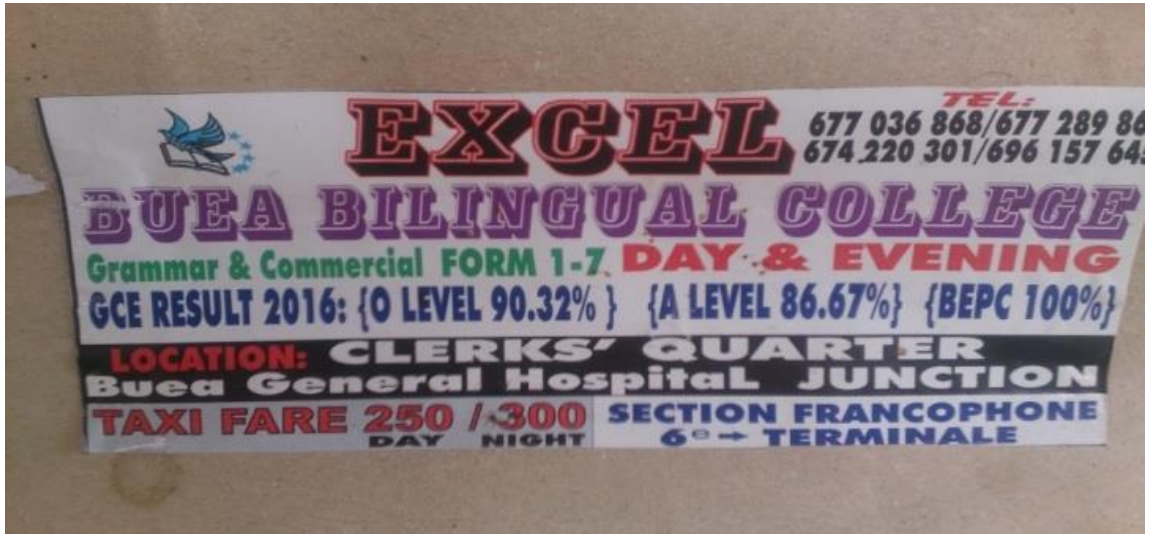


Figure 2: Flyer of Buea Bilingual College

Review of Literature

Arguments for the use of league tables

Given that schooling accounts for a major proportion of public and private expenditure, as well as generating a considerable quantity of paid employment for teachers and administrators, the enduring interest by government and the media in the relative effectiveness of school education and its improvement is not unexpected (Motimore, 1998; Tucker & Coddling, 1998; Rowe, 2000; Watson, 1996).

A fundamental support for league tables is that it provides information about school performance that would create competition between schools and would motivate teachers to change their instructional practices (Chapman and Snyder 2000). It is also understood that the ranking would create a market situation in which parents would shop around for the best available schools for their children. League tables are fine as they represent the school and overall success which helps new comers choose right schools.

There is substantiation to support the fact that assessment data, published in league tables, can affect the behavior of schools. In the 1990s, Senegal introduced a results

oriented management system in which information on school performance was published in the press. Between 1995 and 1998, the success rate for the examination at the end of primary school rose from 30 percent to 48 percent with the enrollment of girls moving from 40 percent to 77 percent (ADEA 2000). Educational accountability is, therefore, the strongest argument for the publication of league tables. Research shows that pupils benefit from having a level of scrutiny from their teachers. It was found that the decision in 2001 by Welsh Assembly to stop secondary school performance data or “league tables” resulted in a significant deterioration in GCSE performance in Wales.

Arguments against League Tables

Although league tables continue to generate enormous and controversial debate in academic cycles, many drawbacks of the practice abound. Wilby (2002) posits that since the creation of league tables in England, cheating has arrived in the schools adding that many academics see it as a phenomenon that will grow further in future. This is also supported by Lippset (2007) who holds that a negative consequence of league tables is that both teachers and schools are under enormous pressure from local authorities to the extent that for some staff,

cheating has become an option. Teacher Support Network (2008) opines that league tables exert a lot of stressful moments on teachers who are sometimes less productive and feel unfilled.

Hutchings in her book titled 'The impact of accountability measures on children and young people' has equally lamented that the pressure to help pupils succeed in high stakes tests leads teachers to engage in a variety of practices which she refers to as 'gaming the system'. According to her, gaming the system includes a

wide spectrum of activities in the school system ranging from legitimate practices such as question spotting and teaching the topics you expect to come up in the examination, to practices that are clearly cheating, such as giving students hints during a test. Though she concludes that many of the practices are perfectly legitimate, she, however, observes that they are at the same time ethically questionable.

Leckie&Golstein (2009) have cautioned that the merits of such practices notwithstanding, league tables come with a huge uncertainty when it comes to statistics. They thus posit that due to league tables, it is now all about figures and no longer about education which they contend should be more than numbers only. More so, they lament that it could be very devastating if a child is making improvements and working hard when his/her school is at the bottom of the league tables.

Findings from several research studies point to the fact that the publication of results according to league tables does not increase student achievement. This is corroborated by Hannaway& Hamilton (2008) who also indicate that a lot of concerns have been raised in US studies about the narrowing of subject areas in response to state testing and publication of school results. Cobbold (2010) clearly states that far from improving education, media publication of school results and league tables are likely to harm education in a number of ways. Citing experience from a number of countries, he contends that league tables distort teaching practice, disadvantage low and high achieving students, unfairly stigmatize low achieving students; discourage co-operation and collaboration between schools and teachers; and

increases social segregation and inequity in education. Cobbold (2010) concludes that league tables turn classrooms into test preparation factories where weeks and months are devoted to test preparation at the expense of an education that empowers learners for deeper and wider learning experiences.

Literature on the subject is emphatic that league tables provide imprecise and misleading information. Even with the use of Contextual Value Added data, the work that schools serving disadvantaged areas do to narrow the achievement gap is not fully reflected in the tables and

therefore gives a misleading impression of the quality of learning that is taking place in the school system. The information provided by the tables fails to reflect the character, ethos and catchment area of a school.

ABC News in 2010, reacting to the negative practices that schools indulge in as a consequence of the National Assessment Program – Literacy and Numeracy (NAPLAN), lamented:

We've got the ridiculous situation of schools just setting aside a whole range of good curriculum offerings just to concentrate on NAPLAN so the school would be seen to be in some way better than the school next to it.

The governments of Northern Ireland and Wales have abolished league tables because of their perverse effect on schools and the limited information they provide. In Scotland, league tables have never been published. The other UK home countries have recognised that such tables neither contribute to school improvement nor support schools with the greatest need. This has not affected either the quality of education or parents' satisfaction with the level of information they receive about schools.

According to Hallgarten (2001), one of the crucial drawbacks of league tables is that schools use and manipulate the tables to demonstrate their performance in a very good way. He adds that they do so by changing priorities and picking the data which suit them best. West & Pennell (2000) are in support of this when they corroborate that parents are exposed to league tables and

cannot do anything against their manipulation as they are exposed to the already highlighted schools' promotional material.

The Problem

The Cameroon GCE Board is still using league tables and the media in the publication of the General Certificate of Education (GCE) examination results, in spite of the fact that many assessment systems have abandoned these practices as a result of the negative consequences such practices exert on the ethics of the assessment processes and on teaching and learning in general. The consequence

of such a practice, this paper argues, is the inclination to cheat and the unorthodox practice by school administrators who often screen candidates before registration into internal and external candidates. External candidates are those who sit for the GCE examinations as independent candidates with no relationship with any particular school while internal candidates are products of particular schools and whose success or failure in the GCE examinations means a lot to the public and the image of the schools.

Objectives of the Study

This study was out to investigate if:

- i) the publication of the General Certificate of Education (GCE) examination results according to league tables encourages examination malpractice;

Findings

Table 1:

Participants of the study

| Internal Candidates | | External Candidates | | Total |
|----------------------------|--------|----------------------------|--------|--------------|
| Male | Female | Male | Female | |
| 150 | 175 | 80 | 115 | 520 |

Of the 520 participants, 325 (62.5%) were internal candidates while 195 (37.5%) were external candidates. Concerning their choice of school, 350 (67.3%) of the candidates indicated that it was influenced by the past

- ii) the media publication of the General Certificate of Education (GCE) examination results introduces unethical practices in the examination processes.

Methodology

The survey research design was adopted for this study. The population of the study consisted of all the students who have written either the General Certificate of Education (GCE) Ordinary or Advanced Level examinations in Cameroon. A combination of random stratified and purposive sampling techniques were used to select a sample of 520 candidates from the fifth, lower-sixth and upper-sixth forms of both Government, confessional and lay-private schools in the South West Region of Cameroon. Respondent characteristics (internal/external candidates) in the examinations provided the strata. Of those selected, 62.5 % (325) were internal and 37.5% (195) external candidates. Instrument for data collection was a closed and open-ended questionnaire of 9 items, administered to the 520 students by the researchers during field visits. The use of the questionnaire was appropriate in the study since it enabled the researchers to probe into the perceptions of candidates about the effects of media publication of GCE examination results according to league tables. The closed-ended items were analyzed quantitatively while the open-ended items were qualitatively analyzed using simple percentages.

GCE examination results. Some of them pointed out that though it was the choice of their parents, their parents took the decision in due consideration of the past good results of the school.

All the students affirmed that the decision to register them as internal or external candidates in the GCE examination was that of the school. The decision was based on candidate performances on the pre-mock and mock GCE examinations, which are considered by most school administrators to be strong indicators of success or failure on the GCE examinations.

Majority of the candidates who were forced to write the GCE examinations as external candidates regretted that

the decision to register them as external candidates was taken at the time of registration by their school authorities when they had no other possibility for another internal centre. Many of them saw this separation, especially in cases where the external centre was different from their school, as a stigma since this was an indication that they were not good enough for the GCE examinations for that year. 425 (81.7%) of the 520 students affirmed that the publication of GCE examinations results over the media (radio and newspapers) causes anxiety amongst candidates, while only 95 (18.3%) of them said they were not concerned about whichever way their GCE results were communicated to them. 390 (91.7%) of the 425 who indicated that media publication of GCE results causes anxiety said they would prefer their GCE results forwarded to them confidentially through their centres or to their private mail box.

Some of the negative things they believe publishing GCE results over the media can cause were:

- Some students hide knowledge from their peers so that only their names should appear on the honour roll in the GCE examinations results;
- Teachers assist their students in order to enable their school obtain the 100% success rate in the GCE examinations;
- Cheating takes place when students want to avoid the shame of not hearing their names over the radio, so they do everything to secure a pass;
- Some parents put a lot of pressure on the children so that they should not disgrace the family;

- Publication of GCE examination results over the media encourages 'toll-gate' classes. Toll-gate classes are organized on the eve of the examination, some times at midnight to coach candidates in order to obtain higher percentages.

Discussion of Findings

Although 350 (67.3%) of the candidates corroborated that their choice of school was influenced by past GCE examination results, Taylor and Nguyen (2006) have cautioned that by using current results of examinations to predict the future performance of the schools, a crucial statistical uncertainty can arise. They contend that this is due to the fact that the performing tables neither use pupil's individual scores to show how schools have improved nor bear in mind their prior attainment. Besides, Hallgarten (2001) has also indicated that schools often use and manipulate the tables to demonstrate their performance in a very good way by changing priorities and picking the data which suit them best. As a consequence, parents are exposed to the league tables and cannot do anything against their manipulation as they obtain the already highlighted schools' promotional material which is part of schools' improvement work (West and Pennell, 2000).

There is a lot of contention that even with the use of Contextual Value Added data in the British system, the work schools that serve disadvantaged areas do to narrow the achievement gap is not fully reflected in the tables and therefore gives a misleading impression of the quality of the school. The information provided by the tables fails to reflect the character, ethos and catchment area of a school.

Since 425 (81.7%) of the students affirmed that the publication of GCE examination results over the media (radio and newspapers) causes anxiety amongst candidates, these feelings of anxiety expressed by majority of the participants are in consonance with findings of a report of a research project commissioned by the National Union of Teachers in the UK titled "Exam Factories?: The impact of accountability measures on

children and young people.” In the report, the researchers concluded that children and young people in UK were suffering from increasingly high levels of school-related anxiety and stress, disaffection and mental health problems due to the implementation of accountability measures, one of which was the publication of league tables. This, they posited, was caused by increased pressure from tests/exams; greater awareness at younger

ages of their own ‘failure’; and the increased rigour and academic demands of the curriculum. They equally lamented that increasingly, children and young people see the main purpose of schooling as gaining qualifications, because this is what schools focus on.

With regards to the numerous negative consequences that league tables exert on education (cheating, mad rush after 100% pass, toll gate classes etc.), research literature is emphatic that schools and teachers tend to respond to publication of school results and league tables by focusing more on teaching test taking skills and practicing for tests at the expense of deeper learning experiences that develops analytical skills and greater understanding. League tables turn classrooms into test preparation factories. Weeks and months are devoted to test preparation and often times on techniques to undercut the examination system through malpractices.

Findings of this study have equally raised critical issues in the domain of functionality, comprehensibility and relevance. With regards to functionality of league tables within the Cameroonian context, the statistical vagueness of the practice introduces no relevant information beside the overall percentage pass of the schools. More so, the fact that students are screened for internal and external centres greatly weakens any arguments for the

functionality of the rankings in Cameroon. Cobbold(2010) also holds that far from improving education, publication of school results and league tables are likely to significantly harm education. Citing experience from many education systems, he posits that league tables are likely to harm education in the following ways:

- Narrow the curriculum;
- Distort teaching practice;
- Disadvantage low and high achieving students;
- Unfairly stigmatise low achieving students;
- Make it more difficult for low performing schools to retain high quality teachers;
- Discourage co-operation and collaboration between schools and teachers; and
- Increase social segregation and inequity in education.

Even though parents/students admitted that their choice of school was informed by league tables, such information is considered misleading, especially as the tables are reflecting a situation 5 years ahead of the student’s entry into the establishment. Changes in the administration and other variables in the organizational set-up of the school over time may seriously compromise the relevance of such league tables. The relevance of league tables in this context is further eroded when we consider that students differ in their level of achievement when entering school. Any subsequent measure of achievement that does not take this into account will be inequitable and misleading. Moreover, the general public is increasingly not in support of league tables. In a survey of 10,000 parents in 2009, 75% of them believed that Key Stage 2 league tables should be abolished (NAHT). The study further corroborated previous studies by indicating parents feel that league tables do not assess the quality of teaching in a school or the development of an individual child (GTC 2007 survey)

Student achievement is dependent on a variety of influences other than those brought to bear on the school.

Comprehensibility negates the use of league tables on grounds that they do not make use of value added and contextual value added. By using current results to predict future performances of the school, a statistical uncertainty can arise.

Conclusion

The study is emphatic that league tables greatly harm education. Apart from creating unnecessary anxiety, they are less effective in indicating high quality learning outcomes. Allen and Burges (2010) have equally insisted that school performance tables remain a controversial part of education policy, asserting that their usefulness in informing school choice is strongly debatable.

Recommendations

- The current practice of publishing GCE examination results over the media and ranking of schools according to percentage pass should be reviewed.
- The different purposes of testing should be separated out so that tests intended to measure pupils' progress and attainment are not used for school accountability.
- There should be a renewed focus on broadly based curriculum objectives which foster creativity, curiosity, and enthusiasm to learn in the students. Collaboration should be encouraged, rather than competition.

References

- ADEA, (2002). *Reaching Out, Reaching All. Sustaining Effective Policy and Practice for Education in Africa. Fighting HIV/AIDS*. Proceedings of the ADEA Biennial Meeting, Arusha, Tanzania, October 7-11, 2001. Paris: ADEA.
- Allen, R. and Burgess, S. (2013). Evaluating the provision of school performance information for school choice, *Economics of Education Review*, 34(2013) pp.175-190
- Chapman, D.W., and C. W. Snyder (2000). "Can High Stakes National Testing Improve Instruction? Reexamining Conventional Wisdom." *International Journal of Educational Development* 20 0: 457-474.
- Hallgarten, J. (2001). School league tables: Have they outlived their usefulness?. *New Economy*, 8: 189–196.
- Leckie, G. and Goldstein, H. (2009). The limitations of using school league tables to inform school choice, *Journal of the Royal Statistical Society: Series A*, 127(4) pp.835 - 52.
- Lipsett, A. (2007). League table predicts RAE forerunner. *Education Guardian*. Accessed 30 October 2007, <http://education.guardian.co.uk/RAE/story/0,,2201416,00.html>
- Mortimore, P. (1998). *The road to improvement: Reflections on school effectiveness*. Lisse, Netherlands: Swetz & Zeitlinger.
- Rowe, K.J. (2000). Educational performance indicators. In M. Forster, G. Masters and K.J. Rowe, *Measuring learning outcomes: Option and challenges in evaluation and performance monitoring (pp.2-20)*. Strategic Choices for Educational Reform; Module IV-Evaluation and Performance Monitoring. Washington, DC: The World Bank Institute.
- Tucker, M. & Coddling, J. (1998). *Standards for our schools: How to set them, measure them and reach them*. San Francisco, CA: Jossey-Bass.
- Watson, L. (1996). Public accountability or control? Benchmarks of performance in Australian schooling. *Australian Journal of Education*, Vol. 40, pp. 104- 123.
- West, A. and Pennell, H. (2000). Publishing school examination results in England: Incentives and consequences, *educational studies*, 26(4) pp. 423-436.

Impact of District Performance League Tables on Performance at the Basic Education Certificate Examination in Ghana

Felix Akuffo-Badoo
felixakuffob@yahoo.com

The West African Examinations Council, Accra, Ghana

Abstract

In 2004, the Ministry of Education introduced the School Performance League Table for the Senior Secondary School Certificate Examination (SSSCE) and it was published widely in the National dailies. The publication of the league tables allowed the comparative study of a particular school's performance over the years. It also served as a challenge to school authorities as it publicized the performance of their schools. In 2005, a Committee of the Council observed that the publication led to a sharp increase in the number of malpractice cases recorded at the subsequent examinations and recommended that the publication be stopped. Another Committee of the Council however was of the opinion that the publication had a positive impact and disagreed that it should be stopped. The league table for the SSSCE was not published again after 2005. The Ministry of Education has however since 2013 presented the District Performance League Table for the Basic Education Certificate Examination (BECE) at yearly Stakeholders' Conferences with District Chief Executives and Directors of Education. The purpose of the study was to find out the impact of the Performance League Tables on the performance of districts over a period of three years. One hundred and thirty (130) districts were purposively sampled for the study. The performance of 43 deprived districts and 50 poor performance districts before and after the introduction of the publication of the League Tables were compared. One of the findings of the study was that the league table had a positive impact on the performance of a few districts in subsequent examinations. Recommendations have been made on how to use the league tables to offer support to districts for improved performance at the BECE.

Key words: District Performance League Tables, Basic Education Certificate Examination (BECE), Ministry of Education, Stakeholders' Conference

Introduction

The West African Examinations Council (WAEC) has been responsible for the certification of candidates in Ghana at the pre-tertiary level since it was established in 1952.

The Council conducted the Common Entrance Examination (CEE) for selection of pupils to secondary schools (from 1953 to 1986) and the Middle School Leaving Certificate Examination (MSLC) for certification of pupils at the end of basic school education (from 1957 to 1992).

However, an educational reform programme that was implemented by the government of Ghana in 1987 led to the replacement of the CEE and the MSLC with the Basic Education Certificate Examination (BECE) in 1990. The BECE therefore serves a dual purpose – for certification

of pupils at the end of 9 years of basic education and for the selection of candidates into senior high schools. The Council also administered the Senior Secondary School Certificate Examination (SSSCE) from 1993 to 2005.

In 2004, the Ministry of Education (MOE) introduced the publication of a School Performance League Table in the National dailies using the then SSSCE. The publication was to allow the comparative study of a particular school's performance over the years. It also challenged school authorities as it published the performance of their school.

However, in 2005, the then Ghana National Aptitude Tests and Examinations Committee of the Council observed that there was an upsurge in the malpractice cases recorded after the publication. The Committee also observed that there were inequalities in the provision of

school facilities and therefore recommended that the publication be stopped. A higher Committee of the Council however disagreed with the recommendation that the publication be stopped because in its view, it had a positive impact. The School Performance League Table using the SSSCE was however not published again but the MOE computed a District Performance League Table using the BECE for internal use.

The MOE has however since 2013 presented District Performance League Tables using the BECE results at yearly Stakeholders Conferences organised for Metropolitan/ Municipal/ District Chief Executives and Directors of Education. According to Dr. S. Adu, a former Director of Basic Education, Ghana Education Service (GES), the objectives of the Stakeholders' Conferences were

1. To enable the MOE/GES account to the District Chief Executives and Directors of Education the average performance of BECE candidates by district in the form of a league table.
2. To enable the MOE/GES to appreciate and recognize the efforts of high performing districts and encourage low performing districts to improve.
3. To allow for opportunity for the respective District Chief Executives and Directors together to assess the performance of their BECE candidates and to forge common strategies to improve performance.
4. To provide opportunity for respective District Chief Executives and Directors to compare the performance of their BECE candidates with one another as a means of motivating the individual districts to commit resources to improve their performance.
5. To provide opportunity for the MOE/GES and District Chief Executives/Directors to discuss common strategies on how best BECE performance could be improved across the country.

The opinion of educational researches is however divided over the usefulness or otherwise of the publication of league tables.

Reed & Hallgarten (2003) noted that information captured in league tables have been used to determine whether

- schools should be given light or heavy touch inspection;
- schools should be given (now defunct) improvement awards;
- teachers qualify for extra pay through passing the threshold;
- private companies managing Local Education Agencies should gain performance bonuses;
- the Secretary of State for Education should resign if certain public service agreement targets are not met.

Baku (2005) also states the necessity for performance league tables as follows:

- Recognising and appreciating good performing schools.
- Introduction of competitive spirit – among schools, teachers and students (competition among un-equals).
- Making the school authorities accountable to the stakeholders.
- Attracting stakeholder interest in and support for schools.
- Introducing relatively unknown but good performing schools for recognition.

However, Cobbold (2010) is of the view that the publication of league tables is likely to significantly harm education. He said findings from major researchers show that they

- narrow the curriculum;
- distort teaching practice;
- disadvantage low and high achieving students and unfairly stigmatise low achieving students.
- make it more difficult for low performing schools to retain high quality teachers and
- increase social segregation and equity.

Cobbold noted that there was evidence of some of these effects in Australia after only one year of publishing school results.

There is therefore the need for educational authorities to carefully weigh the advantages and limitations of league tables before publication.

Statement of the Problem

The introduction of District Performance League Tables (DPLT) based on the performance of districts at the BECE in Ghana at Annual Stakeholder Conferences for District Chief Executives and Directors of Education has generated some controversy. While proponents argue that it effectively leads to improved performance of schools/districts, critics point out that it rather has negative effects. This controversy has not been resolved because there are no empirical studies to determine the impact of the league tables on performance in schools/districts. The thrust of the study is therefore to find out the effect of the league tables on students' performance.

Purpose of the Study

The study sought to investigate the differences in performance in the years immediately before and after the introduction of District Performance League Table for the BECE at the Annual Stakeholders Conference for District Chief Executives and Directors of Education.

Research Questions

1. Is there a significant difference between the mean percentages of candidates that qualified for entry into senior high schools before and after the publication of the League Tables for 2013 BECE for
 - (a) total sample;
 - (b) sub-sample from the high and low performance categories;
 - (c) sub-sample for the deprived districts?
2. What is the trend in performance of the districts over the period from 2011 to 2016 for
 - (a) total sample;
 - (b) sub-sample from the high and low performance categories;
 - (c) sub-sample for the deprived districts ?

3. What is the effect (improve/decline) of the publication of the District Performance League Tables in terms of qualification for entry into senior high schools for
 - (a) total sample;
 - (b) sub-sample from the high and low performance categories;
 - (c) sub-sample for the deprived districts?

Research Design

The descriptive research design was used for the study.

Population

The population constituted all 216 districts in Ghana.

Sample

Districts in Ghana have been reorganized periodically in an attempt to decentralise the governance system. Some districts have been split and new ones have been created. One hundred and thirty (130) districts that have remained intact for the period (2011 – 2016) were purposively sampled for the study. Forty-three (43) deprived districts which were identified by the Ghana Partners for Educational Grant (GPEG) projects were sub-sampled for the study. Out of the 130 districts sampled for the study the top 50 districts and the lowest 50 districts in terms of performance were identified based on their ranking for the period 2011 – 2013. The selection of the top and lowest performing districts was for the purpose of comparison.

Data Collection

The Basic Education Division of the GES provided the District Performance League Tables for the period 2011 – 2016 with the corresponding percentages of candidates who qualified for entry into senior high school (i.e. percentages of candidates who obtain aggregate 06 – 30 in 6 subjects including 4 core subjects).

Data Analysis

1. The percentage of candidates from
 - (a) each district that qualified for entry into senior high school was compiled yearly for the period 2011 – 2013.

- (b) the BECE results from 2011 to 2013 was used to compute the mean percentages of candidates who qualified for entry into senior high schools before the publication of the League Table for 2013.
 - (c) the results of the BECE from 2014 to 2016 were used to compute the mean percentages of candidates who qualified for entry into senior high school after the publication of the League Table for 2013.
2. T-tests were used to find out if there were significant differences between the mean of percentages of candidates that qualified for entry into senior high school before and after publication of the League Table for 2013 BECE
- (a) total sample,
 - (b) sub-samples from the high and low performance categories,
 - (c) sub-sample for the deprived districts.
3. A line graph showing the mean percentages of candidates that qualified for entry into senior high school per years drawn for
- (a) total sample,
 - (b) sub-samples from the high and low performance categories,
 - (c) sub-sample for the deprived districts.
4. A table showing differences in the means percentages of candidates that qualified for entry into senior high school before and after the publication of the 2013 BECE results were computed for the
- (a) total sample,
 - (b) sub-samples from the high and low performance categories
 - (c) sub-sample for the deprived districts.

Findings and Discussion

Research Question 1: Is there a significant difference between the mean percentages of candidates that qualified for entry into senior high school before and after the publication of the League Table for 2013 BECE for the total sample, sub-samples from the high and low performing districts for 2013 and sub-sample for the deprived districts?

Results

The results are presented in Tables 1, 2 and 3. This question was answered using t-tests statistics.

Table 1: Differences between mean percentages of candidates that qualified for entry into senior high school before and after the publication of the League Table for 2013 BECE for the Total Sample

| Number of Districts | Mean Percentages of Candidates | | p |
|---------------------|--------------------------------|-------------------|------|
| | Before publication | After publication | |
| 130 | 51.48 | 36.47 | .000 |

Significant $p < .05$

Table 1 shows that there was a significant difference in the mean percentages of candidates that qualified for entry into senior high school before and after the publication of the League Table for the total sample of districts.

Table 2: Differences the between mean percentages of candidates that qualified for entry into senior high school before and after the publication of the League Table for 2013 BECE for the various performance categories of districts

| Performance Category | Number of Districts | Mean Percentages of Candidates | | P |
|----------------------|---------------------|--------------------------------|-------------------|------|
| | | Before publication | After publication | |
| High | 50 | 70.63 | 53.57 | .000 |
| Low | 50 | 33.88 | 22.53 | .000 |

Significant $p < .05$

Table 2 shows that there was a significant difference between the mean percentages of candidates that qualified for entry into senior high school before and after the publication of the League Table for 2013 BECE. The mean percentages of candidates qualifying for entry into senior high school from the high performance category remained higher than the low performance category over the period of study. The Ghana National Aptitude Tests

and Examinations Committee (2006) however noted that the distribution of school facilities was not uniform and therefore comparison of performance of schools was inappropriate. Sammons (1998) also noted that a key concern is the failure of league tables to compare “like to like”. He said the most damaging consequence has been the demoralization of schools working in disadvantaged areas.

Table 3: Differences between the mean percentages of candidates that qualified for entry into senior high school before and after the publication of the League Table for 2013 BECE for the sub-sample of deprived districts

| Number of Districts | Mean Percentages of Candidates | | p |
|---------------------|--------------------------------|-------------------|------|
| | Before publication | After publication | |
| 44 | 42.01 | 24.92 | .000 |

Significant $p < .05$

Table 3 shows that there was a significant difference between the mean percentages of candidates that qualified for entry into senior high school from the deprived districts before and after the publication of the league table. The mean percentage of candidates that qualified for entering into senior high schools from the deprived districts is low compared to the national average. This is similar to the findings of Trevor (2010) who reported that most of the schools serving in the most disadvantaged communities in Australia were given red flags on a website, meaning their results were substantially below average.

Research Question 2: What is the trend in performance of the districts over the period from 2011 to 2016 for the total sample, the sub-sample from the high and poor performance categories for 2013 and the sub-sample for the deprived districts?

Results

In answering research question 3, the mean percentages of candidates that qualified for entry into senior high schools per year were computed and line graphs drawn. The results are shown in Tables 7, 8, 9 and Figures 1, 2 and 3 below:

Table 7: Mean percentages of candidates who qualified for entry into senior high schools for the total sample

| Year | Mean Percentage |
|------|-----------------|
| 2011 | 53.21 |
| 2012 | 55.28 |
| 2013 | 45.95 |
| 2014 | 31.53 |
| 2015 | 32.61 |
| 2016 | 45.26 |

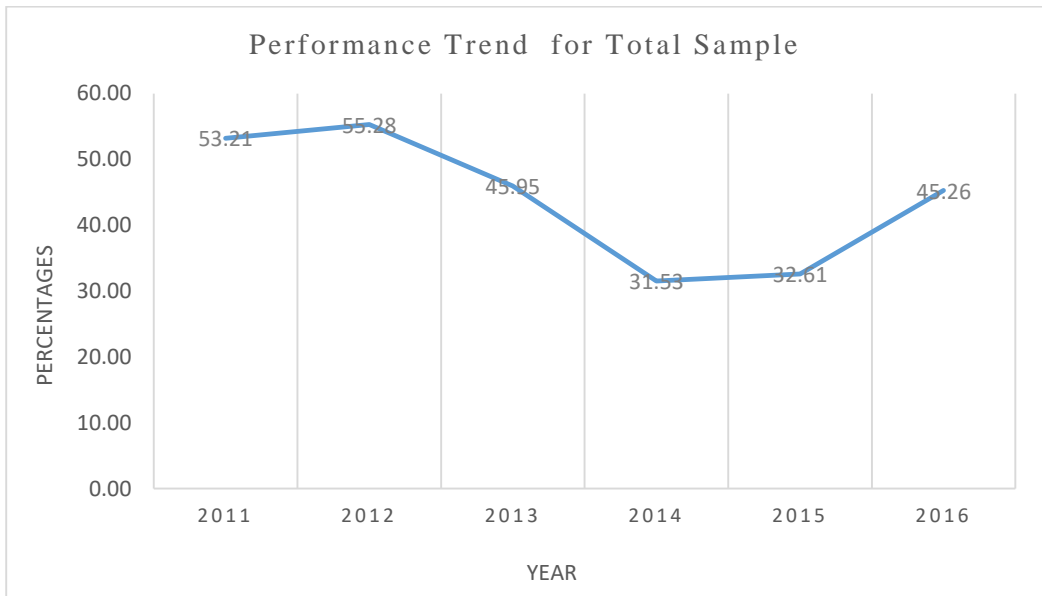


Figure 1

From Table 7 and Figure 1, the mean percentages of candidates who qualified for entry into senior high schools fluctuated during the period of study.

Table 8: Mean percentages of candidates who qualified for entry into senior high schools from the sub-samples of high and low performance categories

| Year | Mean Percentages | |
|------|------------------|--------------|
| | High Category | Low Category |
| 2011 | 69.04 | 36.84 |
| 2012 | 75.89 | 36.50 |
| 2013 | 66.96 | 28.31 |
| 2014 | 49.00 | 17.67 |
| 2015 | 50.09 | 19.03 |
| 2016 | 61.62 | 30.88 |

Figure 2

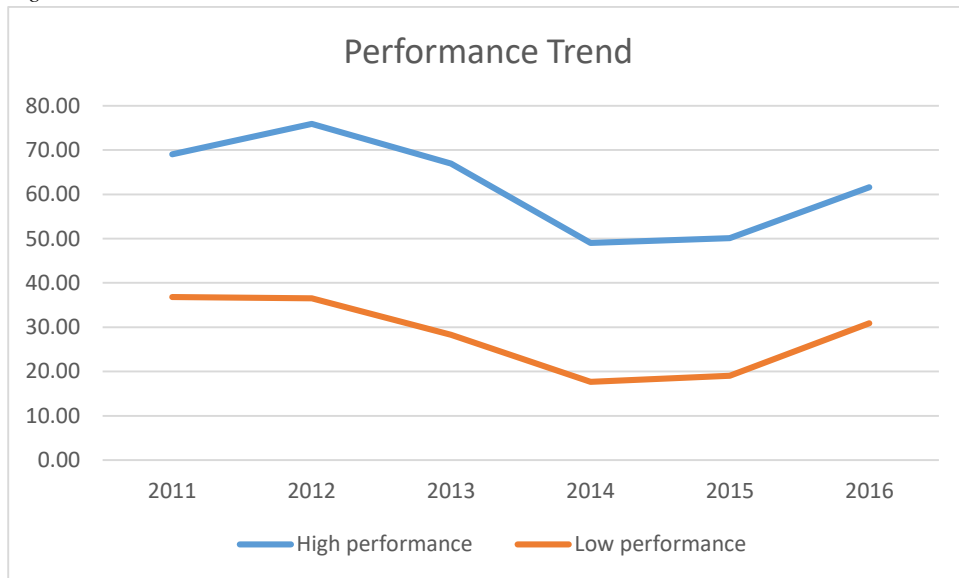


Figure 2

From Table 8 and Figure 2, the mean percentages of candidates who qualified for entry into senior high school for both performance categories fluctuated during the period of study.

Table 9: Mean percentages of candidates who qualified for entry into senior high school for the sub-sample for the deprived districts

| Year | Mean Percentages |
|------|------------------|
| 2011 | 44.10 |
| 2012 | 44.63 |
| 2013 | 37.30 |
| 2014 | 21.57 |
| 2015 | 21.04 |
| 2016 | 32.14 |

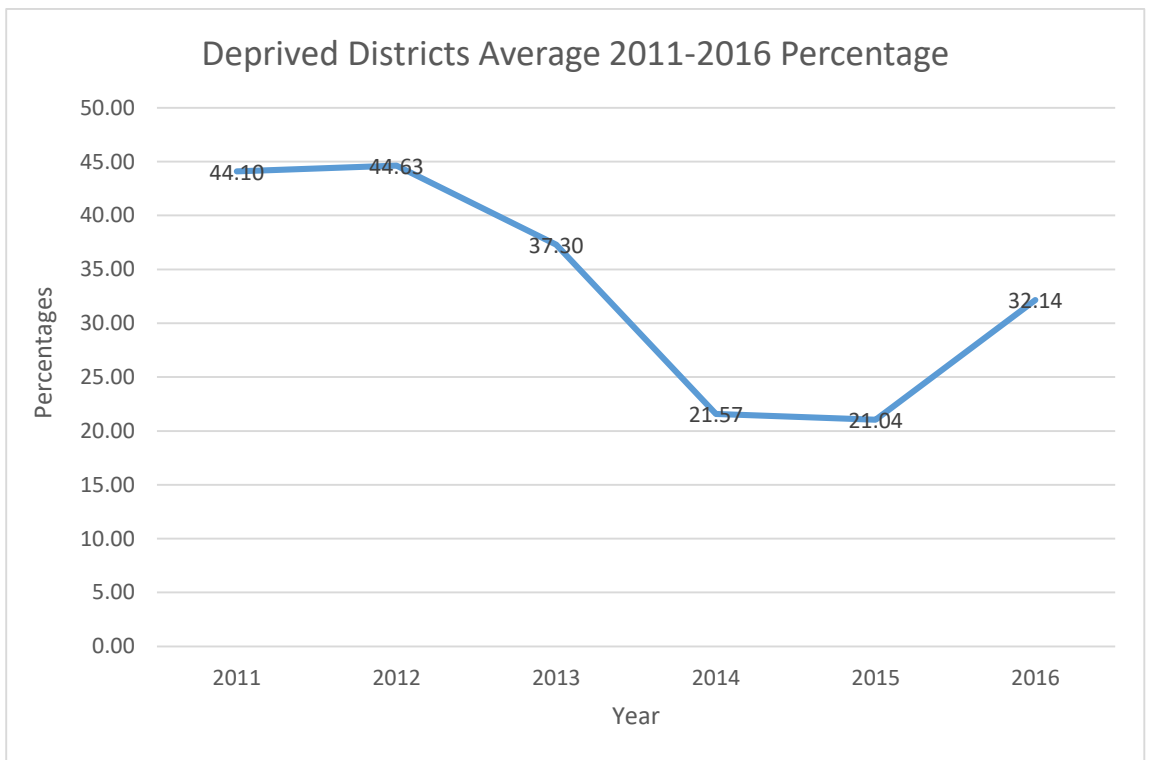


Figure 3

From Table 9 and Figure 3, the mean percentages of candidates that entered senior high schools for the sub-sample of deprived districts fluctuated during the period of study.

Research Question 3: What is the effect (improve/decline) of the publication of the District

Performance League Tables in terms of qualification for entry into senior high schools for the total sample, the sub-sample from the high and low performance categories, and sub-sample for the deprived districts?

Results

The results are presented in Tables 10, 11 and 12.

Table 10: Trend in performance of the mean percentages of candidates that qualified for entry into senior high school before and after the publication of the League Table for 2013 BECE for the total sample

| Number of Districts | Effect | |
|---------------------|---------------------|---------------------|
| | Percentage Improved | Percentage Declined |
| 130 | 6.92 | 93.08 |

Table 10 shows that only 6.92% of the districts recorded improvement in their mean percentages after the publication of the league table for 2013 BECE for the total sample.

Table 11: Trend in the mean percentages of candidates that qualified for entry into senior high school before and after the publication of the League Table for 2013 BECE for the high and low performance categories

| Performance Category | Number of Districts | Effect | |
|----------------------|---------------------|---------------------|---------------------|
| | | Percentage Improved | Percentage Declined |
| High | 50 | 8.00 | 92.00 |
| Low | 50 | 8.00 | 92.00 |

Table 11 shows that for both high and low performance categories of districts recorded the same trend in mean percentages of candidates that qualified for entry into senior high schools before and after the publication of the league tables.

Table 12: Trend in performance of the mean percentages of candidates that qualified for entry into senior high school before and after the publication of the League Table for 2013 BECE for the poor performing deprived districts

| Number of Districts | Effect | |
|---------------------|---------------------|---------------------|
| | Percentage Improved | Percentage Declined |
| 43 | 6.98 | 93.02 |

Table 12 shows that only 6.98% of the deprived districts recorded improvement in the mean performance before and after the publication of the league tables. All 43 districts that constituted the sub-sample for deprived districts received support from the government of Ghana through the Ghana Partnership for Education Grant (GPEG).

Summary of Findings

1. There were significant differences between the mean percentages of candidates from the districts that qualified for entry into senior high school before and after the publication of the league table for 2013 BECE.
2. Majority of the districts recorded declined in the mean percentages of candidates that qualified for entry into senior high school after the publication of the league table for 2013 BECE.
3. The mean percentages of candidates that qualified for entry into senior high school from the sub-sample deprived districts was low compared with that for the total sample.

Conclusion

Generally, the presentation of BECE District Performance League Table at Annual Stakeholders Conferences in Ghana did not have much positive impact on the performance of districts over the period covered by the study even though a few low performing districts recorded improved performance.

Recommendations

1. There should be a follow up research to identify factors responsible for the decline in performance with respect to the percentage of

candidates who qualified for entry into senior high schools.

2. Participation at the Stakeholders Conferences for District Chief Executives and Directors of Education at which the District Performance League Tables were presented should be expanded to include other interest groups.
3. The Ghana Education Service (GES) and the National Inspectorate Board (NIB) should intensify monitoring and supervision of activities of schools/Districts.
4. District Assemblies should commit more resources for improving performance of weak districts upon identifying such districts using the District League Tables.

References

- Baku J. K. (2005). The SSSCE League Table: Is it saying what it is meant to say? WAEC monthly Seminar, April 2005.
- Cobbold T (2010). Publication of School Results and League Tables Harm Education <http://www.saveourschools.com.au>.
- Ghana Partnership for Education Grant Project Implementation Manual (2013).
- Reed J., Hallgarten J. (2003). Time to say Goodbye? The Future of School Performance Tables. <http://www.leads.ac.uk/educol/documents>
- Sammons P. (1998). School League Tables, 1998: Deprived arrears still at a disadvantage. <http://www.independent.co.uk/life-style>
- The West African Examinations Council, (2005 & 2006).

Effectiveness of Measures by Uganda National Examinations Board to Control Malpractice: Perspectives of Secondary School Students

Dan Nokrach Odongo

Uganda National Examinations Board

Dan.Odongo@uneb.ac.ug; danodongo22@yahoo.co.uk

Abstract

An examination measures students' acquisition of knowledge and skills, and is an integral part of any education system. Examination malpractices, however, pose a serious problem to examination bodies because of their impact and consequences on validity and reliability of the examination. Like other examination bodies the world over, Uganda National Examinations Board (UNEB) has instituted measures to curb this vice. However, cases of malpractice continue to be reported annually. This study was conducted to establish the level of secondary school students' awareness of examination malpractice and the effectiveness of the measures instituted by UNEB from their perspective. A survey design was adopted for the study. A stratified random sample of 100 secondary school students was selected to participate in the study. A questionnaire with both open and closed ended items was constructed and administered to the respondents to fill. Ninety-one usable questionnaires were analyzed. The results of the study indicated that secondary school students are aware of examination malpractice, the causes and the perpetrators of the vice. Respondents also indicated that the measures put in place by UNEB to curb malpractice are effective. Recommendations are made on what other measures can be taken.

Key words: Examination, malpractice, control, perspective

INTRODUCTION

According to Animasahun & Ogunniran (2014), an examination is a measurement of proficiency in knowledge and skills, either in oral or written forms, and evaluating the adequacy of these properties possessed by the candidates. Akaranga & Ongong (2013) also observe that examination is a very integral part of any education system. The evaluation of the teaching learning process and the success of any educational system will depend on the efficacy of the examination system. Citing Wilayat (2009), they contend that when examination malpractice occurs, the validity and resulting outcome of an examination is questionable. Odongo (2014) observes that examination malpractice presents a serious problem to all examination bodies because of the negative impact it has on examinations. Not only does examination malpractice compromise the technical qualities – validity

and reliability of the examination, it also raises authenticity and credibility issues concerning the certificates and decisions based on the results of such an examination (Nworgu, Uchekwe & Nworgu, 2013). The view that examination malpractice is a threat to validity and reliability, and undermines credibility is further held by Jimoh (2009); Emaikwu (2012); the Joint Council of Qualifications (cited in Utlwang, 2013); Malami (2013) and Chaminuka & Ndudzo (2014). Adegoke (2010) states that the occurrence of examination malpractice at any level is the greatest threat to validity and reliability, and consequently, the authenticity and recognition of certificates issued. He further notes that examination malpractice is a problem that has been affecting the education system for many years, and seems to have defied solutions as all antidotes applied so far have been faulted by the fraudsters. Animasahun & Ogunniran

(2014) also state that examination malpractice is a cankerworm that poses great threat to authenticity of educational qualifications, and poses a challenge to examination bodies.

Examination malpractice is often described as any deliberate act of wrong doing contrary to the rules governing the conduct of examinations, the purpose of which is to place a candidate at an unfair advantage, or able frequently, to place a candidate at a disadvantage (Malami, 2013). Oyekan (1996), (cited by Adegoke, 2010) views examination malpractice as any act of indiscipline adopted by students or their privileged accomplices to secure facile success and advantage before, during or after the administration of a test or examination. According to Joint Council for Qualifications (2014), malpractice includes maladministration and non-compliance, and is an act, default or practice which is a breach of the regulations. Malpractice, the Council notes, compromises, attempts to compromise or may compromise the process of assessment, the integrity of any qualification or the validity of a result or certificate; and/or damages the reputation or credibility of any awarding body ...' (p.3). Whichever way examination malpractice is defined or described, the ingredients are the same. That it is an act of wrong doing committed intentionally or otherwise; it is done by an examination candidate(s) or other persons; it is done in violation of established rules and regulations; the purpose of the act is for the candidate(s) to gain an unfair advantage, although often the act also puts the candidate(s) at a disadvantage (Odongo, 2014).

Examination malpractice is known to occur before an examination is done such as when candidates get foreknowledge of the questions through a leakage, or during the conduct of the examination such as when invigilators allow external assistance to candidates, or even after the examination has been done (substitution of scripts, inflated marking or forgery of certificates (Odongo, 2014). Available literature indicates that

examination malpractice is widespread (Soyemi, 2002; Ogwang, 2007; Maheshwari, 2011; Obidoa, Onyechi & Okere, 2013 and Shakubanja & Nkoya, 2013), and no examination body appears to be safe from it and its side effects.

Odongo (2014) reported that in Uganda, examination malpractice occurs in all examinations every year. The number of cases varies from year to year. For example, in 2013 1,812 Uganda Certificate of Education (UCE) candidates were confirmed to have been involved in examination malpractice and had their entire results cancelled. In the 2014 UCE examination, almost a similar number, 1,936 had their results withheld pending completion of investigations. Most of the cases reported were as a result of assistance given to candidates in science practical examination papers and Mathematics. Teachers also disclosed to candidates the list of requirements sent confidentially by UNEB for the preparation of the practical tests, while other mounted last minute practical exercises based on the confidential lists.

Statement of the Problem

Like other examination bodies, UNEB has instituted measures to control involvement of students in examination malpractice. These measures include, issuing detailed regulations on conduct of examination and ensuring that the regulations reach the candidates; sensitization briefings conducted countrywide for all persons involved in examination field management; detailed briefing of all candidates before the start of the examination, use of invigilators drawn from other schools and intense monitoring by scouts. Despite these measures, candidates are still consciously being involved in examination malpractice. How effective are these measures from the point of view of the candidates?

Purpose of the study

The purpose of this study was to investigate the effectiveness of measures taken by UNEB to curb malpractice from the perspective of the students

themselves with a view to improving the management of examination.

Objectives of the study

The objectives of the study were:

1. To find out if students in secondary schools were aware of the nature and causes of examination malpractice.
2. To seek the views of the secondary school students on the effectiveness of the measures UNEB has in place to control examination malpractice.
3. To seek the views of the secondary school students on what other measures UNEB could use to control examination malpractice.

Scope of the study

The target population from which the sample was taken consisted of the 2015 students from all over the country in Senior 4 class preparing for the UCE examination and the Senior 6 class preparing for the Uganda Advanced Certificate of Education (UACE) examination. The study was restricted to establishing the respondents' level of awareness about malpractice, their perspective on the effectiveness of the measures UNEB is taking to curb malpractice over the years, and their views on other measures UNEB could take to curb malpractice.

Significance of the study

The views and recommendations made by the respondents should help UNEB to appraise its examination security system and procedures adopted in field administration of examinations with a view to improving on the control of examination malpractice. The study should also help schools to review their methods of preparation of their students for public examinations and institute counseling services for their students.

Literature Review

Causes of malpractice among students

Sigauke (2004) in a study designed to gather views of Harare (Zimbabwe) secondary school students on

examination malpractice reported that students indulged in examination malpractice so as to pass with high grades. Others did so due to pressure from their parents, or because they had not prepared themselves adequately for the examination, or because they found the examination difficult. Alutu & Aluede (2006) also cited pressure from parents and teachers on their children or students to perform well as being a driving force behind examination malpractice. Ngungu (2011) observed that students are motivated to cheat due to lack of time for preparation for the examination, the failure of teachers of teachers to prepare the candidates adequately, pressure from parents and teachers and poor invigilation that enables candidates to enter the examination room with written materials.

Akaranga & Ongong (2013), in a study of malpractice cases in two universities in Kenya reported non – completion of the course outline by the end of semester as a cause of malpractice. They further noted that public release of examination results were rankings are released also fuel examination malpractice. Then there are lazy student who do not put in necessary efforts but still want to aspire for success without corresponding hard work. Poor teaching and learning, poor invigilation and lack of seriousness have also been reported by Obioda, Onyechi & Okere (2013) as causes of malpractice among students in public secondary schools in Anambara State, Nigeria. In an address to a Stakeholders' forum, Mallami (2013) observed that many parent will resort to various corrupt practices to ensure that their children secure the all – important certificate which represents an avenue for poor students to secure non menial jobs. He noted that were students and parents perceive conditions of learning to be inadequate, they may feel that they have no option than to resort to malpractice.

Animasahun & Ogunniran (2014) observe the most important factor responsible for malpractice is the student personality factor, which culminates in the students' desire to pass at all costs – fueled by their lack of preparation, lack of confidence and poor school facilities.

They note that while some candidates get there intentionally, others are driven there by peer pressure, ignorance of regulations, fear of scoring low grades, inadequate coverage of the syllabus and poor sitting arrangements where candidates sit close to each other. Chaminuka & Ndudzo (2014), in their study on students and staff perceptions on malpractice and fraud in higher education in Zimbabwe, also indicated inadequate preparation of students, peer pressure, inadequate school facilities, lack of self-confidence due to laziness of the student and poor academic performance as causes of malpractice. Odongo (2014) cited three earlier studies by UNEB that found out the main drivers of malpractice in Ugandan schools were the desire for good results by schools, candidates and their parents, financial gain and moral degeneration. Success in public examinations is often seen as a stepping stone to success in life, and this exerts pressure on the students to pass and increases the propensity to engage in malpractice (Wasanga & Mwiruri, 2002; Ogwang, 2007; Maheswari, 2011 cited in Odongo, 2014). This study tries to find out if the students in Uganda Secondary schools are aware of these causes of examination malpractice.

Effectiveness of measures to control malpractice

In view of the devastating effects of examination malpractice on the education system, examination authorities have taken various measures to control the menace. What is certain is that, despite the measures in place, the vice thrives. Jimoh (2009), commenting on the situation in Nigeria observed that as examination bodies institute control measures, every examination season witnesses the emergence of new and ingenious ways of cheating. As Adegoke (2010) puts it, the vice seems to have defied solutions as all the antidotes applied so far have been faulted.

Legislation has been used by countries as a measure to control examination malpractice, for example The Examination Malpractice Act of 1999 in Nigeria; the Kenya National Examinations Council Act, 2012. These

provide for relatively stiff penalties for offenders. In Uganda, the law, (Uganda National Examinations Board Act, Cap 137) is being revised extensively to provide for more appropriate penalties. The persistence of examination malpractice despite the existence of legislation may be attributed to non-enforcement of the law (Nworgu, Uchekwe & Nworgu, 2013, citing Uwadie, 2003). In Uganda, the court process is long and frustrating. Until quite recently, the officers of the Criminal Investigations department of the police and officers of the directorate of public prosecutions did not appreciate the nature and consequences of examination malpractice. As a result, reported cases received little attention and were not pursued seriously. The few convictions secured only attracted minimal fines. For example, a secondary school student who was arrested writing the 2014 Primary leaving examination for a registered candidate was convicted and sentenced to a caution!

Measures taken by UNEB to control examination malpractice by secondary school students

As pointed out by Animasahun & Ogunniran (2014), some candidates get intentionally involved in examination malpractice. Others may do so through ignorance of regulations, carelessness or forgetfulness (Joint Council for Qualifications, 2014). UNEB has put in place measures aimed at controlling malpractice from the side of the candidates. Detailed rules and regulations on the conduct of examinations are printed and published. Crucial sections relevant to the candidates are extracted and printed along with the examination timetables and issued to each candidate. Prohibited acts and the consequences of breaches are clearly spelt out. The timetable sets aside one day, immediately before the start of the examination, for the head teachers to take their candidates through a briefing session during which the regulations and the repercussions of involvement in examination irregularities are explained.

UNEB officers conduct countrywide seminars with heads of examination centres and Chief invigilators (supervisors) appointed to take charge of the management of the examination at various examination centres. During the seminars, the rules and regulations are discussed in detail with great emphasis is placed on examination malpractice and other security breaches, the consequences of such malpractice and security breaches, and how they should be avoided, or dealt in case of occurrence.

Chief invigilators and the invigilators are trained and registered teachers in active service, or recently retired. These are appointed and deployed to examination centres with which they have no links to ensure that they do not know the candidates and can be independent of the school administration. To avoid complacency and possibility of compromise, the invigilators are regularly transferred to other examination centres during the course of the examination. The chief invigilators are provided with albums containing the photographs and other biodata of the candidates. They use these to check and verify the identity of each candidate as they enter the examination halls. They are also under instruction to thoroughly check the candidates to ensure that candidates are not carrying any unauthorized materials, notebooks, textbooks or any other media on which notes are written. Invigilators also check the seating arrangement in the halls to ensure that each candidate has a desk alone spaced in such a manner that a candidate is not less than 1.2m from the next one to avoid possible collusion or copying ('giraffing'). An invigilator is allocated not more than 35 candidates to watch over to ensure close supervision.

To supplement the external invigilators, UNEB deploys scouts throughout the country to make unannounced spot checks on examination centres. The scouts are UNEB staff, Directorate of Education Standards inspectors, senior examiners and other educators UNEB may appoint on their individual merit. They monitor the conduct of the examination at the centres they visit and oversee the

implementation of rules and regulations on the conduct of the examination and report any suspected cases of malpractice or security breaches to security agencies or UNEB for further action. Similar strategies to control examination malpractice have been adopted by other assessment bodies. Aworanti (2012) writing about the efforts of National Business and Technical Examinations Board (NABTEB), cites intensive campaigns; monitoring of the examination process by governing council members, ministry of education officials and the officials of the board; training and appointing credible persons as supervisors before the start of the examination; and sanctioning offenders. The Joint Council for Qualifications (2014) issued a *Warning to Candidates* with a directive that it must be displayed in a prominent place outside each examination room. The warning contains what candidates must or must not do.

Improvements have been made by UNEB in the marking of candidates' scripts by using the 'conveyor belt' system where a script of a candidate is scored by more than one marker to avoid bias and compromise. Certificates of candidates of the Uganda Advanced Certificate of Education are personalized using the photographs submitted at registration to curb cases of impersonation, which was particularly rife at this level. Despite the measures, malpractice in examinations has persisted. This study sought to establish the effectiveness of these measures from the point of view of the students.

Other measures to control examination malpractice

Various studies have made recommendations on the control of examination malpractice including use of the police, banning of culprits, counseling (Sigauke, 2004); strict and thorough invigilation, counseling, enforcement of penalties (Oniye & Alawaye, 2008); checking of students' pockets as they enter the examination rooms, cancellation of results of the candidates involved (Adeyemi, 2010); attractive remuneration of examination supervisors to reduce chances of being bribed, de-emphasizing the supremacy of certificates over skills and

professional competence (Malami, 2013); good coverage of the syllabus, providing adequate equipment in schools, counseling (Obidoa, Onyechi & Okere, 2013). This study also sought to find out from the students what other measures they would wish to see implemented by UNEB to curb examination malpractice.

Research Questions:

1. Are students in secondary schools aware of the nature and causes of examination malpractice?
2. What other measures do students in secondary schools recommend to UNEB to curb examination malpractice?

Hypothesis One: There is no significant difference in the number of students who assessed the measures taken by UNEB as effective and those who assessed the measures as not effective.

$$H_0: \sum \left(\frac{O-E}{E} \right)^2 = 0$$

Using Chi – Squared, one variable test, two tailed.

Methodology

A survey research design was adopted for the study as it centered on individuals and their opinions. The study population consisted of all secondary school students in Senior Four and Senior Six classes in Uganda who are due to sit for the 2015 UCE (O Levels) and UACE (A Levels). A stratified random sampling technique was used. First, the country was divided into the traditional four regions of Northern, Eastern, Central and Western regions. Then the schools in each region were separated into two categories; those that had only UCE level and those that had both UCE and UACE levels. Five schools with both UCE and UACE were randomly selected from each region to make a total of 20 schools. Then from each school, five students were selected from the combined list of both UCE and UACE students. The total sample size was 100 students.

Instruments

A questionnaire with both closed-ended and opened-ended items was constructed. The closed-ended items had a two point response scale of Effective, and Not Effective. The instrument was validated by having it peer reviewed by a senior researcher. The Cronbach alpha coefficient was 0.70, indicating that the instrument was suitable for use.

Data collection and analysis of data

The questionnaires were administered to the sampled respondents by research assistants. A total of 91 filled and useable questionnaires were returned and analyzed. Results are presented using descriptive statistics of frequencies and percentages for the closed-ended items. The Chi-square, one variable test was used to establish the significance of difference in the responses to levels of effectiveness of the control measures. For the open-ended items, responses are grouped according to common themes and their frequencies and percentages reported according to the research questions.

Results

All the 91 respondents who filled the questionnaires were students preparing to sit for a UNEB examination. They had sat at least one UNEB examination (Primary Leaving Examination (PLE) for the UCE students, and both PLE and UCE for the UACE students) before. They were, therefore, considered knowledgeable enough to respond adequately to the questionnaires. Forty (44.0%) of them were UCE level students while 49 (53.8%) were UACE students. Two respondents did not indicate their level. Fifty six (61.5%) of the respondents were male while female students were 34 (37.4%). One respondent did not indicate the gender.

Research Question 1: Are secondary school students aware of the nature and causes of examination malpractice?

All respondents, 91 (100%) were aware of malpractice in examinations conducted by UNEB. Respondents were asked whether they had ever witnessed, or known of the occurrence of malpractice. Fifty seven (62.6%) of the respondents reported to have witnessed or known of the occurrence of malpractice, while 34(37.4%) said they had not witnessed any form of malpractice. Of those respondents who had witnessed or known about examination malpractice, 30 (52.6%) mentioned external assistance given by teachers or invigilators; 27 (47.4%) mentioned smuggling of notes, books and material

written on handkerchiefs or even body parts. Foreknowledge, especially of science practical examinations was mentioned by 24 (42.1%) respondents. Some of these mentioned night teaching done on the eve of the practical examination as the way students got foreknowledge of the examination. Collusion between candidates and copying (‘giraffing’) was mentioned by 15 (26.3%) of the respondents, while 11 (19.3%) mentioned entering with mobile phones/use of the internet. Others with lower frequencies were impersonation, inflation of marks by markers related to certain candidates, and substitution of scripts.

Respondents were asked what, in their opinions, were the reasons for students engaging in examination malpractice. Table 1 summarizes their responses.

Table1:
Reasons for *Malpractice*

| Reason | Frequency | % | Rank |
|---|-----------|------|------|
| Inadequate preparation of students | 60 | 65.9 | 1 |
| The desire by students to pass with high grades | 43 | 47.3 | 2 |
| Fear by students to fail examinations | 22 | 24.2 | 3 |
| Lazy students who did not study for the exam to enhance the school reputation | 19 | 20.9 | 4 |
| To appear in the media as best students | 14 | 15.4 | 6 |
| Peer pressure | 13 | 14.3 | 7 |
| Pressure from parents for good results | 12 | 13.2 | 8 |
| To join higher institutions | 11 | 12.1 | 9 |
| Poor invigilation | 11 | 12.1 | 9 |
| Lack of self-confidence | 10 | 11.0 | 11 |
| Exams are hard | 08 | 8.9 | 12 |
| To get government scholarship | 07 | 7.7 | 13 |
| Academically weak students | 06 | 6.6 | 14 |
| School culture/tradition | 03 | 3.3 | 15 |
| Lack of guidance | 03 | 3.3 | 15 |

N = 91

Table1 shows a number of reasons why students get involved in examination malpractice. Respondents

(65.9%) indicated poor preparation of students due to poor teaching or low syllabus coverage as the main cause

of malpractice. The desire by students to pass with high grades (47.35 %) is the next major driving force behind students' involvement in examination malpractice. The students also mentioned the fear of failure (24.2%), the lazy students who did not prepare well for the examination (20.9%); the urge to appear high in the student rankings published by newspapers (15.4%); peer pressure (14.3%) and pressure from parents for good

results (13.2%). Respondents also mentioned the wish to join higher institution, poor invigilation, lack of self-confidence and hard examinations, among others, as the causes of examination malpractice.

Respondents were also asked who, in their opinion, were involved in or promoting malpractice. Seventy seven respondents answered the item. Their responses are summarized in Table 2.

Table 2:
Categories of Persons involved or promoting Examination Malpractice

| Category involved | Frequency | Percentage | Rank |
|---------------------------------|-----------|------------|-----------|
| Teachers | 49 | 63.6 | 1 |
| Invigilators and supervisors | 29 | 37.7 | 2 |
| Head teachers | 22 | 28.6 | 3 |
| Parents | 15 | 19.5 | 4 |
| Peers (fellow students) | 11 | 14.3 | 5 |
| Markers | 08 | 10.4 | 6 |
| School directors (owners) | 06 | 7.8 | 7 |
| Bribed Exam officials | 05 | 6.5 | 8 |
| Scouts who are bribed | 04 | 5.2 | 9 |
| Setters | 03 | 3.9 | 10 |
| Lab attendants | 01 | 1.3 | 11 |
| Police | 01 | 1.3 | 11 |
| Those against government | 01 | 1.3 | 11 |

N=77

Table 2 indicates that from the student's point of view, teachers (63.6%) are the leading promoters of malpractice in examinations conducted by UNEB, followed by the invigilators (37.7%) who are said to be compromised by the schools. Head teachers (28.6%) have featured prominently as have the parents (19.5%) who are said to contribute the money to bribe invigilators or secure help for their children. Fellow students (14.3%) are also said to influence each other to commit malpractice. Respondents have mentioned school directors, bribed

examination officials and scouts, setters of examination items as persons fueling malpractice among students.

Respondents were asked if, given the chance to do so, they would cheat in a UNEB examination. Twenty of the respondents (22.0%) said yes, they would if they got the chance, while 71 respondents (78%) said No, they would not cheat. Of those who admitted they would cheat, 13 (65%) gave the desire to pass well as the reason why they would cheat. Four respondents (20%) said they would do it because other students do it, get good results and are

not punished. They also mentioned the quest for government scholarship and competition with other schools as other reasons for their willingness to cheat.

The majority, who said they would not cheat even when there was a chance to, gave various reasons, which are summarized in Table 3 below.

Table 3:

Reasons for not cheating in examinations

| Reason | Frequency | % | Rank |
|--|-----------|------|------|
| I want to know how far I have achieved | 13 | 18.3 | 1 |
| Cheating lowers my dignity | 08 | 11.3 | 2 |
| You may be given wrong answers and you fail | 07 | 9.9 | 3 |
| It is bad to pass without your own efforts | 07 | 9.9 | 3 |
| My results may be cancelled | 06 | 8.5 | 5 |
| The gain is short-lived one may live to regret | 06 | 8.5 | 5 |
| I have enough material to pass | 05 | 7.0 | 7 |
| Cheating will kill my creativity and capacity | 05 | 7.0 | 7 |
| It is not the way to achieve your dreams | 04 | 5.6 | 9 |
| I may become someone I do not want to be | 04 | 5.6 | 9 |
| It breaks the Ten Commandments | 02 | 2.8 | 11 |
| It is against the law and purpose of the examination | 02 | 2.8 | 11 |
| Lowers academic competence among students | 01 | 1.4 | 13 |
| Shows your mental weakness | 01 | 1.4 | 13 |

N = 71

Table 3 shows several reasons why the students will not cheat. The main reason given by 18.3 percent of the respondents is that they would genuinely wish to see what they have achieved as measured by the examination. Students are also worried that cheating would lower their dignity (11.3%). Other respondents (9.9%) think that answers given may even be wrong and one would fail as a result. A similar number do not believe in passing without their own efforts, while 8.5 percent fear that the cheating may be discovered and their results may be cancelled or that the stigma may remain with them and make them regret later. Other respondents do not see the need to cheat because they are confident that they have enough material. Respondents also mention other negative consequences of cheating.

Hypothesis One: There is no significant difference in the number of students who assessed the measures

taken by UNEB to control malpractice as effective and those who assessed the measures as not effective.

$$H_0: \sum \left(\frac{O-E}{E} \right)^2 = 0$$

Respondents were asked for their opinion on the effectiveness of some of the measures being taken by the Board to control malpractice. Respondents were to choose either ‘Effective’ or ‘Not effective’ and respondents grouped accordingly. The SSPS programme was used to compute the χ^2 values.

A chi-squared, one variable test at $\alpha = 0.05$ was adopted to test the level of significance of the difference. Table 4 presents the results.

Table 4:

Frequency and percentage distribution and Chi square test of secondary school students' responses to effectiveness of measures taken to curb malpractice in examinations.

| Measures taken | Responses on effectiveness | | Test of significance | | |
|---|----------------------------|------------|----------------------|----|------|
| | Not effective | Effective | Chi square | df | sign |
| Publishing regulations on conduct of exams | 23 (25.8%) | 66 (74.2%) | 20.775 | 1 | .000 |
| Briefing of headteachers and invigilators | 20 (22.0%) | 71 (78.0%) | 28.582 | 1 | .000 |
| Briefing of candidates before UNEB exam starts | 13 (14.4%) | 77 (85.6%) | 45.511 | 1 | .000 |
| The script of a candidate is marked by more than one examiner | 10 (11.0%) | 81(89.0%) | 55.396 | 1 | .000 |
| Sending scouts to spot check on schools during exam time | 23 (25.6%) | 67 (74.4%) | 21.511 | 1 | .000 |
| Using invigilators from other schools or places | 18 (18.8%) | 73 (80.2%) | 33.242 | 1 | .000 |
| Using photo albums to check candidates' identity | 13 (14.3%) | 78 (85.7%) | 46.429 | 1 | .000 |
| Searching or checking candidates | 23 (25.3%) | 68 (74.7%) | 22.253 | 1 | .000 |
| Printing certificates bearing candidates' photos | 14 (15.6%) | 76 (84.4%) | 42.711 | 1 | .000 |

*p>0.05

Table 4 shows that measures taken by UNEB to curb malpractice are effective from the point of view of the secondary school students. The Chi-square values are all higher than expected minimum values at $df=1$ and $p = 0.000$. The null hypothesis was, therefore, rejected.

Respondents who thought that publishing regulations was not effective indicated that head teachers and invigilators do not adhere to the regulations and no action is taken against those who do not adhere to the rules and regulations. While conducting countrywide briefing for head teachers and invigilators before the examination is effective, respondents who think otherwise pointed out that many head teachers, teachers and invigilators are corrupt and the briefing does not change them. Respondents also think that using several markers to mark a script of a candidate is effective. Those with contrary views hold that some markers may know the names of some candidates and still mark them favourably. Sending scouts to monitor the examination in the field is

effective but those who thought this measure was not effective had various reasons, for instance, that the scouts are bribed; they are at a centre only briefly and at a specific period, or arrive at a centre after malpractice has already occurred. The use of personnel from other schools as invigilators is also reported as effective but those who think otherwise stated that those teachers are bribed, or could be part-time teachers at the centres to which they are posted as invigilators. Searching the candidates as they enter the examination hall is also effective. Respondents who did not think so cited lazy invigilators who do not check properly. They also pointed out that some candidates hide material in places that cannot be checked easily, for example, female candidates may hide notes in their bras. Use of photo albums and personalization of certificates were said to be effective. Those respondents who did not think so did not give any clear reasons for their responses.

Research Question 2: What other measures do secondary school students recommend to UNEB to curb examination malpractice?

Respondents were asked to recommend measures which, in their views would help curb examination malpractice. Seventy eight of them answered the item. Their responses are summarized in Table 5.

Table 5:

Recommendations made by the secondary school students on measures to curb malpractice

| Recommendation of students | Frequency | Percentage | Rank |
|--|-----------|------------|------|
| Ban or close centres found in malpractice | 13 | 16.7 | 1 |
| Improve security of exams to avoid leakages | 13 | 16.7 | 1 |
| Have effective invigilation and supervision | 10 | 12.8 | 3 |
| Ban candidates involved | 09 | 11.5 | 4 |
| Imprison offenders for up to 10 years | 08 | 10.3 | 5 |
| Use police officers, not head teachers, to collect exams | 07 | 9.0 | 6 |
| Hire invigilators from different districts and rotate them | 06 | 7.7 | 7 |
| Students should be prepared well | 06 | 7.7 | 7 |
| Classroom teachers should not set or mark examinations | 05 | 6.4 | 9 |
| Pay invigilators well to avoid being corrupted | 04 | 5.1 | 10 |
| UNEB should set easier exams to decrease candidates' fear | 04 | 5.1 | 10 |
| Ban invigilators found assisting candidates | 03 | 3.8 | 12 |
| Sensitize people on the evils of malpractice | 03 | 3.8 | 12 |
| Use examination officials who have integrity | 03 | 3.8 | 12 |
| Ensure that each candidate has a copy of the rules | 02 | 2.6 | 14 |
| Stop using teachers to set examinations | 02 | 2.6 | 14 |

N = 78

Respondents saw the banning or closing of centres involved in malpractice (16.7%) and improvement in the security of examinations to avoid leakages (16.7%) as the most important measures UNEB should adopt to curb examination malpractice. Ten of the respondents (12.8%) indicated that the people appointed to do so should effectively do invigilation and supervision of the examinations. Nine respondents (11.5%) recommended a ban on candidates caught indulging in malpractice. Severe punishments, including jail terms of up to 10 years for offenders, have also been recommended by eight (10.3%) respondents. Other measures recommended in with decreasing frequencies have included use of police officers, and not head teachers to collect examination question papers from storage stations; use of invigilators from different districts and rotating them; adequate preparation of students for the examination; banning

invigilators found abetting malpractice; adequate payment of invigilators so that they are not corrupted; sensitization on the evils of malpractice; use of persons with high levels of integrity in examination work and not using classroom teacher as markers of UNEB examinations. Four respondents think UNEB should set easier examinations to decrease the candidates' fear of examinations.

Discussion

The results of this study have shown that students in secondary schools in Uganda are aware of examination malpractice. The majority (62.6%) in this study have either witnessed or known of the occurrence of malpractice. The most common type they have encountered or known of is external assistance (52.6%), where candidates receive assistance from their teachers,

invigilators or other persons in the course of doing the examination. This finding is in conformity with that of Adegoke, (2010) who reported that students accept assistance offered by teachers by writing answers on chits which are then circulated from one student to another, dictation of answers or writing the answers on chalk boards for the students to copy. The finding also agrees with the observation of Adeyemi (2010) who cited examination officials assisting students as a serious type of malpractice. The second most common type of malpractice mentioned by the respondents (47.4%) was smuggling or bringing into the examination room unauthorized materials such as notes, sometimes written on body parts, textbooks, etc. available literature indicate that this is a common type of malpractice (Sigauke, 2004; Oyekan, 2010 cited by Adegoke, 2010; Adeyemi, 2010; Ngungu, 2011;). Respondents have also mentioned foreknowledge, especially instances where teachers responsible for the preparation for practical science examinations disclose aspects of the specimens, chemicals or equipment to the candidates. Respondents have reported night teaching, most probably using the confidential information, the night before. Collusion or copying from one another has been mentioned by a significant number of the respondents.

Respondents in this study have mentioned several categories of people as drivers of malpractice. Teachers are the leading promoters, mentioned 63.6 percent of the respondents. This is not surprising, given that the leading cause of malpractice, according to this study, is inadequate preparation of candidates by the teachers, especially when they do not complete the syllabus or course outline (Kagete, 2008; Adeyemi, 2010; Ngungu, 2011; Akaranga & Ongong, 2013; Obidoda, Onyechi & Okere, 2013; Chaminuka & Nduzo (2014; Odongo, 2014; Uzochukwu, 2015). They will offer assistance to cover for their deficiencies, and students who are aware that they were not adequately prepared will be tempted to cheat. The desire to pass highly has been cited by students in this study as the next major driver of malpractice. This

is in agreement with Animasahun & Ogunniran (2014) in their observation that the most important factor responsible for malpractice is the student factor culminating in the students' desire to pass at all costs. The desire to pass highly may also be a result of the students aiming at selection into institutions of higher learning (Ngungu, 2011), or as in the case of Uganda, competition for the limited government university scholarship places available. Invigilators and supervisors, school heads, parents and peers are the other players mentioned as promoters of malpractice. This supports literature cited in this study, for example Adeyemi (2010) and Ngungu (2011) who mention that poor or lack of proper supervision of students during examinations will promote malpractice. Ojerinde (2002), cited by Adegoke (2010), had earlier noted that examination malpractice was no longer a desperate students' affair, rather, school teachers and even principals were now involved in the perpetration of the vice. The Joint Council for Qualifications (2014) mentions the candidates themselves, and third parties – parents/guardians and peers as promoters of malpractice. In his study on factors influencing examination malpractice by students in Kenya, Ngungu (2011) observed that a student might cheat if he believes that others are cheating. Indeed some respondents in this study, who admitted that they would cheat if given the chance, said they would do so because some others had done so and got away with it.

The majority, 78 percent, of respondents said they would not engage in malpractice even if they had the chance. Their reasons are laudable, and include the wish to know their real level of achievement. The fear of failure as a result of receiving wrong answers or loss of results due to their being cancelled by the examination authority is also expressed, while others feared for the loss of dignity. Respondents in the study by Adegoke (2004) had expressed similar sentiments why they regretted cheating. One said 'Cheating is bad. I actually failed the test'; while another said, "It is bad for my reputation". "I feel that God

will punish you when you cheat”was yet another response. (p.61).

Hypothesis One

Respondents in this study found the measures taken by UNEB to curb examination malpractice effective. There was a significant difference between secondary school students who found the measures effective and those who did not. The null hypothesis was, therefore, rejected. Students appear to be concerned that bribery of some examination officials and poor invigilation by lazy invigilators undermine the effectiveness of some of the measures.

Research question two sought to find out what other measures the secondary school student would recommend to UNEB to curb examination malpractice. The secondary school students who responded suggested banning schools found abetting malpractice; increasing security of examinations; putting in place effective invigilation; banning candidate who take part in examination; and prison terms for offenders, among others. These recommendations are in agreement with those of Sigauke (2004), Oniye & Alawaye (2008), Adeyemi (2010); Aworanti (2012); Obidoda, Onyechi & Okere (2013) who recommend strict and thorough invigilation by dedicated supervisors, expulsion of culprits, enforcement of penalties, proper preparation of candidates and counseling, among others.

Conclusion

Respondents in this study were aware of examination malpractice. Most of them had either seen it happen or had heard of its occurrence. The students mentioned external assistance give by the teacher and invigilators, smuggling unauthorized materials into the examination room, foreknowledge of examinations, especially the practical examinations and copying from each other as the major types. The students gave inadequate preparation, their desire to pass with high grades, the fear of failure and students who do not work hard enough but

want to pass as the main reasons why they indulge in malpractice. Teachers, invigilators, head teachers and parents were mentioned as the main promoters of malpractice.

The measures taken by UNEB to control examination malpractice were assessed as effective by a significant majority of the students. The hypothesis that there was no difference between the students who found the measures effective and those who did not was rejected. The students feel that malpractice could be curbed if centres involved in malpractice were banned; security of examinations was improved; invigilation and supervision was improved; candidates involved in malpractice were banned from taking other examinations; stiff punishments, including prison terms were meted out to offenders. The students also felt that adequate preparation of the examinations would curb malpractice.

Recommendations

Based on the results of these findings, the following recommendations can be made:

1. Schools should ensure that the students are adequately prepared for the examinations so that they have confidence to take the examinations without recourse to malpractice.
2. The examination body should audit the whole process of examination administration to determine points of vulnerability that give rise to foreknowledge of examinations, especially the practical tests.
3. Measures should be put in place to improve the quality of invigilation and supervision of examinations in the examination room
4. Stiff punishment should be given to perpetrators of malpractice, including the banning of centres.
5. Counseling services in schools should be strengthened to assist students resist pressures to cheat.

6. Examination officials should be paid adequately.
7. Sustained campaigns against malpractice should be continued through all available means.

REFERENCES

- Adegoke, B. (2010). A survey of examination malpractice among secondary school students – causes, effects and solutions. Retrieved on 1st June, 2015 from www.grin.com/en/e-book/178869/a-survey-of-examination-malpractice-among-secondary-school-students.
- Adeyemi, T. O. (2010). Examination malpractices among secondary school students in Ondo State, Nigeria: Perceived causes and possible solutions. *American Euroasian Journal of Scientific Research*, 5 (1). Retrieved from [www.idosi.org/aejsr/5\(1\)10/10/pdf](http://www.idosi.org/aejsr/5(1)10/10/pdf) on 29 May, 2015.
- Akaranga, S.I., & Ongong, J.S. (2013). The phenomenon of examination malpractice: An example of Nairobi and Kenyatta universities. *Journal of Education and Practice*, 4 (18) 87 - 96
- Alutu, A. N. G., & Aluede, O. (2006). Secondary school students' perceptions of examination malpractices and examination ethics. *J Hum. & Ecol.*, 20 (4), 295 – 300 Retrieved from krepublishers.com/02-Journals/JHE/JHE on 1st June, 2015.
- Animasahun, R. A., & Ogunniran, J. O. (2015). Correlates of examination malpractice among secondary school students in Oyo State, Nigeria. *International Journal of Educational Administration on Policy Studies*, 6 (9), 181 – 189.
- Aworanti, O.A. (2012). Strategies for managing examination malpractice in public examinations www.nabtebnigeria.org/wp-content/uploads/2012/07strategies-for-managing-examination-malpractice-in-public-examinations.pdf. Retrieved on 1st June, 2015
- Chaminuka, L & Ndudzo, D. (2014). Students and staff perceptions on malpractice and fraud in higher education in Zimbabwe. *Asian Journal of Humanities and Social Sciences*, 2 (2), ISSN: 2320-9720. Retrieved on 10th June, 2015 from www.ajhss.org
- Emaikwu, S.O. (2012). Assessing the Impact of examinational malpractices on the measurement of ability in Nigeria. *International J. SocSci& Education*, 2 (4), 748 – 757. Retrieved on 3 May, 2014 from ebscohost.com/c/articles/90519353/assessing-impact-examination-malpractices-measurement.
- Jimoh, B.O. (2009). Examination malpractice in secondary schools in Nigeria: What sustains it? *European Journal of Educational Studies*, 1(3), 101- 108. Retrieved on May 5th, 2014 from ozeelacademy.com/EJESV1N3_1.pdf
- Joint Council for Qualifications (2014). Suspected malpractice in examinations and assessments. Retrieved on 22 June, 2015 from www.jcq.org.uk/exams/malpractice/suspected-malpractice-in-exams
- Joint Council for Qualifications. (2014). Warning to candidates. Retrieved on 22 June, 2015 from www.jcq.org.uk/exams-office/exam-room-posters/warning-to-candidates
- Joint Council for Qualifications. (2014). How to deal with candidate malpractice. Retrieved on 22 June, 2015 from [www.theexamsoffice.org/userfiles/files/How to deal with candidate malpractice.pdf](http://www.theexamsoffice.org/userfiles/files/How%20to%20deal%20with%20candidate%20malpractice.pdf)
- Kagete, P. M. (2008). Examination malpractices in Kenya – psychological characteristics of a

- cheat. M. Ed dissertation, Nairobi University. Retrieved on 16th June, 2015 from http://www.academia.edu/11625168/examinati_on_malpractices_in_kenya_psychological_characteristics_of_a_cheat
- Malami, U.T. (2013). Examinations malpractice, causes, effects and solutions. Paper presented at the Stakeholders Forum for Raising Integrity in the conduct of examinations in the Nigeria Education System. Retrieved on June 6th, 2014 from http://www.academia.edu/3819778/Examinati_ons_-_malpractice_-_effect_and_solutions.
- Ngungu, J. K. (2011). Factors influencing examination malpractices by students in Kenya Certificate of Secondary Education in Kitui West district, Kenya. University of Nairobi Digital Repository. http://erepository.uon.ac.ke:8080/handle/1234_56789/4183. Retrieved on 30th May, 2015
- Nworgu, B.G., Uchekwe, J.O. &Nworgu, L.N. (2013). Stakeholders' perception of the effectiveness of the strategies for curbing security breaches in public examinations in Nigeria. Paper presented at the 31st Annual Conference of the AEAA held on August 5 – 11, 2013 in Arusha: Tanzania.
- Obidoa, M. A., Oyechi, K. C., &Okere, A. U. (2013). Examination malpractice and control in public secondary schools in Anambara State, Nigeria: *Implications for the counselor. Journal of Education and Practice, 4(28), 69-73*. Retrieved on June, 2 2014 from www.iiste.org/Journals/index.php/JEP/article/view/9925
- Odongo, D.N. (2014). Effectiveness of measures by Uganda National Examinations Board to control malpractice: Perspectives of teachers and school administrators. A paper presented at the 32nd Conference of the Association for Educational Assessment in Africa, Zambezi Sun,Livingstone, Zambia 11 – 14 August, 2014
- Ogwang, S.G. (2007). The persistence of malpractice in examinations conducted by Uganda National Examinations Board at PLE, UCE, and UACE levels. UNEB: Kampala.
- Oniye, A. O. & Alawaye, A. S. (2008). Female students' perceived causes of and solution to examination malpractice in Asa local government: Implication for counseling. *Sokoto Educational Review, 10 (2),June, 2008*. Retrieved on 2nd June, 2016 from https://www.unilorin.edu.ng/publications/oniy_e/female student perceived(DR.ONIYE).htm
- Sigauke, A. T. (2004). Examinations malpractice in schools: Views from secondary school students in the Harare region: Zimbabwe. *The Zimbabwe Bulletin of Teacher Education, 13 (1), 52 – 72*.
- Soyemi, M. O. (2002). Controlling examination irregularities. The Experience of the West African Examinations Council in Nigeria. Paper presented at the 20th Annual Conference of the Association for education Assessment in Africa, held in Arusha, Tanzania. October 7-11, 2002.
- Uzochukwu, M (2015). Examination malpractice and causes. HubPages. Retrieved from uzochukwumike.hubpages.com/hub/Examinati_on-malpractice-and-causes.

Level of Effectiveness of Scout in Combating Examinations Malpractice:

The Case for Uganda National Examinations Board

Thomas Kagoro
Uganda National Examinations Board
E-mail: kagoroakiiki@gmail.com
Telephone: +256772474221

Abstract

Examination malpractice cases, which could have been avoided, continue to occur despite the deployment of permanent scouts at sitting centres. This research therefore set out to establish the level of effectiveness of permanent scouts deployed to monitor Primary Leaving Examination, with a view to finding ways of improving effectiveness in the conduct of PLE. Cross sectional survey research design was used. The participants in the study included scouts and head teachers in the participating schools, district inspectors of schools in the sampled districts and Uganda National Examination Board staff. The researcher used observation guides and questionnaires to obtain data. Responses got from questionnaires and observation were analyzed and described qualitatively with the use of content analysis and descriptive statistics. The study found out that scouts were performing well on most of the tasks save for a few weak areas; and most of the perceived important attributes of scouts were in line with what the Board recommends, although new ones were brought out. The findings will be of value to the Board and other stakeholders in improving on the effectiveness in the conduct of public examinations. It is recommended that before personnel is deployed, they should be thoroughly trained instead of just briefing them.

Key words: *Effectiveness, Permanent Scouts, Honesty, Integrity, Quality*

Introduction

During the era of the East African Examinations Council (EAEC) and the first few years of Uganda National Examinations Board (UNEB), there was a lot of trust in schools in handling examinations. Headteachers received all the examinations for the day/days of sitting and kept them in their offices until they were written and thereafter kept the scripts and awaited collection or delivery to the storage stations. Primary school teachers from neighbouring schools were engaged at no cost to supervise examinations. No malpractice cases were reported in the process of conducting the examinations.

Over time, cases of examination malpractice at Primary Leaving Examination (PLE), a high-stake examination

conducted after seven years of the primary education in Uganda, started cropping in. It then became necessary to employ teachers from neighbouring secondary and primary schools to work as supervisors (to take charge of the conduct of the examinations at the centre) and invigilators (to assist the supervisor). In addition, UNEB staff would be deployed to monitor the conduct of the examinations over a wide area in various parts of the country. This system seemed to work for some time until it became ineffective as the examination malpractice cases started increasing. Table 1 shows the number of Primary Leaving Examination (PLE) malpractice cases (2000-2009).

Table 1:
Number of PLE Examination Malpractice Cases from 2000-2009

| | | | | | | | | | | |
|--------------------------|-------|--------|------|------|------|------|------|------|------|-------|
| Year of examinations | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
| No. of malpractice cases | 6,796 | 12,065 | 63 | 76 | 24 | 574 | 926 | 290 | 616 | 1,499 |

Source: UNEB records.

In 2002, permanent scouts at PLE sitting centres were introduced owing to the high number of malpractice cases recorded in 2000 and 2001 as shown in Table 1. In this research permanent scouts are people deployed at a sitting centre to oversee the general conduct of the examinations. They will hereafter be referred to as scouts. These are normally examiners or senior educationists and citizens of high integrity, drawn from distant districts to avoid any conflict of interest. UNEB (2008) describes a scout as one with a high level of integrity, vigilance, honesty, knowledge and one very conversant with the regulations. This is a person who is able to exercise good judgment and take correct decisions on matters regarding field conduct of the examinations.

Some of the duties scouts are expected to perform are: check the examination rooms to ensure that arrangements conform to the regulations of the Board; brief the candidates on the proper conduct of examinations; monitor the checking of candidates before entry into the examination room; check the state of the question paper envelopes before they are opened; monitor the movement of candidates and their teachers before and after the papers; sign and enclose the attendance certificate in the scripts returning envelopes; and finally ensure that all the Board regulations are observed when the examination is in progress (UNEB, 2010).

The impact of scouts can be vividly seen from the statistics in Table. 1. The number of PLE malpractice cases drastically dropped from a hefty 12,065 in 2001 to a mere 63 in 2002, an improvement caused by this innovation. However, as seen in 2005, the number of

malpractice cases started increasing despite the increasing number of scouts.

The malpractice cases candidate get involved in i.e.. collusion, smuggling of information, impersonation, and external assistance – can be curbed if only the scouts effectively perform the duties they are employed to do. There were increasing complaints about the work of scouts in some aspects of the conduct of examinations which raised questions about their ability to monitor examinations. There was need to investigate whether scouts were complying with the rules and regulations and were knowledgeable about their tasks.

Statement of the problem

There is increasing unease about the effectiveness of scouts deployed to ensure the smooth conduct of PLE and yet the cost of conducting examinations is increasing. Like other stakeholders, Ogwang (2007) expressed concern about the deteriorating quality of scouting. Cases of malpractice of a nature that could have been abated in the examination rooms continue to occur despite the increasing number of examination monitors and cost to the Board. The number of malpractice cases first decreased from 12,065 in 2001 to 24 in 2004 after the introduction of scouts. However, the number of cases have kept on increasing. If this is not investigated, the numbers may continue to rise and yet the Board is devoting a lot of resources on examination monitoring. This study therefore set out to examine the level of effectiveness of scouting during the conduct of PLE and suggest ways of improving their effectiveness.

Purpose of the study

The study sought to establish the level of effectiveness of scouting in the UNEB PLE with a view to finding out areas where there is ineffectiveness so that suggestions can be made for improvement.

Objectives

The objectives of the study were to:

1. Find out the level of effectiveness of scouts in performing their roles in the management of PLE.
2. Find out the perceived desirable personal qualities of an effective scout.

Scope

The study looked at whether scouts perform their roles effectively and the perceived qualities of an effective scout. The study involved UNEB staff, headteachers and District Inspectors of Schools (DIS) handling PLE in various districts sampled from 16 sub regions of Uganda.

Significance of the Study

The findings are expected to benefit UNEB through finding out areas of weakness that need strengthening for effective PLE monitoring and stimulate more guidelines on the field conduct of examinations and recruitment of scouts. Other examination organizations may also utilize the findings to strengthen their systems of monitoring examinations. This study is likely to stimulate more research in the area of field conduct of examinations.

Conceptual Framework

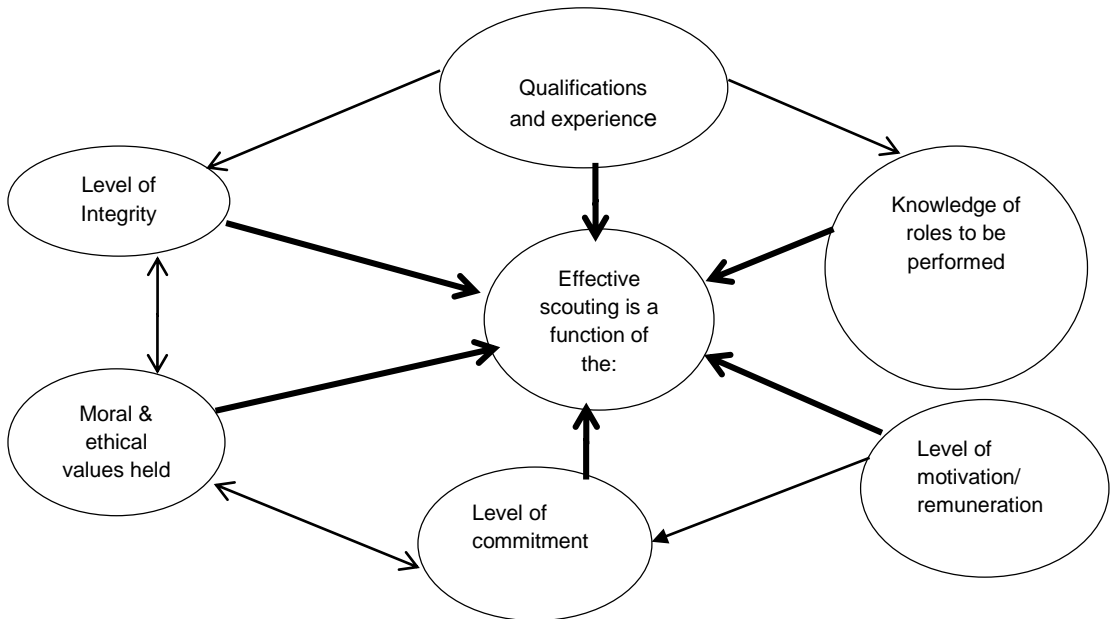


Figure 1: Factors that may influence the field conduct of PLE

Some of the factors that influence effectiveness of UNEB examination scouts are shown in Figure 1. The standard of scouting of UNEB examinations is dependent on a number of factors namely: professional qualifications, knowledge of one's work, level of motivation and commitment, integrity, and moral and

ethical values held. The model in Figure 1 presupposes that when a person is rated highly in the indicated attributes, then scouting will be done more effectively and when rated low, they are less effective. The current monitoring model for the management of the conduct of PLE is shown in Figures 2.

Contextual Framework

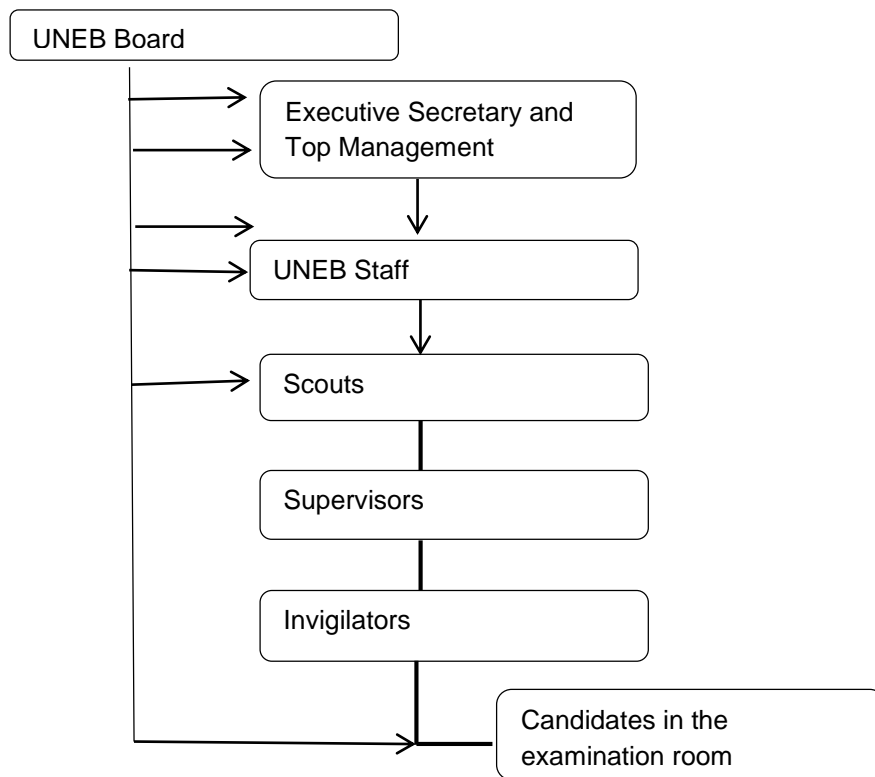


Figure 2: Current Monitoring Model of UNEB PLE

Invigilators are at the end of the monitoring hierarchy. Scouts were introduced in this chain when invigilators and supervisors had failed to control examination malpractice. They have since then more or less played the role of an overseer at the sitting centres. It is perplexing to see cases of malpractice which could be curbed by invigilators and supervisors still occurring in the examination room.

Literature Review

This chapter is intended to look at the related literature in the areas of the role and effectiveness of scouts in the administration of examinations and personal qualities of effective scouts.

Roles of scouts in the field conduct of examinations

The main duty of all officers involved in the field conduct of examinations is to ensure that candidates taking examinations abide by and conduct themselves according to the examination regulations. In order to enforce the regulations on the conduct and supervision of examinations, the officers involved must be conversant with the regulations of the Board (UNEB, 2010). This, to a great extent, is likely to determine their effectiveness in the monitoring of examinations. The University of Cambridge (1995) describes effectiveness as a situation where something produces the results that it was intended to. Similarly, Harvey (2009) and Fraser (1994) also indicate that effectiveness is the extent to which an activity fulfills its purpose or function. This research considers effectiveness as a measure of the match between stated guidelines and their achievement by the scouts.

In this research, therefore, effectiveness is considered as the ability of scouts to ensure smooth conduct of examinations by following the guidelines issued on the conduct and supervision of examinations. Success or failure of the conduct of any examination rests solely on the monitors and how effective or otherwise they are (Akanbi, 2008).

The purpose of this study is to find out whether scouts are able to fulfill the roles stipulated in the rules and regulations.

Personal qualities of an effective scout

Quality is a characteristic or feature of someone or something, or the standard of excellence of something (University of Cambridge, 1996). It is generally believed that the quality of scouts recruited to participate in the monitoring of UNEB examinations, to a great extent, influences the level of their effectiveness. Most examination bodies run examinations depending on the trust they put in the personnel they employ (Akanbi, 2008). The protection of examinations during administration will greatly depend on the personality of the individuals involved. UNEB (2010) and Malikat, Kolajo and Oyediran (2008) all agree that personnel engaged in monitoring of examinations should be of high integrity, firm, fair, honest, committed, with high moral and ethical values, smart, sharp and alert.

One of the key attributes considered very important when selecting field officers to participate in monitoring UNEB examinations is level of integrity. Kate (2008) defines integrity as doing the right thing because it is the right thing to do. People of integrity behave according to the values, beliefs and principles they hold. Choice of personnel to participate in UNEB field conduct of examinations is driven by the fact that all those recommended, and eventually appointed, are people of unquestionable integrity and personality. The question is whether these people have gone through the right environment where integrity flourishes.

This research was intended to assess the effectiveness of scouts vis a vis the existing UNEB examination regulations for the monitoring of PLE at sitting centres.

Research questions

1. How effectively do scouts perform their roles in managing UNEB examinations?

2. What are the perceived desirable personal qualities of an effective scout of UNEB examinations?

Methodology

This chapter presents the study design, sample size, sampling techniques, data collection instruments, validation of instruments, procedures for data collection and analysis.

Research design

Cross sectional survey research design was used to find out the level of effectiveness in the monitoring of UNEB examinations. The survey mainly employed the observation method. The observational survey design used was structured, disguised and natural. This design was found vital in assessing the effectiveness of scouts because they needed to be observed as they did their work using a guided observation tool (structured), without knowing that they are being observed (disguised) and in a real (natural) working environment. This personal observation approach permitted recording of actual behavior as it occurred. Since observational design is rarely used as a sole method of obtaining data (Mann 2003), it was employed in conjunction with cross-sectional survey. The cross sectional survey design was also used because it was able to generate a sizable amount of data from a cross-section of the target group –scouts, headteachers, UNEB staff and district education officials. It helped in generating data on perceptions about desirable personal qualities of an effective scout, which could not be obtained through observation.

Both qualitative and quantitative methods of data collection and analysis were used. The qualitative

methods were hoped to bring out a holistic inquiry through observation guides, interviews and questionnaires. The purpose of qualitative research is to give a deep understanding of the way things are and also why they are the way they are (Amin, 2005). On the other hand, the quantitative methods were employed in order to

establish the level of effectiveness of scouts from the recorded observations.

Sample size

This study covered the whole country, taking into account the four traditional regions (Central, East, North and West) and Kampala city. These regions were sub-divided into 16 zones (UNEB, 2007). This gave a total of 17 districts, (including Kampala). From the selected districts, four primary school sitting centres were used. This made a total of 68 centres. The study sample included 17 inspectors of schools, 68 heads of centre, 68 scouts from the sampled districts and 44 UNEB staff members.

In addition, reports from scouts, district monitors and district Inspectors of Schools were studied. This helped to scrutinize the performance of scouts using the information given in the reports.

Sampling techniques

Multi-stage, stratified and purposive sampling techniques were used in the research. Multi-stage random sampling was used first to select one district from each of the 16 zones in order to have a fair representation of the population across the country.

Four schools from each of the main storage stations in the sampled 16 districts were then chosen considering the urban-rural location factor. One urban and three rural sitting centres were chosen using a stratified random sampling technique to ensure a fair representation between the rural and urban schools since there are more rural sitting centres than urban. Kampala district was purposively included in the study because it is a unique urban district from which four primary sitting centres were chosen. The scouts and heads of centre from the

sampled schools, and inspectors of schools from the sampled districts were automatically included in the study. The UNEB staff were also randomly selected to tap their field experiences in the conduct of PLE.

Research instruments

The items included in the observation guides, questionnaires and the interview guide to assess the effectiveness of scouts were drawn from the duties and responsibilities of scouts as stipulated in UNEB Rules and Regulations on the conduct of examinations.

The observation guides were used as a primary data collection tool to answer research question one about whether scouts do their work effectively during the progress of PLE. The guides were necessary because they enabled the research assistants to obtain data in a standardized manner. Actual observation of what happens in the field was seen to come in handy to capture real life activities of the scouts. The observation guides had both closed and open-ended-items.

The researcher used both close and open-ended questionnaires to obtain data from scouts. The questionnaire was used to find out whether scouts effectively perform their role in managing examinations and the perceived qualities of an effective scout.

District Inspectors of Schools (DISs) were interviewed to give their overall assessment of performance of scouts in the field. Interview guides were appropriate here because they are flexible in their use and fit for in-depth discussion with a small number of District Inspectors of Schools. A check list was also used to obtain specific information from the reports got from the field.

Validation of the instruments

Observation guides were the prime instruments for data collection. These were constructed to observe the work of scouts based on the activities expected of them as stipulated in the Regulations on the Conduct and Supervision of PLE. Since this is a standard document of

the Board with bye-laws for the smooth field conduct of examinations, the researcher deemed the items included in the guide valid and reliable. A validation of the open-ended questionnaires for scouts and UNEB staff was

carried out before administering them. They were subjected to close scrutiny, discussions and criticism from UNEB Research Officers and some members of the UNEB Primary Department. Those items that were deemed irrelevant were dropped.

Data collection procedure

A letter of authority was obtained from the Executive Secretary to allow data collectors to access examination premises during the progress of the examinations

A total of 36 data collectors (research assistants) was engaged. They comprised examiners who have ever worked as scouts and had an experience in research work, at least at their first degree. Before the appointed officers were dispatched to the various districts, the researcher took them through a one-day training on how to collect data from the field. In addition, a guide was prepared for them on how to do their assignment. On the first day of the examination, the research assistants went to the sitting centres, where they observed all activities performed by the scouts before, during and after the paper. Finally, questionnaires were administered to scouts, headteachers and UNEB staff. Data collected from the field and reports were then assembled for analysis.

Data Analysis

Responses got from questionnaires, interviews and observations were mainly analyzed and described qualitatively, save for the close-ended items which were quantitatively analyzed. Basic descriptive statistics were used to report responses to the items within the questionnaires and observation guides which were close-ended. Tables and percentages were utilized to present data.

Content analysis of the open-ended items in the questionnaire, interview recordings and observations was done (Harwood & Garry 2003). The process of content

analysis included open coding, creating categories and developing themes. During the coding process, information was classified into themes and issues and

then data was systematically arranged and presented (Burn 1997).

A further analysis was done of the observations made about the work of the scouts to establish their levels of effectiveness. Occurrence (Yes) of an observed activity during the examination was scored one while a 'no' occurrence was a zero. A total score was obtained for each scout observed.

Presentation, Analysis and Interpretation of Findings

Introduction

In this chapter, the response rate, and the quantitative and qualitative findings of the study investigating the level of effectiveness of scouts in monitoring of UNEB examinations are presented. Presentation, analysis and interpretation is done per research question.

Research Question one: How effectively do scouts perform their roles in managing UNEB examinations?

The quantitative data from the observation guides and questionnaires are reported using frequency tables and percentages.

The observations on the work of scouts during the conduct of PLE are presented in Table 2.

Table 2: Observations on the effectiveness of Scouts at PLE Sitting Centres

| Duty | Yes | | No | |
|--|-------|-------------|-------|-------------|
| | Freq. | %age | Freq. | %age |
| Did he/she possess an identity card? | 65 | 95.6 | 3 | 4.4 |
| Did he/she present an appointment letter from the Executive Secretary of UNEB? | 68 | 100 | 0 | 0.0 |
| Did he/she carry albums for the school(s) at the sitting centre and use them? | 62 | 91.2 | 6 | 8.8 |
| Did he/she check the examination rooms/halls to ensure that sitting arrangements conform to the regulations of the Board? | 54 | 80.6 | 13 | 19.4 |
| Did he/she sign and enclose the attendance certificate in the scripts returning envelope of each paper? | 59 | 86.8 | 9 | 13.2 |
| Did he/she brief the candidates on the proper conduct of examinations? | 41 | 60.3 | 27 | 39.7 |
| Did he/she monitor the checking of candidates before they enter the examination room? | 55 | 80.9 | 13 | 19.1 |
| Did the scout ensure that candidates are mixed? | 51 | 87.9 | 7 | 12.1 |
| Did he/she check the state of the question paper envelopes before they are opened? | 62 | 91.2 | 6 | 8.8 |
| Did he/she check on the left-over question papers to ensure that they have been safely locked away? | 49 | 73.1 | 18 | 26.9 |
| Did he/she ensure that the duration for the paper as stipulated on the timetable and question paper is observed? | 66 | 97.1 | 2 | 2.9 |
| Did he/she ensure that sitting plans are drawn and enclosed by the Supervisor? | 57 | 96.6 | 2 | 3.4 |
| Did he/she check to ensure that all the information on the returning envelope is accurately filled? | 53 | 84.1 | 10 | 15.9 |
| Did h/she witness the sealing of candidates' scripts? | 60 | 96.8 | 2 | 3.2 |
| Did he/she monitor the movement of candidates and their teachers before and after the first paper of the day (lunch time)? | 42 | 67.7 | 20 | 32.3 |
| Did he/she ensure that the attendance registers are packed and enclosed in the Science Script returning envelopes? | 50 | 82.0 | 11 | 18.0 |
| Average Score | | 85.8 | | 14.2 |

All the scouts (100%) presented appointment letters to the observers. High compliance levels of over 90% were recorded in the use of albums (91.2%), checking the state of the question paper envelope before they are opened (91.2%), ensuring that the duration for the paper is observed (97.1%), ensuring that sitting plans are drawn and enclosed by a supervisor (96.6%) and witnessing the sealing of candidates' scripts (96.8%). However, even in these areas where scouts seemed to be doing very well, there were issues of mal-practice noted. For instance there was a case where candidates were required to check the state of the question paper envelope, it was observed that candidates who witnessed the opening signed at the end of the examination and were the same who signed on the return envelope. In yet other two instances, one of the question paper envelopes was opened from outside the examination room, and in another, the scout was busy outside chatting with the headteacher as the examination was in progress. It was also noted that some invigilators drew imaginary and inaccurate sitting plans using identity cards, while sealing of candidates' scripts, at one centre, the sending was done only in the presence of the three witnesses when the rest of the candidates had been released.

There were a number of areas where scouts' level of performance was observed to be between 80% and 90%. This included checking of examination rooms to ensure that the sitting arrangements conform to the UNEB regulations; the signing and enclosing of the attendance certificate in the scripts returning envelope of each paper; monitoring the checking of candidates before they entered the examination room; ensuring that candidates were mixed; ensuring that all the information on the returning envelope was accurately filled; and ensuring that the attendance registers were packed and enclosed in the Science script returning envelopes.

Specific observations indicated that even where checking was seen not to be effective, some of the scouts would just look on or were seated at a distance. One of the

scouts was observed to be too old to play any active role in the monitoring of the examination. In another centre, the headteacher took over the work of seating candidates according to his design. In another incident, a scout was not around at the sealing of the scripts in the last paper because he had left by lunch time to go and meet the district monitor in his/her the home district. Even where scouts seem to be doing well, there are isolated cases which may have far reaching implications.

There are areas where the scouts were clearly not doing well. For instance only 60.3% of the scouts were able to brief candidates on the proper conduct of examinations and 39.7% did not. In a number of sitting centres it was either the headteacher who did it or the supervisor. The second aspect which was not handled well was the checking of the leftover question papers to ensure that they have been safely locked away, where 73.1% was observed to have done it while 29.9% did not. In some cases invigilators were seen to have remained with the question papers and yet scouts did not reprimand them, let alone endeavor to find out how many had remained. In another instance, the leftover question papers were kept in a locker which was in the staffroom. The third one was where 67.7% were observed to monitor the movement of candidates and their teachers before and after the first paper of the day while 32.3% did not seem to care. In centres where scouts did not care, some candidates rode their bicycles away and in another instance the candidates from a neighbouring school were left to go back for lunch.

General assessment of the scouts' effectiveness by the observers indicates that whereas a good number of the scouts (85.8%) were able to follow the rules and regulations for the conduct of PLE, there were a few who fell below the mark. In centres where there were many scouts, they tended to engage in conversations rather than do the monitoring as required. Some of them lacked vigilance and commitment to their work to the extent that one sitting centre did not have a scout completely while a

few would leave before the end of the exercise. One scout was observed not to have entered the examination room at all during one of the papers. In the words of one of the observers making a general assessment of a scout she said,

“She was not effective at all since she did not do any of her work. She mistook me for a fellow scout and possibly thought I was to act instead of her. She spent almost all the time outside the examination room chatting with the headteacher, who instead would sometimes play the role of a scout when the situation would demand”.

The observation of the duties of scouts indicates that on average 85.8% was performed as required by the Board and 14.2% was not. Although 85.8% seems to be a high level of effectiveness, 14.2% lack of compliance is big enough to negatively affect the conduct of PLE.

Research question Two: What are the perceived desirable personal qualities of an effective scout of UNEB Primary Leaving Examinations?

It is generally believed that personnel to monitor UNEB examinations need to be of certain qualities as spelt out in the Regulations on the Conduct and Supervision of examinations booklets. This study therefore sought to find out what some of the stakeholders perceive to be the desirable qualities of an effective invigilator, supervisor and scout. Table 3 shows what DISs perceive to be the important attributes of an effective scout.

Table 3: Important attributes of an effective scout as perceived by selected UNEB Staff

| Attribute | Frequency | %age |
|----------------------------------|-----------|------|
| Professional teacher | 12 | 30 |
| Honesty | 8 | 20 |
| Vigilance | 4 | 10 |
| Examiner | 4 | 10 |
| Years of experience as a teacher | 4 | 10 |
| Firmness | 4 | 10 |
| Commitment | 4 | 10 |
| Above 30 years of age | 2 | 5 |
| Secondary school teacher | 2 | 5 |

Thirty percent of UNEB staff who participated in the study considered being a professional teacher as the most important attribute of an effective scout. The other attributes perceived to be important are honesty (20%), vigilance (10%), being an examiner (10%), years of experience as a scout (10%) and firmness (10%).

Scouts were also asked to give the most important attributes of effective officers who are deployed as scouts.

Table 4 gives the details of the qualities scouts think effective officers deployed should possess.

Thirty percent of UNEB staff who participated in the study considered being a professional teacher as the most important attribute of an effective scout. The other attributes perceived to be important are honesty (20%), vigilance (10%), being an examiner (10%), years of experience as a scout (10%) and firmness (10%).

Scouts were also asked to give the most important attributes of effective officers who are deployed as scouts.

Table 4 gives the details of the qualities scouts think effective officers deployed should possess.

Table 4: Important attributes of an effective scout as perceived by scouts.

| SCOUTS | | |
|----------------------|------|------|
| Attributes | Freq | %age |
| Professional teacher | 29 | 42.6 |
| Honesty | 29 | 42.6 |
| Commitment | 24 | 35.3 |
| Knowledgeable | 5 | 7.4 |
| Integrity | 20 | 29.4 |
| Vigilance | 11 | 16.2 |
| Years of experience | 15 | 22.1 |
| Time consciousness | 2 | 2.9 |
| Firmness | 3 | 4.4 |

Scouts considered the attributes of being a professional teacher and an honest person as the most important for an effective scout. As many as 42.6% of the scouts who filled the questionnaire, showed that being a professional teacher and an honest person are important attributes influencing ones scouting effectiveness.

Discussion, Conclusions and Recommendations

Introduction

This chapter discusses the findings of the study in relation to the reviewed literature. It presents the conclusions arising from the findings. It finally gives the recommendations in relation to the conclusions drawn from the study.

The discussion of the findings is presented per research question.

Research Question One: How effectively do scouts perform their roles in managing UNEB examinations?

The major method used to assess the level of effectiveness of scouts in performing their roles was observational survey, (Harvey, 2000 & Fraser, 1994). The main tool used was an observation guide.

Scouts showed high performance levels in the use of albums, checking the state of the question paper

envelopes before they are opened, ensuring that the duration for the paper is observed, ensuring that sitting plans are drawn and enclosed by supervisors and witnessing the sealing of candidates' scripts. This high performance was despite the few weaknesses observed in those areas. It is, however, disappointing to note that in some cases invigilators and supervisors were not performing as required of them according to UNEB guidelines and yet scouts could not do their role of ensuring that things are done the right way.

Observations indicated that scouts were not performing well in the following areas: briefing candidates on the proper conduct of examinations, checking on leftover question papers and monitoring of candidates' movements during lunch breaks. These observations seem to shed light on the fact that they were not in control in some instances. This is probably why some of them left their duties to be performed by other people. For instance, some scouts left briefing of candidates to headteachers or supervisors. Some of them did not exude enough confidence in what they were doing, indicating a lack of authority. This is why in some instances the headteachers took over the process of the conduct of examinations and yet primary headteachers are not allowed in the examination rooms. This raises a question

whether scouts were in control of the examination process. The headteachers in this case were in control of the management of the conduct of the examinations instead of the scouts. This leaves an open chance for the headteacher to do what he/she wants with the examinations.

Research Question Two: What are the perceived desirable personal qualities of an effective scout of UNEB Primary Leaving Examinations?

The research explored to see what stakeholders perceive to be the most desirable personal qualities of an effective scout. It was revealed that the most important attribute is being a professional teacher. The other highly rated attributes are integrity, honesty and commitment. This is in agreement with the laid down regulations (UNEB, 2010), and Malikat, Kolajo & Oyediran (2008), who also agree that personnel engaged in monitoring of examinations should be of unquestionable character.

These findings underpin the fact that all those involved in the conduct of PLE have to be professionally trained teachers, honest and of high integrity. Integrity as an important attribute is supported by Anthony (2008), although he admits that this is a very difficult subject. The only challenge, whereas it is easier to identify a professionally trained teacher, it may not be easy to know who an honest person or a person of integrity is. The Board has to rely on the recommendation of those who know the officers deployed to monitor PLE. This requires that those involved in recommending and recruitment of the monitoring staff should be conversant with the requirements for recruitment.

The study also brought out other important attributes not considered in the UNEB (2010) guidelines. For instance, years of experience as a teacher were considered important. They opined that one should have a reasonable experience in teaching before he/she is deployed to monitor PLE, but should not be too old to cope with the rigours of the monitoring exercise. Secondary school teachers were also preferred by UNEB staff to work as scouts.

In a nutshell, the most highly rated attributes of scouts that influence their level of effectiveness are being a professionally trained teacher and an honest person, one's level of commitment and integrity, and years of experience in the task deployed to handle. The study came up with some attributes which are not normally emphasized by the Board like: knowledge level in the task assigned, years of experience as a teacher, years of experience in the task deployed to handle, being a secondary school teacher, time consciousness, confidence and years of maturity.

Conclusions

The main problem in this study was reoccurrence of malpractice cases of a nature that could have been avoided in the examination room after the introduction of scouts. Therefore, it examined the level of effectiveness of scouts during the conduct of UNEB PLE. Following the findings of the study, the researcher made the following conclusions:

1. Scouts showed high performance levels in the use of albums, checking the state of the question paper envelopes before they are opened, ensuring that the duration for the paper is observed, ensuring that sitting plans are drawn and enclosed by supervisors and witnessing the sealing of candidates' scripts. This high performance was despite the few weaknesses observed in those areas.
2. Scouts had low performance levels in the following areas: briefing of candidates on the proper conduct of examinations, the checking on the leftover question papers to ensure that they have been safely locked away, and monitoring the movement of candidates and their teachers before and after the first paper of the day.
3. The most highly rated attributes of scouts are being a professionally trained teacher, being an honest person, level of commitment, integrity

Amin,
and years of experience in the task deployed to
handle.

Recommendations

1. The scouts should be thoroughly trained before deployment. Particular attention should be paid to those areas of weakness identified.
2. Recruitment criteria should include; one being a professionally trained teacher, being an honest person, high level of commitment, integrity and years of experience in the task deployed to handle.
3. Other examination bodies may consider adopting this monitoring model.
4. More research should be carried out to investigate effectiveness of the other field officers.

References

ACCEL – Team (2008) Employee Motivation: Theory and Practice.
<http://www.accel-team.com/motivation/> Retrieved on 14/10/2009.

Akanbi, A. O. (nd). Towards a Crisis Free Examination Management in Nigerian Colleges of Education. Kwara State College of Education: Ilorin. Retrieved on 01st October 2009 from http://www.naere.org/journal/TOWARDS_CRISIS_FREE_EXAMINATION_MANAGEMENT_NIGERIAN_COLLEGES_EDUCATION.pdf.

Argborbechem, P, T., Akoko, M, A., Kagoro, T, and Nantumbwe, J, S. (2009).

Protection in Test Item Production, Banking and selection in Examination Bodies Across Africa. Paper presented at the 27th Annual Conference of the Association for Educational Assessment in Africa in Yaounde-Cameroon, August 24-28, 2009.

M, E. (2005) Social Science Research: Conception, Methodology and Analysis. Makerere University Press: Kampala

Anonymous. (n.d) Nine Theories of Ethics that Rule the World. Retrieved from <http://jan.ucc.nau.edu/~dgs2/theories.pdf>

Bennett, AE & Ritchie (1975) Questionnaires in Medicine: A Guide to Their Design and Use, Oxford University Press: London.

Best, J, W & Kahn, J, V. (1993). Research in Education 7th Edith, Allyn and Gacon, Boston.

Burn, R, P. (1997). Introduction of Research Methods, 3rd edition, Longman, Canberra.

Fraser, M. (1994). Quality in higher education: an international perspective. In Green, D. (Ed), 1994, What is Quality in Higher Education? Pp. 101-111 (Buckingham, Open University Press & Society for Research into Higher Education.

Harwood, T, G. & Garry, T. (2003). An Overview of Content Analysis. The marking Review, Vol. 3, No. 4, pp. 479-487.

Harvey, L. (2009). Analytic Quality Glossary, Quality Research International. <http://www.qualityresearchinternational.com/glossary/>

Hornby, A, S. (2000). Oxford Advanced Learner's Dictionary of Current English. Oxford University Press: New York.

Kagoro, T. (2008). Assessment Considerations for Special Needs Candidates at Primary Leaving Examinations, Uganda Certificate of Education Examinations UNEB: Kampala.

- Mann, C. J. (2003). Observational Research Methods. Research Design II; Cohort, Cross-sectional, and Case-control Studies. Emergency Medicine Journal. 20 (1) 54-60.
- MBA Here I come! (Saturday 19th May 2007). Survey and Observation: Business Administration and Marketing class notes. Destinydawnmarie.blogspot.com/2007/.../survey-and-observation.ht...
- NIST/SEMATECH (6/23/2010). E-Handbook of Statistical Methods <http://www.itl.nist.gov/div898/handbook/>,
- Ogwang, S. G. (2007). The Persistence of Malpractice in Examinations Conducted by Uganda National Examinations Board at PLE, UCE and UACE Levels. UNEB: Kampala.
- Polit, F, D & Beck, C, T. (2008). Nursing Research: Generating and Assessing Evidence for Nursing Practice. 8th Edition, Lippincott Williams & Wilkins, Philadelphia.
- UNEB. (2007). The Achievements of Primary School Pupils in Uganda in Literacy in English and Numeracy, NAPE Report Kampala.
- UNEB. (2008). Regulations on the Conduct and Supervision of Primary Leaving Examination. UNEB: Kampala.
- University of Cambridge. (1996). Cambridge International Dictionary of English. Cambridge University Press: London.

SUB THEME C: CURRICULUM REFORM AND EDUCATIONAL ASSESSMENT

Home based and school factors that influence pupil academic achievement at Class 3 level in Kenya.

Doreen Kawira Mbwiri

dkawira@knec.ac.ke

The Kenya National Examinations Council, Kenya

Abstract

One key deliverable of the social pillar within Kenya's Vision 2030 is the provision of quality education; for it to be attuned national assessments should be undertaken. This paper presents finding of a study (national assessment) conducted in 2016 to determine home based and school factors influencing learner achievement among class 3 pupils in Kenya. The purpose of the study was to identify home based and school factors that influence pupil achievement in order to enable parent's teachers and school administrators cater for these factors to improve learner achievement and the overall quality of education. The study covered all the 47 counties of the Republic of Kenya, 250 schools across the country were sampled using the SAMDEM sampling software. 5, 522 pupils who were randomly sampled participated in the study. The descriptive survey design was used for the study. Achievement tests and questionnaires were used for data collection. The data captured using CSPro was thereafter transferred to STATA and SPSS for data cleaning, analysis and tabulation. The findings indicated that Orphan hood had influence on learner achievement. Pupils whose parents had died performed worse than those whose parents were alive. The socio-economic status of pupils also had implication on the performance. Pupils who had possessions that enhanced learning such as radio, newspapers and computers in their families performed better than those who did not have these possessions. Parental involvement in their children academic work was found to have great significance in performance where pupils who were assisted in their homework performed well than those who did homework on their own. On school based factors, pupil textbook ratio had great significance on achievement. Pupils who shared textbooks with others performed worse than those who had their own text books. School libraries are very important and from the study only a small percentage indicated that they had libraries. Guidance and counseling in schools is very vital as pupils who were free to talk to their teachers performed better than those who were not free to talk to their teachers. In conclusion there are home based and school factors that had influence on pupil achievement and it's the role of parents, teachers and school administration to work together in ensuring these factor are catered for so as not to have negative effect on pupils achievement.

1.0 Background

Assessments are important drivers to learning as they portray learners' attitudes towards their studies. Assessment is the process of gathering and discussing information from multiple and diverse sources in order to develop a deep understanding of what learners know, understand and can do with their knowledge as a result of their educational experiences.

There are many types of assessment which include: diagnostic assessment often undertaken at the beginning of a unit of study to assess the skills, abilities, interests, experiences, and levels of achievement. Informal assessment can be referred to as the systematic monitoring of learners during class learning. Formal assessment is the use of specific assessment strategies to determine the degree to which students have achieved the learning outcomes. Summative assessments are used to evaluate learners learning, skill acquisition, and academic achievement at the conclusion of a defined instructional period. It's conducted to, promote learners, and certification. Summative assessment is also referred to assessment of learning.

Formative assessment refers to frequent, interactive assessments to measure learners' progress. It's conducted to identify learning needs so as to adjust teaching appropriately. (OECD/CERI. 2008). It is also referred to as assessment for learning as it happens during learning. It involves learners understanding exactly what they are required to learn, what is expected of them and are given feedback and advice on how to improve their work.

Assessment for learning (AFL) is an approach to teaching and learning that creates feedback which is then used to improve learners' performance. Learners become more involved in the learning process and from this gain confidence in what they are expected to learn and to what standard. This paper focuses on assessment for learning

as it looks at home based and school factors that influence learner achievement.

The paper will discuss on home based and school factors that influence learner achievement. On home based factors it will focus on orphan hood, socio-economic status at home and parental involvement in their children school work as the key factors that influence learner achievement. On school based factors, the paper will focus on availability of textbooks, school library and provision of guidance and counseling service as the major factors that influence learner achievement.

1.1 Statement of the Problem

The overall aim of education in Kenya is to produce members in the society who have adequate skills, knowledge and attitude that can be used to contribute positively to a productive society.

To achieve this aim and improve pupils' performance at class 3 level, the learning process must be guided and controlled; and the learning environment must be conducive and supportive. The parents, teachers and the school administration play a key role in ensuring a learner gets conducive environment both at home and school.

According to Maicibi (2005) a good environment should be provided at home if children must learn, if the administration must be successful and if the school must develop. Maani (2009) observes that students' success and passing in their examinations at school is closely related to their home backgrounds. Maku (2005) observes that home based factors such as the size of the household, income of parents, education level of parents and type and level of interaction between parents and children with regard to school work affect academic performance.

Learning involves interaction of learners with the environment. Teaching and learning resources include classrooms, laboratories, libraries, playing fields, textbooks among others. Indeed physical resources go a long way in creating conducive environment that promote

effective teaching and learning. Physical materials in terms of adequacy and quality have been noted to have a great impact on performance of students in the examination. A school that has adequate instructional materials is likely to post better quality grades than a school which has poor quality physical resources

Therefore, this study sought to explore the home based and school factors that influence learner achievement.

1.2 Objectives of the Study

- Determine the home based factors that influence learner academic achievement at class 3 level;
- Determine the school based factors that influence learner academic achievement at class 3 level;

1.3 Purpose of the Study

This study aimed at finding out the home-based and school factors that influence the pupil achievement at Class 3 level.

1.4 DELIMITATION OF THE STUDY

The study was limited to the Class 3 pupils attending registered mainstream public and private primary schools

excluding the pupils in special needs schools it also excluded all schools that had class sizes of less than fifteen pupils. The special schools were excluded because they used an adapted syllabus. As for the class size of less than fifteen, the study largely used a class size of 25 and above.

1.5 Operation Definition of Terms

1.5.1 **Learner/pupil Achievement:** It describes what the learner knows and will be able to do by the end of a course or programme.

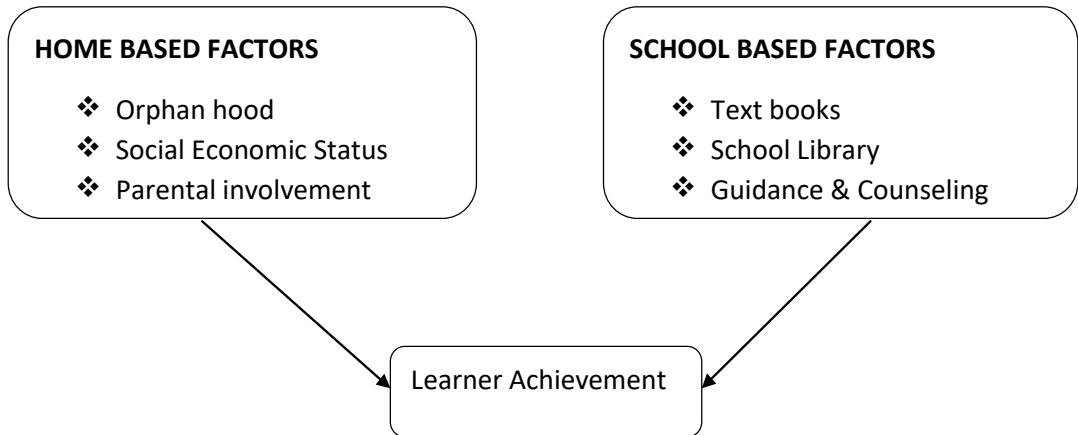
1.5.2 **Home based factors** refer to conditions inherent to the home environment that either limits or affects the academic performance of pupils.

1.5.3 **School based factors** refer to physical, human and financial resources that either limits or affects the academic performance of pupils.

1.5.4 **Socio-economic status** refers to the income, parental education levels, and parental occupation based on mother's education, father's education, mother's occupation, father's occupation and combined income.

1.6 Conceptual Framework

The figure below shows how home based and school factors influence learner achievement.



2.0 Literature Review

2.1 Learner Achievement

Learner achievement measures the amount of academic content a student learns in a determined amount of time. Each grade level has learning goals or instructional standards that educators are required to teach. Learner achievement is about how much the learners are learning at school, whether it's enough and if there are gaps in their learning that needs to be addressed.

Indicators of learner achievement are specific, observable and measurable characteristics that can be used to show progress towards achieving an objective. According to Abagi and Odipo (2007) the indicators of academic performance in education include good examination results, positive exhibition of knowledge, skills and attitudes.

There are several factors that affect performance in schools such as home based and school based

2.2 Home Based Factors that influence Learner Achievement

2.2.1 Orphan hood

An orphan is a child who has lost either one or both parents according to UNICEF (2015) and according to Rubaha (2008) neglect, abuse, lack of parental love, lack of food, high labor demand at home, stigmatized at school because of lack of school uniforms and learning materials to mention but a few are some of the challenges facing orphaned students. Kafwa (2005) argued that compared to non-orphans, orphaned students are more likely to be malnourished, more likely to drop out of school and have limited access to health social services as well as being prone to exploitation. These challenges

create emotional stress making the pupils difficult to concentrate and learn in the classroom due to trauma. Hewlett [2006] argued that motivation is the drive of learning and in the case of orphans, since they lack basic needs such as love from parents, good shelter, food and

clothes then they will be less motivated to perform well in class.

2.2.2 Social Economic Status (SES)

Parental income is the key indicator of Social Economic Status (SES) According to Becker & Tomes if those children from [poor family background and those from the rich family) have the same endowments and their parents have the same value and norms, the former are less likely than the later to succeed because the poor parents cannot afford to buy things that their children need such as food and medical care, and things that could help their children to get ahead, such as computers, music lessons and trips to interesting and educational places. Children whose basic materials need are not met have hard times acquiring the skills that help them succeed, and children whose parents cannot buy them the “extras” are at a competitive disadvantaged.

Parents are often faced with unique challenges that hinder them from meeting the learners’ needs. These include; insufficient time, job type, home rules, level of education, order of priority, set home environment, opinion to voluntary work at school, time taken to respond to school activities, buying instructional materials, attending parents meetings, conferences, sports, academic clinic day, and discussing the academic progress of the child. If the above needs are not attended to, the child may not perform well because of inadequate parental support.

2.3 School Based Factors that influence Learner Achievement

2.3.1 Textbooks

According to yala (2010), textbooks provide the only source of information for students as well as the course of studies for the subjects Squire (1991) writing on teachers reliance on textbooks stated that those seeking to improve the quality of education in instructional materials would inevitably lead to changes in actual teaching. While the selection of a textbook has been judged to be of vital importance to academic achievement, it is sad to say that relevant books are not available for teaching and learning activities.

2.3.2 Libraries

A study by Akinsanya (2010), revealed that physical facilities like libraries were inadequate which affected students’ performance. Similarly, Oni (1995) said that availability and the quality of materials facilitates enhance smooth operation of any school and thereby enhancing effective teaching/learning activity and when this is so, there is higher educational attainment by students.

2.3.3 Guidance and Counseling

Counseling is an activity that utilizes interpersonal relationships to enable people to develop self understanding and to make changes in their lives. According to a study by Aalieh&Mohammadyari (2011). Counseled students showed a superior retention as compared to their peers it’s also proved that it had a positive effect on the children research has proved that learners who attended the counseling showed improvement in their grades as compared to those who did not.

3.0 Research Methodology

The study adopted a descriptive survey design. This design allows for data collection from a wide geographical coverage. In addition, the survey design entails the study of situations in their natural setting without manipulation of variables. A representative sample was used as it was not possible to collect data from all schools and teachers in the whole country.

3.1 Target Population

The desired target population for this study was all the Class 3 pupils of 2015 in Kenya. The defined target population was all the Class 3 pupils of 2015 attending registered mainstream public and private primary schools.

3.2 Sampling

The study involved all the 47 counties in the Republic of Kenya. The sample of schools for study was generated using the Sample Design Manager (SAMDEM). SAMDEM is a specialized software system that enables generation of a range of sampling options (Syllaet *et al*, 2003) From the software 250 schools were sampled for the study.

Sampling of students at school level was random sampling using a scientific random table. 25 Class 3 pupils from each sampled school participated in the study. A total of 5, 522 students participated in the study.

3.3 Data collection instruments

This study used the following instruments:

- a) Achievement Tests: Numeracy, Literacy (English and Kiswahili);
- b) Pupil questionnaire.

The study also utilized several administrative documents to facilitate effective sampling and conducting of fieldwork. These included:

- a) Pupil name form;
- b) Data collectors manual;
- c) Scientific random sampling table.

3.3.1 Achievement tests

English and Kiswahili: The tests had sampled open ended questions across all the topics in the syllabus.

Mathematics: A team of Mathematics specialists developed a test which covered all the competency levels at class 3.

These tests were piloted and revised to ensure all the questions were good and also for reliability and

validity. All the tests were administrated to the sampled pupils at class 3.

3.3.2 Questionnaire for Pupils

Pupils: Questionnaire explored the pupil's personal characteristics, home based factors, and school based factors and attitudes to schooling.

3.4 Data Collection

Data collection was conducted for two days in the sampled schools. On the first day sampling of students using the class register was done using the random sample table, administration of Mathematics and English test was also carried out to the sampled pupils in two sessions. On the second day, administration of Kiswahili test and pupil questionnaire was done.

3.5 Data Entry Cleaning and Analysis

Data entry and cleaning for both qualitative and quantitative data was done using Census and Survey Processing System (CSPRO) version 6.3. This program enabled accurate data capture. double entry, conversion from and to different programs, and also performs tabulations.

The data from CSPRO were thereafter transferred to STATA (version 14) and later SPSS (version 22) for, analysis and tabulation. Descriptive analysis was done to report specific variables such as pupil characteristics. In addition, a comparison of achievement means among different groups such as pupils from different socio-economic backgrounds and counties was carried out. To determine factors contributing most to learning achievement, a multilevel analysis of the data was carried out using STATA 14 software. This software allows for simultaneous examination of the effects of variables at different levels of the hierarchy.

4.0 PRESENTATION AND DISCUSSION OF FINDINGS

Table 4.1 Results for Multi Level Analysis

The table presents results of the multi level analysis in the three subjects. It shows how the variables both from home and school were significant to learner performance.

| Variable | Mathematics | | | | English | | | | Kiswahili | | | |
|--|-------------|----------|--------|------|---------|----------|--------|------|-----------|----------|--------|------|
| | Coef | <i>P</i> | 95% CI | | Coef | <i>p</i> | 95% CI | | Coef | <i>p</i> | 95% CI | |
| Orphanhood status | | | | | | | | | | | | |
| Orphaned | -4.5 | | -9.9 | 1 | -8.1 | ** | -12.9 | -3.3 | -5 | | -10.8 | 0.9 |
| Average item possession aggregated at school level | 7.1 | ** | 2.1 | 12.1 | 6.9 | ** | 2.4 | 11.5 | 11 | ** | 6.6 | 15.5 |
| Pupil has textbook on subject | 5.1 | | -0.2 | 10.5 | -0.5 | | -5.3 | 4.3 | 3.8 | | -1.8 | 9.4 |
| Homework (parental involvement) | 10.2 | * | 1 | 19.4 | -0.6 | | -8.2 | 6.9 | 4.4 | | -3.9 | 12.8 |
| Pupil talks to teacher to solve a problem (Guidance and counselling) | 6.8 | | -1.3 | 14.9 | 8 | * | 0.8 | 15.1 | 9.5 | * | 1 | 18 |

Key : * Significant at 0.05 level
 ** Significant at 0.01 level
 ZP*=Pupil variables

The table gives results from the linear regression analysis which is a predictive analysis used to explain the relationship between one continuous dependent variable (in this study, pupil achievement) and two independent variables (in this study, home and school based characteristics). The pupil achievement was from the achievement tests while home and school characteristics were from the questionnaire.

4.1 Home Based Factors

From the above table, orphan hood was reported to influence learner achievement. Orphans scored lower in Mathematics, English and Kiswahili by 4.5, 8.1, and 5.0 respectively than those who were not orphaned.

Socio-economic status was noted to have an influence on learner achievement from the findings. Pupils from homes with a higher socio-economic status scored better in Mathematics, English and Kiswahili by 7.1, 6.9 and 11.0 respectively than those from homes with a lower socioeconomic status. This variable was very significant in all the three subjects.

Giving homework and parental assistance in homework was also observed to be casually related to improved learner achievement. In Mathematics, those who reported to be given homework scored 10.2 better than those who were not given.

4.2 School Based Factors

Availability of textbooks for pupils was found to be positively casually related to learner achievement. As indicated in table 4.1, pupils who had a textbook in Mathematics had better scores by 5.1 than those who did not have. Similarly, pupils who shared a book with others (at the ratio of 5 pupils per book) had a lower achievement score in Kiswahili as compared to those who did not share at all (1:1 ratio).

A good relationship between pupils and teachers was also observed to influence learner achievement from the table presented. Pupils who talked to their teachers whenever they had problems had better scores in English and

Kiswahili by 8.0 and 9.5 respectively. Although a positive score was also observed for Mathematics,

5.0 RECOMMENDATIONS

5.1 Home Based Factors Orphan hood

Schools administration and teachers should work closely with children's officers and the community to support orphans and pupils staying away from their families.

Parental involvement

Guardians and parents should be involved more in activities that promote learning to build their confidence and capacity to supervise homework and monitor children's progress in school.

Socio Economic Status

Parents should be encouraged to have possessions such as radio and television as they are significant in enhancing learning outcomes through transmission of educational programs.

5.2 School Based Factors Textbook sharing

The school administration should continue reviewing the adequacy and utilization of textbooks, so as to ensure that resources are mobilized from the government to ensure that the 1:1 pupil textbook ratio is achieved in all schools.

Availability of libraries

Given the role of reading in enhancing achievement levels in literacy and numeracy, it is important for schools to create libraries/book corners/book boxes so as to inculcate a reading culture in pupils.

Guidance and Counseling

The school administration should ensure that there is guidance and counseling service in the school and the pupils are able to access this service.

References

- Abagi, J.O. (2007) *Status of education in Kenya: indicators for planning and policy formation* Nairobi, institute of policy Analysis and Research
- Akinsanya O: (2010): *Differential Distribution and Utilization of Human and Material Resources on Students Academic Performance in Secondary Schools in Ogun State. African Journal for the Study of educational issues* Vol (3, 4) 2010.
- Becker & Tomes, N. (1979) An equilibrium theory of the distribution of income and intergenerational mobility. *Journal of political economy*, 87,1153-1189.
- Hewlett, L. [2006] *Youth Sexuality and HIV/AIDS Prevention: A Literature Review*, Graduate School of Public and Development Management: University of Witwatersrand Available at
- Kafwa, V. N. (2005) Effectiveness of AIDS Education Programs in Secondary Schools in Busia District Unpublished M Phil Thesis: Moi University
- Maani, B.F.(2009). *Factors affecting academic performance* . Unpublished dissertation, Makerere University, Kampala, Uganda.
- Macibi, N.A (2005). *Pertinent issues in management: Human Resource and Moyieducational management*. Net Media Publishers Ltd.
- Squire,J.R. (1991): *Textbook Publishing in Encyclopaedia of Educational Research*, vol 4(6th Edition),Macmillan,PP,1419.
- OECD/CERI. (2008). *Assessment for learning: Formative assessment*, International Conference, Learning in the 21stCentury: Research, Innovation and Policy. Paris: OECD
- Ogina, A.T(2007) . *Redefining the role educators in managing the need of Orphaned Learners*. University of Pretoria
- Owoeye and Yara: (2010): *School Facilities and Academic Achievement of Secondary School Agricultural Science in Ekiti State Nigeria*. www.ccsnel.org/ass Asian Siciak Sciences.Vol,7, No.7; July 2011
- Rubaha , R.F.(2008) . *Factors influencing good performance in Tanzania school: University of Dares Salaam*.
- Sylla, K., Saito, M., Ross, K. (2003) *SAMDEM (Sample Design Manager Software)*. Paris: International Institute for Educational Planning.
- Wasanga, P. M., Ogle, M. A. and Wambua, R. M. (2010). *Report on Monitoring Learner Achievement Study for Class 3 in Literacy and Numeracy*. Nairobi: Kenya National Examinations Council.

Assessment Reforms at the Lower Secondary Level of Education in Uganda: From Norm Referenced to Criterion-Referenced Testing

James Turyatemba

Uganda National Examinations Board (UNEB)

P.O. Box 7066, Kampala-Uganda;

Tel. +256-772-625-154; +256-752-625-154.

Email: James.Turyatemba@uneb.ac.ug

ABSTRACT

In Uganda, universalisation of education has improved access to educational opportunity. However, more effort is needed to enhance the quality and relevance of education in this era of rapid change. Consequently, there has been a need to reform the curriculum and assessment system in the country if the skills and competences required for the 21st Century and beyond are to be amplified. This paper brings to the fore Uganda's experiences in reviewing the Lower Secondary curriculum in an attempt to cope with the changing times. It highlights the assessment reforms that have been undertaken to make it responsive to these curriculum reforms. Curriculum, assessment and examination reform is being undertaken at the Lower Secondary school level of education with the aim of tailoring Uganda's education system to the contemporary demands of a rapidly changing global environment. There is a deliberate move to reform the assessment processes and practices in order to avert the backwash effects of summative examinations on the teaching-learning process. This entails a fundamental shift from norm-referenced to criterion-referenced testing. The key elements of the reform have been highlighted as well as the merits of criterion-referenced testing. This transition is expected to yield more reliable information on the learner's actual levels of achievement than the current norm-referenced assessments. The innovation calls for the incorporation of School-Based Assessment grades into the final summative grades to give a more nuanced picture of the learner's competences. Envisaged challenges have been highlighted and recommendations have been made if the assessment reforms are to meet the aspirations of the learner in this era of rapid change.

Key works: Norm-referenced testing; criterion-referenced testing; Secondary education; assessment reforms.

Introduction

Many researchers concur that education is intricately linked to society and national development (Pãrgaru, Gherghina and Duca, 2009). Education must be relevant and of the right quality if it is to enhance national development and economic growth (State of Education in Africa Report (2015).

Developing and developed countries tend to have static curricula in which learners, driven by the need to succeed in high-stakes examinations, are forced to learn a mass of

abstract, fact-centred, decontextualized and irrelevant knowledge (Clegg, Ottevanger, Bregman & Nannyonjo, 2007). Consequently, continuous curriculum and assessment reforms have to be undertaken in response to the rapidly changing global labour market demands (Clegg, et al, 2007). Learners require knowledge and skills to live as effective citizens in the world of work, self-employment and further education (UTAMU, 2016).

Worldwide, many educational reforms have been undertaken at primary and tertiary levels than at the

secondary level (UTAMU, 2016). Uganda deemed it necessary to initiate assessment reforms at the Lower Secondary level, which had an overloaded and anachronistic curriculum (Clegg, etal). The teaching styles promoted by the content-heavy curriculum could not generate the skills needed by a 21st century workforce (Clegg, etal). This made it necessary to reform the curriculum and assessment systems to align them with the changing socio-economic terrain of a globalised environment.

Historical Perspective

Globally, educational assessment systems have influenced the curricula content taught and how it is taught by the teachers. In Uganda, high stakes norm referenced examinations have had backwash effects on the teaching and learning process. Norm referenced tests (NRTs) are designed to highlight achievement differences between and among students to produce a dependable rank order of students across a continuum of achievement from high achievers to low achievers (Stiggins, 1994). A test taker's score is interpreted with reference to the scores of other test takers or groups of test takers, rather than to an agreed criterion (Richard & Schmidt 2002). NRTs depict the learners' achievements compared with the performances of other learners sitting for the same test. They focus mainly on a limited range of competencies in the cognitive domain which makes them rather inappropriate. This has necessitated a fundamental shift from assessment *of* learning to assessment *for* learning that indicates the learners' achievement as judged against pre-determined competencies.

Criterion referenced assessment is an evaluative description of what learners know, understand and can do, without reference to the performance of others (Brown, 1988). A criterion referenced test (CRT) measures a test taker's performance according to a particular standard or criterion that has been agreed upon. The test taker must reach this level of performance to pass the test, and a test taker's score is interpreted with reference to the criterion score, rather than to the scores

of other test takers (Richard & Schmidt 2002). This mode of assessment can also be used to provide a feedback to inform and predetermine the course of future teaching and learning needs. This is expected to enhance the quality of teaching and learning at the classroom level. CRTs focus on specific, well-defined (and usually objectives-based) language knowledge or skills (for diagnostic or achievement purposes), unlike NRTs that tend to measure very general skills and abilities for proficiency or placement purposes (Brown, 2014). They yield a more reliable guide on the learner's actual levels of achievement. It is this actual learning acquisition and outcomes that are of educational significance than the mere juxtaposition of a learner's performance relative to other learners (Turyatamba 2007). CRT results provide a basis for determining how much is being learned by students and how well the educational system is producing desired results (Cohen, Manion & Morrison 2004). These demonstrable merits explain the fundamental shift from norm to criterion referenced assessment in Uganda's curriculum and assessment reform.

Contextual Perspective

There is a growing recognition among African governments on the need to invest in and expand access to secondary education (UTAMU, 2016). Secondary education provides the skills required to meet a country's growing demands for highly skilled and educated workers in a globalized world. The challenge for many African governments is not just to expand secondary education, but to enhance the quality of education. (State of Education in Africa Report, 2015). Uganda's secondary school education system has been criticised for being overloaded, lacking key characteristics of the 21st century skills, lacking the needed flexibility to absorb new knowledge, failure to address the social and economic needs of the country, failure to meet the needs of majority of learners who enter the secondary education, declining and poor quality, being too theoretical, and using teaching methodologies which do

not allow active learning and acquisition of skills for solving contemporary problems (UTAMU, 2006). To correct the above evils in Uganda's secondary education sector, the following interventions were proposed in the reform process:

Shortening of the School Day

It was observed that a long school day puts immense pressure and stress on learners and teachers. There is little room for learners to undertake private study which consequently, stifles innovation and creativity. The school day in Uganda has been shortened from eight to six hours.

Reducing and Re-grouping of Subjects

To enhance efficiency and effectiveness, there has been a phenomenal reduction and re-grouping of subjects from forty-three (43) to nineteen (19). Merging and integration of related content across the formerly independent subjects has eradicated duplication, saved time, financial resources and human effort. The remaining subjects and curriculum content have been carefully selected basing on its direct alignment to the learner's self and national development.

Promoting Technology & Enterprise

Unemployment among school leavers is one of the greatest challenges in this era of rapid change. This has necessitated the promotion of contemporary technological and entrepreneurial competences needed in the world of work.

Amplification of Generic and Soft Skills

Due to rapid global changes, formal education must be transformed to evolve new forms of learning required to tackle the complex challenges ahead (Scott, 2015). Today's learners need to be equipped with critical thinking, creativity, collaborative and communication skills to cope with the anticipated economic and global crises (Barry, 2012). Employers require youths with soft skills: the broad set of skills, attitudes, behaviour and

personal qualities to effectively navigate their environment, work with others, perform well and achieve their goals (Ignatowski, 2017). Unfortunately, soft skills are poorly understood, not well assessed, and often overlooked in policy and institutional contexts, including education, training and the workplace (Ignatowski, 2017).

In a survey done in Uganda on employers' expectations from the school leavers, what was needed more were the generic and soft skills such as effective communication, numeracy, literacy, punctuality, honesty, respect, creativity, analytical thinking and teamwork (Scott, 2015). Most of the education system graduates exhibited deficiency in these skills (UTAMU, 2006), hence the need for their amplification in the reformed curriculum.

The Statement of the Issue at Hand

Uganda's education system was characterized by an overloaded curriculum whose content was not in tandem with the contemporary global demands (Clegg, etal). The system was examination driven, with norm-referenced assessments exerting a pervading influence on classroom practice (Turyatamba, 2007). There was a weak link between assessment grades and the knowledge, attitudes and competences acquired by the learner. Such assessment bred stiff competition among schools and learners. Attention shifted from the intended criterion to the school or learner that had performed better than the other in terms of grades. Teachers were conditioned on what aspects of the curriculum to emphasize and the pedagogical tools to employ if the learners were to pass examinations at all costs. The education system was highly academic, with practical and soft skills reduced to a trivial status (UTAMU, 2006). Consequently, there was a need for a paradigm shift from norm-referenced to criterion-referenced testing meant to achieve effective teaching and learning at the classroom level.

Reviewing the Existing Curriculum in Uganda to cope with the Changing Times

The Ministry of Education and Sports (MoE&S) has been implementing a World Bank funded Curriculum, Assessment and Examination (CURASSE) reform project at the Lower Secondary School level since 2011. Uganda has re-directed its focus towards implementing policies and strategic projects geared towards improving classroom instruction and the quality of education. This is in line with Article 4 of the World Declaration on Education for All (1990) which stipulates that:

“The focus of basic education must, therefore, be on actual learning acquisition and outcome, rather than exclusively upon enrolment, continued participation in organized programmes and completion of certification requirements.” (p.5).

The Lower Secondary curriculum reforms in Uganda are geared towards equipping the learners with relevant knowledge, skills, attitudes and values to fit in a rapidly changing environment. It emphasizes the acquisition of generic and soft skills as well as the acquisition and application of learning (Adams 1998). The type of education imparted should respond to the contemporary socio- economic and political needs and meet the minimum national educational standards.

Aligning Assessment Towards the Curriculum

Whenever there is a significant reform in the country’s curriculum, there is need for a corresponding reform in the assessment and examination system. This is intended to maintain the alignment between the curriculum, instruction and assessment.

Assessment of Practical and Performing Arts Subjects

New forms of assessment have been introduced to assess practical competences in Science, Performing Arts (dance and drama), Technology and Enterprise. Because of the large classes, group presentations may be resorted to. This requires constant monitoring by the teachers who use a checklist in the assessment of group performances.

Transition from Norm to Criterion Referenced Assessment

There cannot be any meaningful curriculum reform unless it is complemented by assessment reform. Consequently, there has been a need to redesign the assessment system with a competence oriented structure that is in tandem with the new curriculum reforms. To avert the backwash effects of norm-referenced examinations on classroom instruction, a deliberate shift has been made towards criterion-referenced testing. This reform emphasizes assessment *for* learning more than assessment *of* learning.

Developing new Grade Awarding and Certification Procedures

Uganda’s curriculum and assessment reform involves developing new grading and certification procedures with a wide range of descriptors intended to make all learners achievers. These describe the types and ranges of performances that students at a particular level should demonstrate. Unlike norm-referenced examinations which have arbitrary pass marks or grade boundaries, and rank learners, criterion-referenced certification describes what students are expected to know and can do. The descriptors inform the stakeholders and the users of this information the meaning of any particular grade in a given subject. This is likely to reduce the current stiff competition among learners and schools which negatively impacts on classroom practices and the quality of education.

Incorporating School-Based Assessment Grades into Summative Grades

Unlike the previous norm referenced examinations, this criterion referenced system is to incorporate School-Based Assessment (SBA) grades into the UNEB summative grades. SBA is formative and diagnostic in nature and is intended to enhance the teaching and learning process. Incorporating SBA grades into the final summative grades tends to reduce the tension among the learners by making examinations more of a classroom

feature. With SBAs, the teacher develops ample skills to make well informed judgements about the learner's achievement, understands the principles of learner progression and uses their judgement to plan ahead for all categories of learners. By having a school based component and being outcome-based, criterion referenced tests are expected to improve teaching and learning because the anticipated learning outcomes are predetermined, a timely feedback to the learner and the teacher is provided, the learners are deeply involved in their own learning through a mixture of interactive activity-based and discursive presentations and teachers adjust their teaching to take into account their observations in the course of assessment. This assessment contributes to a life-long process of learning where the learner is equipped with the basic functional and generic skills that enable them cope with the challenges ahead.

Using Criterion Referenced Assessment to Improve Classroom Instruction

Uganda's move towards criterion referenced testing is based on the principle of teaching and assessing for mastery learning. The intended criterion is predetermined and the expected performance benchmarks are known by both the teacher and the learner. Since these indicators of performance are assigned values that reflect the degree of the learner's mastery, it helps the teacher to make important decisions.

Criterion referenced testing is a powerful tool for enhancing learning and the general quality of education if the learners clearly understand the goals and criteria of their learning, where they are in relation to the intended goals and criteria and how they can achieve them. These issues are central in determining the course of classroom practice and enhance effective teaching and learning.

Criterion-referenced tests determine what test takers can do and what they know, not how they compare to others (Anastasi, 1988). They help to establish how well students have learned the knowledge and skills which they are expected to have mastered. Such information may be used to determine how well the student is learning

the desired curriculum and how well the school is teaching that curriculum.

The current norm referenced mode of assessment tests a limited range of abilities and puts pressure upon teachers for their students to perform well on tests and this has resulted in an emphasis on low level skills in the classroom (Corbett & Wilson, 1991). The criterion mode of assessment is expected to address a wider range of competencies by stressing the acquisition of conceptual understanding as well as the application of practical skills in novel situations to solve problems. This will bring to the surface some of the important learner attributes which are not being captured by the existing pen and paper norm referenced tests. Criterion referenced tests have an advantage of assisting teachers to design teaching programmes that are targeted to improving actual learning in the classroom and providing them with an idea of the areas of difficulty experienced by the learners. This enables the teachers to set realistic improvement targets for individual learners as they progress in the various subjects.

Formative assessment, under normal school conditions, can assess abilities which are not easily assessed in summative examinations, and so tend to be "under assessed". Teachers get to know their learners' abilities better and in depth by tracking each student's progress, whereas formal summative tests, being just a snapshot, give a much less nuanced picture. Once teachers are given clear assessment guidelines with detailed levels of criteria descriptors, they can be able to identify and track which level each learner is working at across each subject.

Implications of the Curriculum and Assessment Reform for Uganda

The MoE&S has had to coordinate the various departments and agencies in the curriculum and assessment reform. Emerging policy and legal issues have to be addressed for the smooth running of the reform process. Funds have to be secured from government and donor agencies to ensure sustainability of this massive reform. There is need to organise teacher refresher and

retooling workshops in order to align them to the reformed curriculum and assessment. A system to monitor the quality of School Based Assessment has to be developed if the grades there from are to be incorporated into summative grades.

There is need for continuous sensitization of the various stakeholders and the general public about the underlying philosophy and principles of the Curriculum reform so that they buy into it. This calls for the intensification of advocacy and communication with relevant departments and Agencies that are involved in this reform process.

A new assessment framework should be developed in order to integrate the changes in the reformed curriculum. There has been need for change in the grading and certificate awarding regulations to harmonize them with the reformed curriculum. The assessment processes and practices have been re-designed so as to flow from the reformed curriculum. The reform makes it imperative to invest in capacity building of test developers, setters, moderators and examiners. Test instruments have to be modified to assess unique areas like Performing Arts, attitudes and values to ensure that they can be timely scored, cognizant of the increasing Universal Secondary Education (USE) learners at this level. Efforts have been made to cater for the interests of learners with special needs as well as private candidates in all the subject areas. The teacher factor is an important cog in the wheel of curriculum and assessment reform. Continuous teacher professional development and appraisal are of paramount importance. There is need to retool teachers who are already in service. Supervising the teacher to prepare standardized formative assessments is another important aspect of this reform. Storage and retrieval mechanisms for students' SBA records have to be put in place by schools. Teachers have to be motivated if they are to engage the learner during the "free period" or else, unintended consequences may arise from administration lapses and/or absence of guidance during this period.

Envisaged Challenges

Uganda is an education hub for students from the Eastern Africa region. The drastic curriculum and assessment reforms are likely to breed some skepticism among the parents and students from those countries whose education systems are quite different. Private school proprietors envisage a drop in the volume of foreign students.

The bureaucratic legislation process to create enabling laws and policies that allow the entire CURASSE reform process to unfold is another challenge.

Strategies to minimise subjectivity in School Based Assessment have to be put in place by the MoE&S. Constant monitoring and evaluation of the progress and quality of SBA is necessary if the grades therefrom are to be incorporated into the UNEB grades.

The reform process requires high level of planning of all assessment and examination reform processes. Criterion-referenced assessment system requires UNEB to adopt new grading and certification procedures. This involves the identification of specific learner achievements on examination components, relating this to grade descriptors and checking decisions against similar evidence from earlier years, where this is available, for purposes of standard setting. In the process of developing a scale of performance descriptors, it is necessary to clearly define the learner's progress at a given level but as Ridgway observed, the definition of mastery is not always clear. It would be impractical to expect candidates to attain perfect scores on every aspect of every task on which they were tested. We are faced with the task of making decisions about the level of success which counts for mastery (Ridgway, 1988).

Teacher motivation is one of the core perquisites for quality education (Bennell 2004). Studies indicate that Professional Development is key to enhancing teachers' motivation (Karabenick & Conley, 2011). Continuous professional development, appraisal and motivation of assessment personnel within UNEB and teachers within the school system are critical for the success of this reform process.

The interpretation and application of holistic performance descriptors in criterion testing involves a risk especially if the criteria become more abstract and complex, requiring greater interpretation, then the assessment can become less objective. Sizmur and Sainsbury (1997) aptly argued that the better we understand the nature of the subject, how it is taught, how pupils learn and make progress in it, the better we are able to rate specific pieces of work in terms of quality. Level descriptions are in one sense a means of imposing coherence on diverse elements of attainment.

Some teachers fear that these reforms may lead to loss of jobs given that some subjects have been dropped. They view SBA as additional burden since it calls for constant learner monitoring, data gathering and remedial instruction of the less gifted learners in an attempt to achieve mastery. They are asking for additional payment for these additional tasks.

Putting in place a fair, valid and reliable standardization process at the school level is another formidable challenge. SBA tests tend to be of low reliability and questionable validity. Teachers may not be able to set tests that meet the national standard and they may fail to explicitly bring out the ways in which SBA can be most effectively used to give a comprehensive picture of the actual learning achievement. SBA is more likely to be abused if teachers' judgements are to be part of the final examination grade.

The other challenge is that of large and overcrowded classes. The African Population and Health Research Center observes that overcrowding in classrooms is a major learning barrier, especially in Uganda's public schools (APHRC, 2016). The high student-teacher ratio coupled with limited science and technology equipment makes it difficult to conduct practical assessments in Science, Creative Arts, Technology and Enterprise related subjects.

Catering for the interests of the learners with special education needs during this reform also poses new challenges. The assessment instruments are adapted to

suit the learner but this has a limit, depending on the nature of the subject and extent of the learner's disability. Poor storage and retrieval of student records is another hurdle given that computer coverage is still limited and a lot of paper work is involved.

CURASSE has not yet been fully appreciated by some stakeholders and there are still some pockets of resistance against the reform initiative. Resistance and sentiments have been met from science and religious education. These sentiments point to lack of adequate consultation or awareness by the MoE&S and the National Curriculum Development Centre. (UTAMU, 2016). Much as consultations with the relevant stakeholders are on-going, there is a tendency for individuals to cling on to old systems and practices. Unless all the stakeholders: parents, policy makers, teachers, learners and the general public are positively thinking about the CURASSE reform, the task ahead remains arduous.

Curriculum and assessment reforms have a technical dimension and also social and political connotations. It is necessary to establish the social and political acceptability of the reform programme right from the outset.

Recommendations

Introducing curricula and assessment reforms to improve the quality of education is one thing, but for the reforms to achieve their intended effect is yet another. If the envisaged reforms are to bear fruit, some sufficient conditions should be fulfilled:

The MoE&S should extend the frontiers of public-private partnership in the funding of the education sector to offset the massive cost of curriculum and assessment reform.

Teacher training is very crucial if teachers are to undertake valid and reliable assessments. Teachers need clear guidelines on how to manage the new curriculum and assessment tasks and how to make judgments of the learners' performances. There is need for local teacher support networks to constantly review SBAs and set standards. A monitoring system to evaluate the progress

and quality of SBA should be developed if the grades generated are to be trusted.

Efforts should be made to avoid wholesale replication of curricula and assessment reforms from other countries. The social, economic and political terrain of the country undertaking educational reforms must be put into consideration.

Curriculum and assessment specialists should form a functional partnership to evolve design principles and test specifications for each subject. Capacity building of Curriculum and Assessment personnel is an indispensable prerequisite for this reform process.

Teacher motivation, professional development and continuous appraisal should be a key component in the reform process. This will go a long way in improving the quality of School Based Assessment.

There is need for active consultation and involvement of the key stakeholders, policy makers, politicians and the general public to enhance the acceptability of the intended reforms. Availing ample information to all the parties at all stages of the reform will help to lubricate the potential frictions at the formative stages of the reform process.

The implementation strategies should not be too ambitious but realistic and achievable.

Conclusion

Other than sorting and ranking learners' performance, norm referenced tests do not explicitly reveal whether the learners have learnt or not the subject content that they were taught. NRTS tend to have adverse effects on the curriculum and instruction. They emphasize memorization of facts and performance of routine procedures, rather than testing problem solving, decision making analytical and social skills. Criterion referenced tests are therefore, an indispensable prerequisite for higher educational attainment. CRTs help in correcting classroom deficiencies that inhibit learning. The results from CRTs provide a basis for establishing how much has been learnt by students and whether the education system

is producing the learner with the desired competences to live and work in this rapidly changing global setting.

REFERENCES

- Adams, D. et al (1994). *Improving Education Quality: A New Approach*. A Paper Presented at Improving Education Quality (IEQ) Seminar, Jinja, Uganda.
- African Population and Health Research Center (APHRC) 2016. *The Quality of Education in Uganda*.
- Anastasi, A. (1988). *Psychological Testing*. New York: MacMillan Publishing Company.
- Barry, M. (2012). *What skills will you need to succeed in the future?* Tempe, AZ, University of Phoenix.
- Bennell P. (2004). *Teacher Motivation and Incentives in Sub-Saharan Africa and Asia*. Knowledge and Skills for Development, Brighton.
- Brown, J. D. (2014). *Statistics Corner. Questions and answers about language testing statistics: Solutions to problems teachers have with classroom testing*. Shiken Research Bulletin, 18(1) 29-33.
- Brown, S. (1988) 'Criterion Referenced Assessment: What Role for Research?' in Black, H.D. and Dockerell, W.D., *New Developments in Educational Assessment*, British Journal of Educational Psychology, Monograph series No. 3.
- Clegg, A. Ottevanger, W. Bregman, J. & Harriet Nannyonjo. H. (2007) *Uganda Secondary Education & Training*

- Curriculum, Assessment & Examination (CURASSE) Roadmap for Reform. Uganda.
- Cohen, L, Manion, L. & Morrison, K. (2004). *A Guide to Teaching Practice*. London: Routledge Falmer.
- Corbett, H.D. & Wilson, B.L. (1991). *Testing, Reform and Rebellion*. Norwood, New Jersey: Ablex Publishing Company.
- Ignatowski, C (2017). What Works in Soft Skills Development for Youth Employment? A Donors' Perspective. Youth Employment Funders Group & Mastercard Foundation.
- Karabenick, S. A., & Conley, A. (2011). Teacher Motivation for Professional Development. Math and Science Partnership Motivation Assessment Program, University of Michigan Ann Arbor, MI 48109
- Pârgaru, I. Gherghina, R. and Duca I. (2009). *The role of education in the knowledge-based society during the economic crisis in Annales Universitatis Apulensis Series Oeconomica, 11(2), 2009.*
- Richard, JC & Schmidt, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics*. London: Pearson Education. Sainsbury, M. and Sizmur, S. (1997) *Level descriptions in the National Curriculum: What kind of criterion-referencing is this?* Oxford Review of Education, 24.2.
- Scott, C. L. (2015). *The futures of learning: what kind of learning for the 21st century? Education research and foresight*. UNESCO Working Papers, 14 November 2015.
- State of Education in Africa Report (2015). *A report card on the progress, opportunities and challenges confronting the African education sector*.
- Stiggins, R.J. (1994). *Student-Centered Classroom Assessment*. New York: Merrill.
- Turyatemba J. (2007). "The Relationship Between Educational Assessment and Learning- The Ugandan Experience".
- A Paper Presented at the 25th Annual Conference of the Association for Educational Assessment in Africa, Kigali, Rwanda.
- UNESCO (1990). *World Declaration on Education for All. Meeting Basic Learning Needs* New York.
- UTAMU (2016) *Curriculum review report on a framework for the harmonization Of Lower Secondary teacher education curriculum with the revised lower secondary school curriculum, assessment and examination reform in Uganda*. Kampala, Uganda

SUB THEME D: ENHANCING THE QUALITY OF PRACTICAL SKILLS ASSESSMENT

Correlation between Practical and Theory Paper Performance in Home Science and Chemistry KCSE Subjects

Maria Atiamuga

**The Kenya National Examinations Council
P.O. Box 73598-00200, Nairobi. Kenya.
matiamuga@kneec.ac.ke**

Abstract

The overall objective of this study was to explore the relationship between the scores in practical and theory papers in Home Science and Chemistry KCSE subjects administered by KNEC. There is general thought that a candidate who has acquired theoretical knowledge and practical skills in a subject will score equally in both areas. This has not been the case and there is concern that in several cases the scores between the practical and the theory papers vary greatly. This study was an ex post facto survey that adopted a desk survey design. Purposive sampling was used to select the KCSE Home Science and Chemistry candidates for the year 2016. The two subjects selected were Home Science because it has two practical papers whose performance varies greatly. Chemistry was chosen for comparison. Data was collected from secondary unpublished sources at KNEC. It was analyzed using SPSS 21.0 and Microsoft excel 2016 for descriptive statistics, correlation and ANOVA. Data presentation is varied. The results showed a difference in the performance of practical and theory papers for both subjects. This study adds to the knowledge gathered on assessment of practical and theory papers. In conclusion there is a difference in the performances of the two practical Home Science papers in relation to the theory paper. It recommends the ministry of education looks into ways of encouraging equal performance in practical and theory papers in every subject.

Keywords: *Practical paper, theory paper, score.*

Background information

The Kenya Certificate of Secondary Education (K.C.S.E) is administered as a national examination at the end of 4 years of secondary education in Kenya. It is administered in all the subjects offered in secondary education including Chemistry and Home Science. This examination is mainly a selection examination to university. The skills and mastery of knowledge are specified in the syllabus and are tested as under Bloom's taxonomy. KNEC develops and administers the K.C.S.E. The examination per subject is designed in such a way that it will test a wide variety of skills the learner has mastered in the best way possible. Therefore a subject

may have more than one paper in the examination to achieve this. In this case the 2 subjects under study have got 3 papers each. Home Science has got 3 papers: Paper 1 (is a theory paper divided into 3 sections and is worth 100 marks. Paper 2 is a practical paper testing skills in clothing construction and is worth 45 marks. Paper 3 is also a practical paper that tests skills in Foods and Nutrition and is worth 25 marks. For purposes of computing the results, the scores of paper 2 & 3 are combined to form paper 2. But in this study the marks of each paper will be used individually to get specific

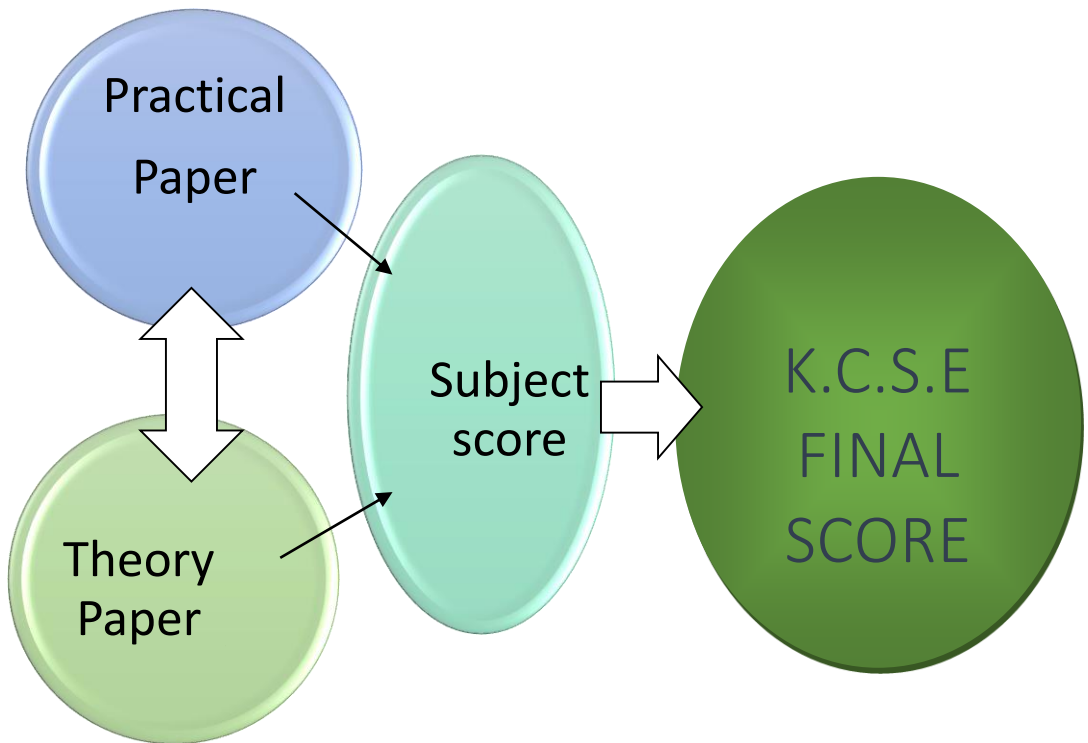
insights about the relationships and the differences. Chemistry is tested using 2 theory papers each marked out of 80 marks and a practical paper marked out of 40 marks.

The two subjects were selected because Home Science has two practical papers whose performance varies greatly compared to the theory paper. One practical paper is in Foods and Nutrition where the scores are normally very high compared to those of Clothing Construction in

relation to the theory paper. Chemistry was chosen for comparison purposes.

Framework

Both scores in the practical paper and the theory paper in a subject add up to make the subject score. The scores of different subjects are then added to give the candidate their final score. It is expected that there is a relationship between the practical and theory paper scores which contribute to the final score of the candidate in KCSE as illustrated in this conceptual framework.



Statement of the Problem

There is general thought that a candidate who has acquired practical skills in a subject is able to translate the same knowledge efficiently to paper in the theory exam and vice versa. However, this has not necessarily been the case for many subjects in KCSE that have both practical and theory examination papers. It has been noted with concern that in several cases the scores between the practical and the theory papers vary greatly. One of the objectives of assessment is to find out the extent to which a student has mastered a subject taught. Does this mean that candidates in such instances are unable to apply knowledge from the theoretical into practical knowledge and vice versa? Is there a relationship in the knowledge taught in both fields or is it a matter of silo teaching whereby learners are not able to apply knowledge across board? As observed by Kolo and Nwaise, 2013; the quality of science graduates produced by African secondary and tertiary institutions is wanting because they somehow possess book knowledge of what they have been taught but lack the practical knowledge.

Purpose of the study

The overall objective of this study was to explore the relationship between KCSE scores for practical and theory papers in both Home Science and Chemistry subjects as administered by KNEC.

Research Questions

1. Is there any relationship between performances of candidates in Chemistry practical and theory papers?
2. Is there any relationship between performances of candidates in Home Science practical and theory papers?
3. Is there any significant difference between the performances of candidates in Chemistry and Home Science?
4. Is there any significant difference between the scores of the two Home Science practical papers?

Research Objectives

The study had the following objectives:

- To investigate the relationship in the performance of Chemistry practical and theory papers.
- To determine the nature of relationship between the performance of candidates in Home Science practical and theory papers.
- To investigate candidates' scores in Chemistry and Home Science.
- To investigate the relationship between the scores of the two Home Science practical papers.

Significance of the study

The importance of this study cannot be underestimated from several vantage points. The study will:

- Add to the knowledge concerning the assessment outcome of practical and theory papers in examinations.
- Make a significant contribution to The Kenya National Examinations Council, to tackle the challenges in assessing theory and practical papers to give a fair assessment of the candidate without over relying on one method of assessment.
- Inform teachers and other stakeholders in education concerning the relationship in performance between practical papers and theory papers of students in order to improve performance.
- Stimulate further research in related areas.

Limitations and Delimitations of the study

The study had an ex post facto design that was a desk review therefore it can only give insights but no definitive conclusion. The study was limited to Home Science and chemistry examinations in secondary schools in Kenya. Therefore, any generalizations and implications of the study findings to other subject areas and examinations must be done with caution.

Literature Review

Learning gaps in assessment

Most classroom instruction takes place with objectives to achieve. One of the ways to determine whether the objectives have been met is through a test. A test is an instrument used to measure sampled behavior. It does not measure the pupil's achievement directly; rather the pupil's characteristics are inferred from the responses to the test questions. Assessment is the process of assigning numerical values to and comparing the values so assigned between two or more. An assumption is made that the scores assigned to the test therefore depict the student's achievement and establishes the strengths and weaknesses of learners.

However, assessment is not just setting a test and producing a test score. When a student scores 7 out of 10 in a class test, what does that tell us? How useful is that score? The learner is unable to successfully tackle the tasks that carry the 3 marks therefore there is a gap in the learning and understanding. 7 out of 10 marks means there is a gap of 3 marks. *A learning gap is the difference between a desired response and what a student provides as a response to an assessment task.* A good assessment defines the size and nature of the learning gap. Is the gap due to *carelessness or oversight, lack of understanding, misconception, inability to apply a concept or confusion?* The variation in the scores could be an indication that the learning gaps in the theoretical and practical areas are different.

It is expected that in a particular subject a candidate is likely to have an almost equal ability in the general content of the subject. Therefore, the learning gaps in the aspects of theory and practical should almost be the same since the content is similar. However apart from the mastery of content the difference could be attributed to other factors that affect the acquisition of knowledge in theory and practical areas.

Assessment of the Home Science and Chemistry practical exams

For effective assessment of acquisition of skills, exams must be valid, reliable, efficient, fair and equal. In Bloom's taxonomy, the 3 domains to be tested are cognitive, affective and psychomotor domain.

Relationship between assessment and Teaching

According to Kithuka (2004), one of the major disadvantages of public exams is the fact that they do not measure the totality of the learner. While they measure the cognitive achievement sufficiently well, they do not measure the affective domain adequately. The best method to use as suggested by Yoloye (1984) would be continuous assessment to give the truest picture of each trainee's ability. Despite all the tremendous benefits of continuous assessment, the question of reliability of the scores from the instructors comes up again. This is also noted by Mutuku (2012), who observed that the coursework marks are highly exaggerated by the instructors because there is no standardized way of scoring the coursework.

Methodology

Description of the study area

The study area includes the whole republic of Kenya as the subjects are done in the whole republic. This includes all the 47 counties in the republic.

Research Design

This study was conducted after the KCSE examination was administered and there no way the variables could be manipulated. Thus it was an ex post facto study that adopted a desk survey design using data collected from secondary unpublished sources at KNEC.

Target Population and Sampling Procedures

The target population of the study was composed of 274,502 girls and 302,751 boys who sat for the 2016 KCSE and attended secondary school in Kenya.

The sample included all the K.C.S.E Chemistry and Home Science candidates for the year 2016.

Their scores for these 2 subjects in theory and practical papers were used.

Purposive sampling was used in order to offer in-depth information about differences and relationships.

Data Analysis

The results will lead to conclusions about associations and relationships.

Frequencies and percentages were used to summarise categorical data which included demographic data. Descriptive statistics were also used to analyse data. For inferential purposes Chi-square was computed to test for associations between the practical papers and the theory papers. The above data was analysed using Microsoft excel 2016 and the Statistical Package for Social Sciences (SPSS-X) 21.0.

Results and Discussions

Introduction

The overall objective of this study was to explore the relationship between KCSE scores for practical and theory papers in both Home Science and Chemistry subjects as administered by KNEC. To achieve this

Background information

purpose the following objectives were formulated and used to guide the study:

- To investigate the relationship in the performance of Chemistry practical and theory papers.
- To determine the nature of relationship between the performance of candidates in Home Science practical and theory papers.
- To investigate candidates' scores in Chemistry and Home Science.
- To investigate the relationship between the scores of the two Home Science practical papers.

The findings of this study are presented in this chapter under the subheadings outlined below:

- Performance of candidates in Chemistry practical and theory papers.
- Performance of candidates in Home Science practical and theory papers.
- Relationship between the scores in practical papers and theory papers.
- Relationship between the two practical papers in Home Science.

Table 1

Distribution of candidates according to gender and subject

| | Home Science | Chemistry |
|---------------|---------------------|------------------|
| Gender | | |
| Male | | |
| Female | | |
| TOTAL | | |

Figure. 1. Bar chart showing distribution of candidates in Home Science according to gender

Figure.2. Bar chart showing distribution of candidates in Chemistry according to gender

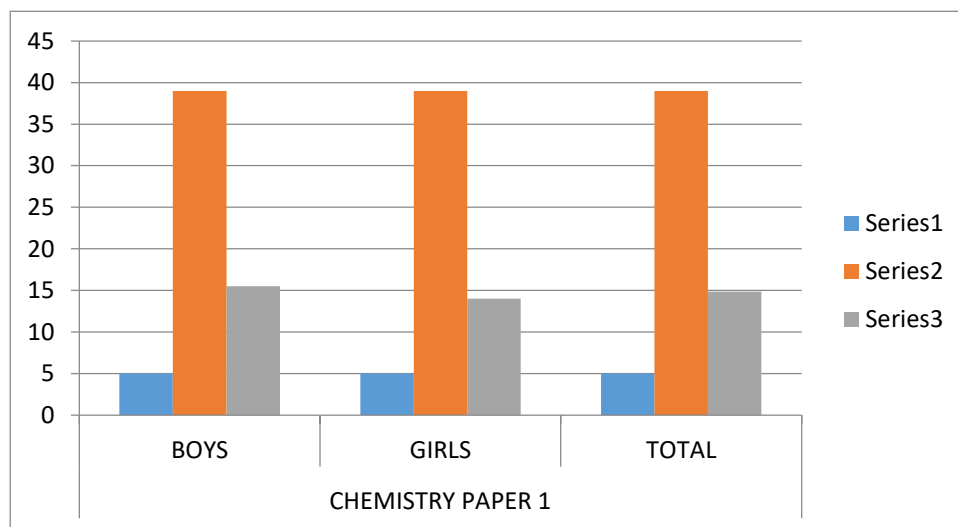
Performance of Candidates in Chemistry Theory Papers

Table 2: Table showing performance of candidates in chemistry Paper 1(Theory)

| | BOYS | GIRLS | TOTAL |
|--------|--------|--------|--------|
| Mode | 5 | 5 | 5 |
| Median | 39 | 39 | 39 |
| S.D | 15.491 | 14.024 | 14.848 |

There was no difference in the performance of girls and boys in this paper. The spread of the scores was good.

Figure. 3. Bar chart showing performance of candidates in chemistry Paper 1



Key: Series 1 – Mode Series 2 – Median Series 3 – Standard Deviation

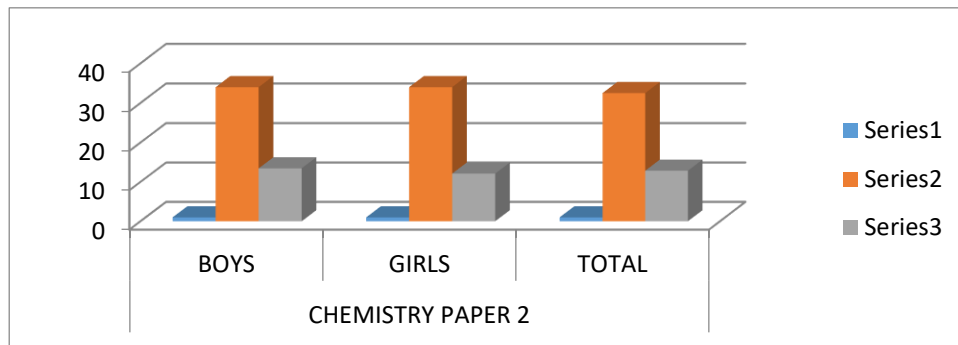
Performance of Candidates in Chemistry Paper 2 (Theory)

Table 3: Table showing performance of candidates in chemistry Paper 2

| | BOYS | GIRLS | TOTAL |
|--------|-------|--------|--------|
| Mode | 1 | 1 | 1 |
| Median | 34 | 34 | 32.5 |
| S.D | 13.45 | 12.114 | 12.852 |

There was no significant difference in the performance of girls and boys in this paper. The scores were well spread as shown by the Standard Deviation.

Figure.4. Bar chart showing performance of candidates in chemistry Paper 2



Performance of Candidates in Chemistry Practical Paper 3

Table 4 Table showing performance of candidates in chemistry Paper 3

| | BOYS | GIRLS | TOTAL |
|--------|-------|-------|-------|
| Mode | 10 | 11 | 10 |
| Median | 19 | 19 | 19.5 |
| S.D | 6.527 | 6.054 | 6.307 |

Although the mode for this paper was better than for paper 1 and 2, the performance in this paper 3 was the poorest in chemistry. There was no significant difference between the performance of girls and boys.

Figure.5. Bar chart showing performance of candidates in chemistry Paper 3

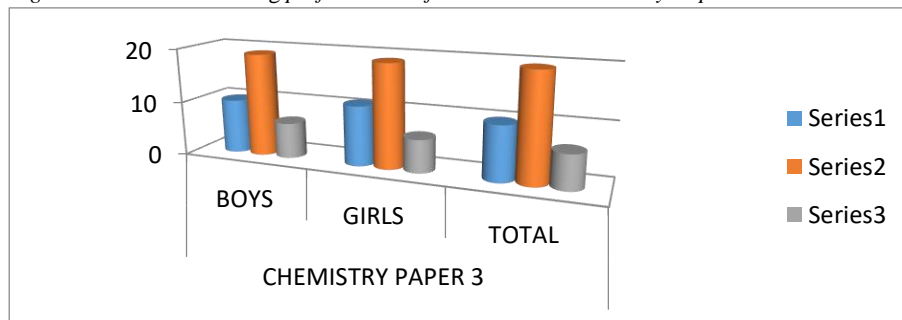
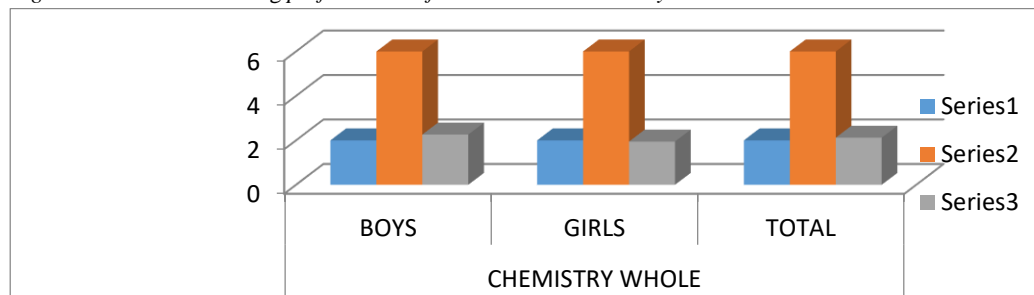


Table 5 Table showing performance of candidates in Chemistry Paper as a whole

Chemistry paper as a whole

| | BOYS | GIRLS | TOTAL |
|--------|-------|-------|-------|
| Mode | 2 | 2 | 2 |
| Median | 2.261 | 1.946 | 2.121 |
| S.D | 2.261 | 1.946 | 2.121 |

Figure.6. Bar chart showing performance of candidates in chemistry as a whole



One way Analysis of Variance (ANOVA) showed that there were no significant differences between the boys and girls. The performance in Paper 1 and Paper 2 which are theory papers is comparable in both mean and median although that of Paper 2 is poorer. However, there is a significant difference in performance between Paper 1 and Paper 3 (Practical) and between Paper 2 and Paper 3

Practical. The mode for Paper 3 is higher in value than for Paper 1 and Paper 2 but Paper 3 has the lowest median of 19. The Standard Deviation (S.D) for paper 3 is the lowest at 6.307; an indication of poor distribution of the scores. The performance of the theory papers is better than that of the Practical paper.

Performance of Candidates in Home Science Theory Paper 1

Table 6 Table showing performance of candidates in Home Science Paper 1

| | BOYS | GIRLS | TOTAL |
|--------------------|--------|--------|--------|
| Mode | 51 | 53 | 53 |
| Median | 46 | 46 | 45.5 |
| Standard Deviation | 13.051 | 12.227 | 12.528 |

The scores are totaled out of 80; this depicts very strong scores by the candidates. There was no significant difference in the performance of girls and boys in this paper. The spread of the scores was good.

Figure.7. Bar chart showing performance of candidates in Home Science Paper 1

Performance of Candidates in Home Science Theory Paper 1

Table 7 Table showing performance of candidates in Home Science Paper 2

| | BOYS | GIRLS | TOTAL |
|--------------------|-------|-------|-------|
| Mean | | | |
| Mode | 26 | 31 | 31 |
| Median | 25 | 25 | 26 |
| Standard Deviation | 7.765 | 7.602 | 7.669 |

The performance in this paper is slightly poorer than that of paper 1 with both the mode and median being of a lower value than that of paper 1. The spread of the scores

is also poorer compared to that of paper 1. The girls had a slightly better performance than that of the boys but the difference was not significant.

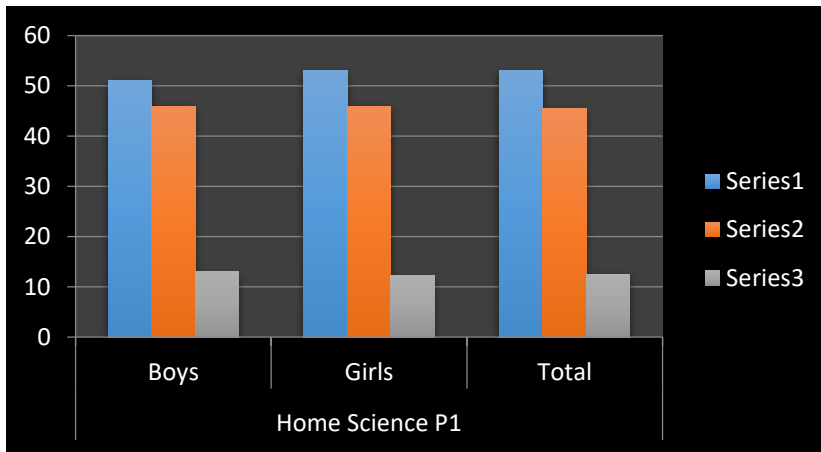


Figure.8. Bar chart showing performance of candidates in Home Science Paper 2

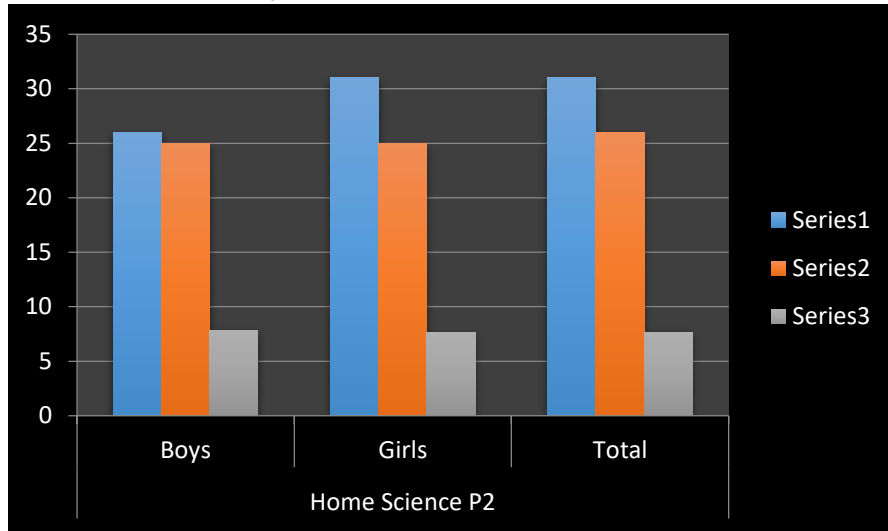


Table 8 Table showing performance of candidates in Home Science Paper 3

Home Science Paper 3

| | BOYS | GIRLS | TOTAL |
|--------------------|------|-------|-------|
| Mode | 21 | 23 | 23 |
| Median | 21.5 | 21.5 | 24 |
| Standard Deviation | 9.87 | 3.858 | 4.773 |

The scores of this paper are given out of 25 meaning that a conversion to percentage would give very high figures. These figures indicate extremely good performance in this practical paper 3 of Foods and Nutrition. Girls performed slightly better than the boys in this paper, however, the spread of the scores for the girls is very poor (3.858) compared to that of the boys who had a standard deviation of 9.87.

Figure.9. Bar chart showing performance of candidates in Home Science Paper 3

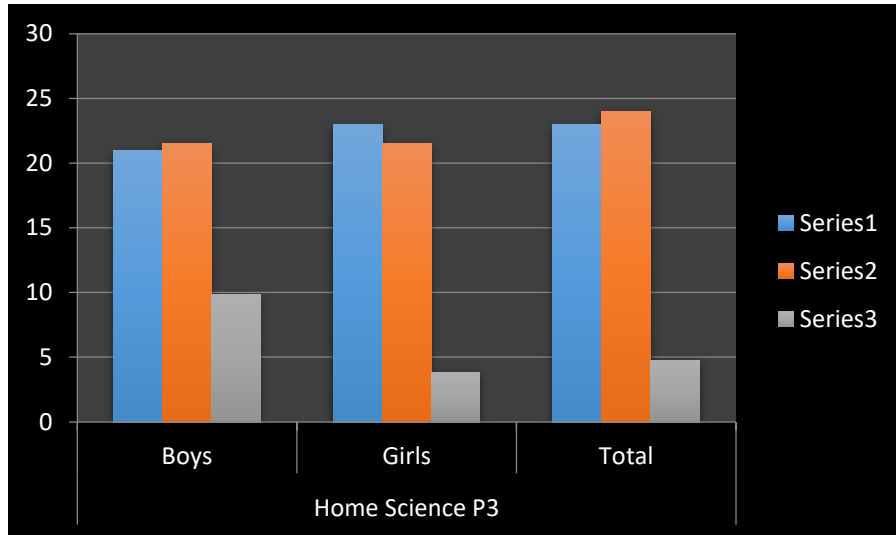
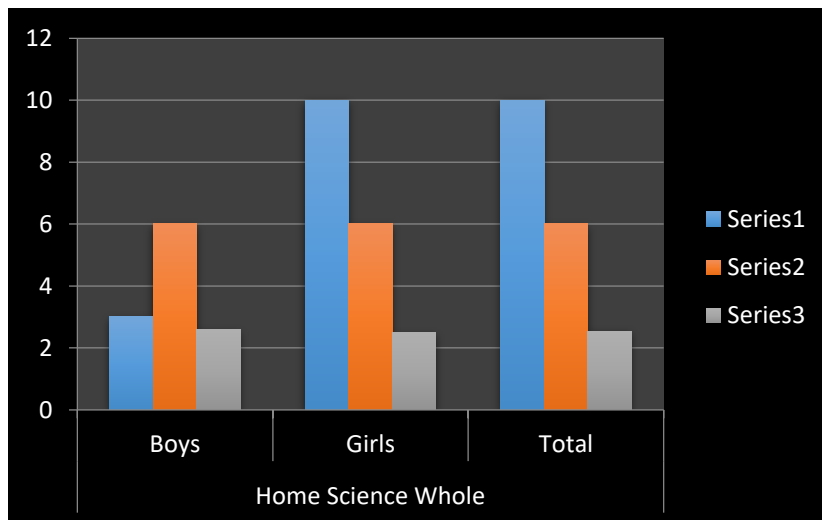


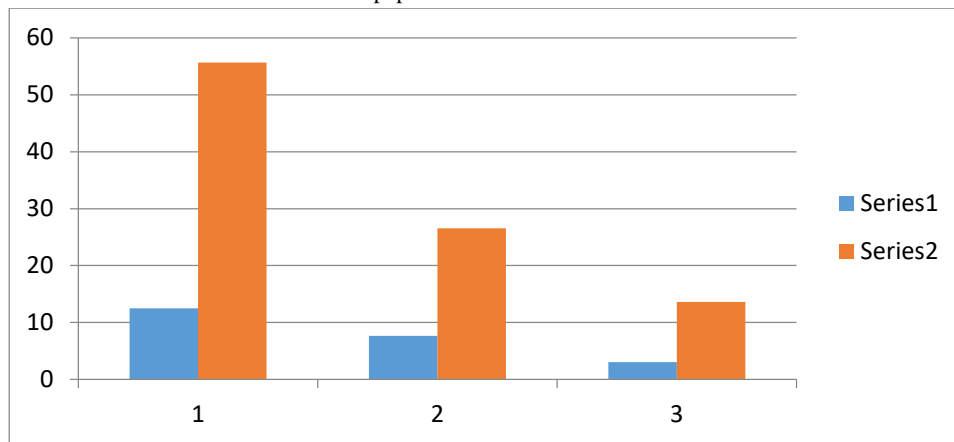
Table 9 Table showing performance of candidates in Home Science as a whole

| | BOYS | GIRLS | TOTAL |
|--------------------|-------|-------|-------|
| Mode | 3 | 10 | 10 |
| Median | 6 | 6 | 6 |
| Standard Deviation | 2.581 | 2.488 | 2.544 |

Figure.10. Bar chart showing performance of candidates in Home Science as a whole



Standard Deviation and mean of all 3 papers in Home Science



Series 1 – Standard Deviation

Series 2 – Mean

There is a clear variation in the performance of the papers. In this graph, paper 3 seems to be the worst performed, but this is because it is marked out of 25 marks. Paper 2 is marked out of 45 marks.

Home Science reveals that there is a difference in performance in the two practical papers. Paper 3 has better performance than paper 2. When compared to the theory paper, paper 3 still has better performance than paper 1. However, it is interesting to note that paper 2 is poorly performed as a practical yet both practicals are in the same subject area.

These findings reveal that there is no relationship in the performance of Home Science and Chemistry practical papers. Therefore the null hypothesis which stated that: there is no relationship in the performance of Home Science and Chemistry practical papers was accepted.

Conclusions and Recommendations

Conclusions

The following conclusions were made basing on the findings of this study:

- There is no significant relationship between performances of candidates in Chemistry practical and theory papers.
- There is a significant relationship between performances of candidates in the Foods and Nutrition practical and the theory paper; however, there is no relationship between the clothing and textiles practical and the theory paper.
- There is a significant difference between the performance of candidates in Chemistry and Home Science.
- There is a significant difference between the performance of females and males in Chemistry and Home Science.

Recommendations

- The Ministry of Education should look into ways of encouraging equal performance in all the papers in every subject.
- Further research should be carried out in this area.

References

- Gupta, S.P. (2014). **Statistical Methods.** New Delhi: Sultan Chand & Sons. Roopak Printers
- KNEC (2015). The year2014 KCSE EXAMINATION REPORT. Nairobi. Kenya
- Kithuka, M. (2004). **Educational Measurement and Evaluation. A guide to teachers.** Egerton University.
- World Bank. (1995). Technical paper No.272. Washington D.C.: Author
- Yoloye, E.A., (1984). **Continuous Assessment. A simple guide for teachers.**Eastbourne: Cassell Ltd.

Improving the Quality of Science Practical Skills Assessment in Secondary Schools in Nigeria

Dr. I. E. Anyanwu¹

Dr. F. I. W. Onwuakpa²

Dr. I. U. Ezenwanne²

Tel: + 234- 8034522829

+ 234-8023806111 + 234-8035063496

Email: aikmanuel@yahoo.com, felixonwuakpa@gmail.com & ucheezen@yahoo.com

Department of Quality Assurance¹ and Psychometrics²,
National Examinations Council (NECO), Nigeria.

ABSTRACT

World over, the acquisition of practical skills for science subjects in secondary schools forms a key part of most science curricula, making the assessment of science practical skills a crucial aspect of quality assurance mechanism in science subjects certification. If the practical skills to be assessed are not clearly defined, their assessment may become merely a process in which students learn how to obtain high marks for summative examinations instead of the opportunities to develop their practical skills. Lack of adequate science facilities and poorly trained teachers affect how science practical skills are assessed in a summative assessment. To this end, the study sought to ascertain the factors responsible for improving the quality of Science Practical Skills Assessment in secondary schools and determine which of the factors prove to be very effective in improving the quality of science practical skills assessment in secondary schools. Hence, the study adopted a survey design and used a sample of one hundred and sixty (160) science teachers in secondary schools in Niger State, Nigeria. Science Practical Skills Assessment (SPSA) questionnaire was used as an instrument for data collection. The analysis of data was carried out using frequencies and percentages. The results revealed that: more than two-third of the students were assessed from their processed data and reports; practical skills assessments were usually not conducted as a regular part of learning science; inadequate practical facilities; unqualified science teachers and lack of clear understanding of what practical skills to be assessed by the teachers from the curriculum. The study recommends that to improve science practical skills assessment, the use of both direct and indirect methods of carrying out science practical skills should be used. Also, schools should be well equipped with adequate and functional science apparatus, training of teachers on the practical skills to be assessed should be on regular basis and qualified teachers should be employed.

Keywords: Practical Skills Assessment, Summative and Formative Assessment, Pedagogy.

Introduction

Science Practical Skills have become very essential in the teaching and learning of sciences. Globally, science curricula in schools have adopted the assessment of practical skills and theory in science subjects for certification in both summative and formative assessment. This is because to strike a balance in learning outcomes, there is the need to integrate both theory and practical aspects of science in the teaching-learning

process as well as in assessment. Repeated practical work or experiment helps to hone practical skills in students and improves the learning of science. Also, without experiments, many students will view science as only a theoretical concept. This is because effective learning of science involves observing and manipulating objects. Some of the objects require special skills to effectively manipulate them. These special skills are known as science practical skills. Science practical work or

experiment has revolutionized the teaching, learning and assessment of science. Practical work refers to any science teaching and learning activity which involves students, manipulating and/or observing real objects and materials, as opposed to the virtual world (Science Community Representing Education (SCORE, 2008)). Central Board of Secondary Education (2009) on the other hand defined Practical work as “an attempt to describe the working of the real world around us, most effective ways to generate interest in science as well as to promote the basic skills and competencies of doing science”(p. 3). These basic skills include procedural skills, manipulative skills, observation skills and skills of representing and interpreting data. Without achieving a certain level of competency in these skills, practical learning of science would be impaired.

Teaching of these science practical skills is widely assumed to be difficult by science teachers, and assessing them is even considered more difficult. Science educators and scholars focus more on the pedagogy of the subject rather than on its assessment. This is because of the difficulty encountered in assessing science practical skills and the teachers’ poor assessment skills. According to Abrahams & Saglam (2010), it is assessment that drives what is taught, to the extent that teachers’ preferences for using different types of practical work are routinely influenced by their considerations of curriculum targets and methods of assessment. This is probably why science practical skills are taught without the underlining skills required to make the students grasp the subjects fully.

Reiss, Abrahams and Sharpe (2012), stated that practical skills are those skills whose mastery increases a student’s competence to undertake any type of science learning activity. Since mastery of science practical skills are essential to science, its assessment to determine the level of students’ competence is equally crucial. They opined that practical skills are, in some cases, best assessed directly. However, they categorized practical skills assessment into direct and indirect assessment. They

referred to direct assessment as direct assessment of practical skills (DAPS); while the indirect assessment is referred to as indirect assessment of practical skills (IAPS). The DAPS refers to any type of assessment where students directly demonstrate a specific or generic skill in such a way that can be used to ascertain their level of competence in that skill. IAPS on the other hand, refers to any type of assessment in which students’ level of competency on a specific or generic skill can be ascertained through inference from student’s data and/or reports of the practical work they undertook.

In recent times in Nigeria, there has been a growing anxiety over the teaching and learning of science in secondary schools. According to Ogunmade (2005), the quality of science teaching and learning has also been questioned over time by parents, science educators, the general public and even by the government. The assessment of science practical skills is not left out. This can be seen before the conduct of public examination in science subjects, assessment tasks are distributed to schools within a certain period prior to the date of the examination as ‘instructions to schools’. Here, the schools are given a list of items to be used for the practical examination or the tasks to be performed. The length of time prior to the examination differs. It ranges from a week to three weeks. These lists propel the teachers to concentrate on the areas that concern the tasks to be assessed. Consequently, the teachers teach only the required practical skills and prepare their students well for the practical tasks ahead, hence ‘teaching to test’. Hoe and Tiam cited in Reiss, Abrahams & Sharpe (2012), see the benefits of such an approach for teachers to include helping the students to understand the skills required for the examination and providing opportunities for professional development of the teachers through internal moderation sessions. However, the aim of ensuring competency in these skills is lost and the “skills” are forgotten as soon as the examinations are over. This study looks at how pedagogy, curriculum, teachers’ quality, quality of science facilities and methods of assessment

can be used to improve the quality of science practical skills assessment in secondary schools in Nigeria.

Statement of the Problem

Science Practical work constitutes a major part in science education. If it is not properly developed in students, teaching and learning as well as the assessment in science subjects would be negatively affected. It requires adequate laboratory with some basic functional equipment to function properly. In Nigeria, this is lacking in many schools. The schools also lack qualified science teachers who possess adequate knowledge of science practical skills assessment. According to Watts (2013), Practical science is considered to be pedagogic and hence the difficulty in defining what should be assessed in a Public examination. There is also the difficulty in defining what should be assessed in a summative examination. Could it be that lack of infrastructure, qualified teachers, and inadequate science facilities and, more importantly, lack of reliable assessment have resulted in the unfortunate neglect of science practical skills acquisition and assessment in most secondary schools in Nigeria?

Purpose of the Study

The study is aimed at determining how quality of science teachers, quality of school science facilities, methods of science instruction, methods of science practical skills assessment and clear understanding of what science practical skills to be assessed in curriculum help to improve the quality of science practical skills assessment for both summative and formative assessment in secondary schools in Nigeria.

Objective

The general objective of the study is to improve the quality of science practical assessment in secondary schools in Nigeria.

Hence, the specific objectives of the study are as follows:

- i. To ascertain the factors responsible for improving the quality of Science Practical Skills Assessment in secondary schools;
- ii. to determine which of the factors prove to be very effective in improving the quality of science practical skills assessment in secondary schools; and
- iii. to make recommendations on how best these factors could be put into practice for better in results effective science practical assessment, teaching and learning.

Research Questions

The following research questions guided the study:

- (1) Does pedagogy help to improve the quality of science practical skills assessment in secondary schools in Nigeria?
- (2) Can clear understanding of science practical skills to be assessed help to improve the quality of science practical skills assessment in secondary schools?
- (3) Can quality teachers and science facilities help to improve the quality of science practical skills assessment in secondary schools in Nigeria?
- (4) Which science practical skills assessment method(s) can help to improve the quality of science practical skills assessment in secondary schools in Nigeria?
- (5) What are the factors responsible for poor science practical skills assessment in your school?

Methodology

Design of the Study: A descriptive survey research design was adopted in this study. This was because none of the independent variables was manipulated.

Population of the Study: The population of the study consists of all science teachers (Agricultural Science, Biology, Chemistry and Physics) teaching in secondary schools in two Local Government Areas (Chanchaga and

Bosso) that make up Minna Metropolis, the capital of Niger State of Nigeria.

Sample and Sampling Techniques: A sample of 160 teachers was used in the study. Simple random sampling technique was used to select thirty (30) senior secondary schools from the population, fifteen (15) from each of the Local Government Areas. At least, four teachers were randomly selected from each school, comprising at least a teacher each for Agricultural Science, Biology, Chemistry and Physics.

Instrument for Data Collection: The instrument used for data collection was Science Practical Skills Assessment (SPSA) questionnaire developed by the researchers. The items were constructed to answer the research questions formulated to guide the study. The questionnaire consisted of two sections: Section 1 elicits information on the respondents' personal data, while Section 2 contained five parts of items structured to provide answers to the major research questions.

Validation of the Instrument: The questionnaire was validated by three measurements and evaluation experts with science education background. A pilot study was also carried out using twenty non-participating teachers from the population and content validity was obtained using Lawshe's Content Validity Ratio (CVR). The CVR coefficient of 0.79 was obtained.

Reliability of the Instrument: The reliability of the instrument was established using the Cronbach-Alpha Coefficient for establishing reliability. Reliability coefficient of 0.85 was obtained for the instrument. This high value showed that the instrument was reliable and suitable for the study.

Procedure for Data Collection and Analysis: The researchers administered 160 questionnaires in the thirty (30) selected senior secondary schools. The researchers and trained research assistants administered the

questionnaire to the respondents. Only 124 questionnaires were received from the respondents. This give a 77.5% return on questionnaire which was adequate for the study. The data for the study were analyzed using frequency counts and percentages.

Results/Findings**Table 1: Teachers' response on the factors responsible for improving the quality Classroom of Science Practical Skills Assessment in secondary schools**

| S/N | Questions | Yes | No | Total |
|-----|--|------------|------------|------------|
| 1. | Do you assess your students' practical skills after practical science lessons (indirectly), based on their processed data and reports?? | 83(66.9 %) | 41(33.1 %) | 124(100%) |
| 2. | Do you assess your students' practical skills during practical science lessons (direct)? | 28(22.6%) | 96(77.4%) | 124(100%) |
| 3. | Are the learning objectives of science practical work well stated? | 85(68.5 %) | 39(31.5 %) | 124(100%) |
| 4. | Can pedagogy help to improve the quality of science practical skills assessment in secondary schools in Nigeria? | 65(52.4 %) | 59(47.6 %) | 124(100%) |
| 5. | Does the curriculum clearly spell out what practical skills are to be assessed? | 56(45.2%) | 68(54.8%) | 124(100%) |
| 6. | Does the curriculum provide for adequate feedback for students? | 36(29%) | 88(71%) | 124(100%) |
| 7. | Do you take part in assessing your students' science practical work in public examinations? | 12(9.7%) | 112(90.3%) | 124(100%) |
| 8. | Should teachers participate in the assessment of their students' science practical skills in public examinations? | 101(81.5%) | 23(18.5%) | 124(100%) |
| 9. | Can clear understanding of science practical skills to be assessed help to improve the quality of science practical skills assessment in secondary schools? | 70 (56.5%) | 54 (43.5%) | 124(100%) |
| 10. | Can quality of teachers and science facilities improve SPSA? | 83(69.9%) | 41(30.1%) | 124 (100%) |
| 11. | Is assessing students science practical skills using both direct and indirect methods the most effective way of improving the quality of science practical skills assessment | 73(58.9%) | 51(41.1%) | 124 (100%) |

Research Question 1: Does pedagogy help to improve the quality of science practical skills assessment in secondary schools in Nigeria?

The responses to the question of which teaching method(s) can be used to improve the quality of science practical skills assessment in secondary schools in Nigeria showed that there are various methods that can be

used to improve the quality of science practical skills assessment in secondary schools in Nigeria. The results indicated more than half of the respondents (50.8 %) held the view that the teaching method that would help to improve the quality of science practical skills assessment was the demonstration method. Other methods are Discussion (8.9%), Group activity (10.4); Role play

(6.5%); Assignment (3.2%); Seminar (2.4%); Brain storming (9.7%) and others (8.1%). On the issue of classroom assessment of science practical skills, 28(22.6%) of the respondents indicated that they assess the science practical skills of their students during practical science lessons while 56(77.4%) of the respondents do not; 22 (17.7 %) of the respondents assess the practical skills of their students after practical science lessons while 52(82.3 %) of the respondents do not; and 83(66.9 %) of the respondents indicated that the assessment of their students should be based on the students' processed data and reports while 41(33.1 %) of the respondents thought otherwise. More than two-third 85(68.5 %) of the respondents held the opinion that the learning objectives of science practical work were well stated, while 65(52.4 %) stated that pedagogy could help to improve the quality of science practical skills assessment in secondary schools.

Research Question 2: Can clear understanding of what science practical skills to be assessed help to improve the quality of science practical skills assessment in secondary schools?

On whether clear understanding of science practical skills to be assessed would help to improve the quality of science practical skills assessment in secondary schools, the results showed that 70(56.5%) of the respondents indicated that having a clear understanding of the science practical skills to be assessed will help to improve the quality of science practical skills assessment in secondary schools while 43.5% indicated otherwise. Responses on the clarity of science practical skills assessment in the current national curriculum indicated that a total of 68(54.8%) of the respondents stated that the current curriculum clearly spelt out the practical skills to be assessed and specified how the students should be assessed 61(49.2%). However, 75(60.5%) held the view that curriculum limits the practical skills to be assessed and does not give room for adequate feedback to students in science practical skills assessment 88(71%).

Research Question 3: Can quality of teachers and science facilities help to improve the quality of science practical skills assessment (SPSA) in secondary schools in Nigeria?

More than two-third 83(69.9%) of the respondents were of the view that the quality of teachers and science facilities can help to improve the quality of science practical skills assessment in secondary schools in Nigeria, while 41(30.1%) opined that they will not. On availability of science practical skills facilities and personnel, 74(59.7%) of the respondents indicated that laboratory facilities in schools were inadequate. Though, 91(73.4%) of the respondents stated that science textbooks were adequate in schools, science practical skills texts and manuals were inadequate 87(70.2%). Seventy-Seven (62.1%) and 67(54.0%) of the respondents indicated that the quantity and quality of science practical skills equipment were very inadequate respectively. Most of the respondents 102 (82.3%) opined that the schools have inadequate number of qualified science teachers and laboratory attendants.

Research Question 4: Which science practical skills assessment method(s) can help to improve the quality of science practical skills assessment in secondary schools in Nigeria?

On how the respondents assess their students in Science practical skills, two-third of the respondents 82 (66.1%) assess their students from the students' processed data and written reports of the practical works they carried out, 20(16.1%) assess their students as they carry out practical work and 22 (17.8%) do so both when carrying out science practical work and from their students' processed data and written reports. On the best method of assessing science practical skills for improving the assessment of science practical skills in secondary schools, 73(58.9%) of the respondents indicated that assessing students while carrying out science practical work as well as from their processed data and written reports was the best. Also, 31(25%) said assessing while carrying out science practical work was the best method and 20(16.1%) stated that assessing from students' processed data and written

reports was the best. Furthermore, on whether the teachers take part in the assessment of their students' science practical skills in public examinations, 112(90.3%) of the teachers stated that they do not take part. On whether teachers should participate in the assessment of their students' science practical work in public examination, 101(81.5%) of the respondents strongly believe that teachers should participate in the assessment of their students' science practical skills in public examinations.

Research Question 5: What are the factors responsible for poor science practical skills assessment in your school?

Factors responsible for poor science practical skills assessment

The most important factors responsible for poor science practical skills assessment in secondary schools indicated by the respondents include:

- i. inadequate laboratory equipment(19%);
- ii. teachers' poor knowledge of the subject matter(14%);
- iii. poorly trained science teachers in science practical work assessment(14%);
- iv. inappropriate methods of assessment(11%);
- v. inadequate practice time for students(10%);
- vi. lack of clear understanding of the science practical skills to be assessed(7%);
- vii. very large science practical classes(7%);
- viii. lack of proper monitoring of practical skills assessment by supervisory authorities(6%);
- ix. absence of laboratory attendants in many school laboratories(5%);
- x. absence of science laboratory(5%); and
- xi. lack of safety tools in school science laboratories(2%).

Factors for Improving Science Practical Skills Assessment

To improve the assessment of science practical skills in secondary schools, the respondents suggested:

- i. procurement of adequate and better science equipment and facilities (20%);
- ii. training of teachers in science practical skills assessment (15%);
- iii. use both direct and indirect practical skills assessment procedures in carrying out students' science practical skills assessment (12%);
- iv. more time should be allotted to science practical works and its assessment (12%);
- v. award more marks to science practical work than the theory(10%);
- vi. curriculum should clearly specify what practical skills students should master and the ones they are to be assessed in (9%);
- vii. encourage group practical work and assessment(8%);
- viii. assessing students when they are carrying out practical work(5%);
- ix. provide moderate class size for science practical work and assessment(5%);
- x. need for feedback on students' science practical skills assessment(2%); and
- xi. employ well trained laboratory attendant/assistant to handle the laboratories(2%).

Discussion

Several methods were used in teaching science practical skills but demonstration method was considered to be the most preferred teaching method for improving the quality of science practical skills assessment in secondary schools in Nigeria. Other methods worthy of mentioning are group activities, discussion and brainstorming. This can be attributed to the fact that practical skills involve observing, manipulating and demonstrating. Furthermore, on the issue of classroom assessment of science practical skills, more than half of the respondents

assess their students' practical skills after practical work lessons while more than two-third assess their students from their processed data and reports. This corroborates the finding of Afemikhe & Imobekhai (2014) that practical assessments are usually not conducted as a regular part of learning science. The teachers conduct assessment only during termly examination while others do not conduct any practical assessment.

Having clear understanding of the science practical skills to be assessed can help to improve the quality of science practical skills assessment in secondary schools. This was the view of most of the respondents. Also, the curriculum is seen as the nucleus of any educational system and from where assessment of practical science skills stems from. However, the practical skills to be assessed and how they should be assessed are not clearly spelt out in the curriculum. Perhaps, the inability to comprehend what the curriculum prescribes for practical skills assessment has put many teachers in a wrong stead. This is why many scholars see science practical work as more pedagogical, giving less credence to assessment. Furthermore, the quality and quantity of science practical facilities and teachers play very crucial role in teaching and learning of sciences. It is often said that science practical skills cannot be mastered without a laboratory and other science facilities as well as qualified science instructor. But the data obtained revealed that many schools in Nigeria have inadequate science practical equipment and qualified science teachers and laboratory attendants. Adequate provision of quality science facilities and qualified science personnel will help to improve the quality of science practical skills assessment in secondary schools in Nigeria.

On the issue of the methods teachers adopt in assessing science practical skills, the study revealed that most of teachers assess students indirectly, based on the processed data which the students obtained from their practical works and written reports. This system limits the amount of direct assessment of practical skills if any at

all. This agrees with the study conducted by Donnelly et al.(1996), which concluded that teachers only assess students in practical science skills indirectly. This means that the teachers not devote enough time and effort to developing students' practical skills since the students are not assessed while carrying out science practical work. However, to improve the quality of science practical skills assessment, the students should be assessed using both direct and indirect assessment methods. That is, the students should be assessed while carrying out practical work as well as from their processed data and reports. Furthermore, teachers' participation in the assessment of their students' science practical work in public examination is crucial in improving the quality of science practical skills assessment in schools. This gives the teachers exposure on how practical skills assessment should be done. This is because many teachers lack the knowledge of how to properly assess science practical skills in school. Also, to improve the assessment of science practical skills in secondary schools requires good science equipment and facilities; increased funding; well trained teachers in science practical skills and its assessment as well as the need for timely feedback on students' science practical skills assessment.

Conclusion and Recommendations

Science practical skills are those skills which students acquire to show their competence to undertake any science activity. As important as these skills are, the current curriculum does not state in clear terms how they can be validly assessed. The curriculum should clearly state which practical skills candidates should develop and how they are to be assessed. The study revealed that these skills were mostly assessed indirectly, but assessing them both directly and indirectly can help to improve the assessment of science practical skills in secondary schools. Consequently, public examination bodies in Nigeria should consider changing from their traditional method of indirect assessment of science practical skills to a combination of direct and indirect assessment to create optimum balance and achieve better results. Public

examining bodies should also state clearly what science practical skills candidates should develop and the ones they are to be assessed in. In the same vein, teachers should participate in the assessment of their students' science practical skills in public examination and it should be done by direct and indirect methods. This means that the direct assessment method should be incorporated into the existing indirect assessment of practical skills in public examination. Classroom assessment of students' science practical skills should also be done by using both direct and indirect methods. And to improve the assessment of science practical skills in secondary schools, more and better science equipment and facilities should be procured, teachers should be well trained in science practical skills teaching and assessment, more time should be allotted to science practical works and its assessment and more marks should be awarded to science practical work assessment.

References

- Abraham, I., Reiss, M.J., & Sharpe, R.M. (2013). The assessment of practical work in school science. *Studies in Science Education*, 49(2), 209-251.
- Abrahams, I., & Sagle, M. (2010). A study of teachers' views on practical work in secondary schools in England and Wales. *International Journal of Science Education*, 32 (6), 753-768.
- Afemikhe, O. A. & Imobekhai, S. Y. (2014). 'Nigeria'. In B. Vlaardingerbroek and N. Taylor (Eds). *Issues in Upper secondary science education: Comparative perspectives* (161-174). New York: Palgrave Macmillan.
- Afemikhe, O. A., Imobekhai, S. Y. & Ogbuanya, T. C. (2016). Assessment of valid science practical skills for Nigerian secondary schools: Teachers' practices and mitigating factors. Central Board of Secondary Education (2009). Class, X. Assessment of Practical Skills in Science and Technology.
- City and Guilds (2003). Guide to the assessment of practical skills in intermediate vocational qualifications. London: City and Guilds of London Institute.
- Donnelly, J., Buchan, A., Jenkins, E., Laws, P., & Welford, G. (1996). *Investigations by order. Nafferton*: Studies in Education Ltd.
- Ogunmade, T. O. (2005). The status and quality of secondary science teaching and learning in Lagos State, Nigeria. Retrieved from <http://ro.ecu.edu.au/theses/86>
- Reiss, M., Abrahams, I. & Sharpe, R. (2012). Improving the assessment of practical work in school science. *A report on the assessment of practical work in school science*. Retrieved from reiss@ioe.ac.uk
- Science Community Representing Education SCORE (2008). Practical work in science: A Report and proposal for a strategic framework. Retrieved from www.score-education.org
- Watts, A. (2013). *The assessment of practical science: A literature review*. Cambridge: Cambridge Assessment.

Benefits, Challenges and Procedures of Introducing Oral Skills Assessment in English Language at Uganda Certificate of Education

Florence Obong Capuc and David M. Weerhe

obongfr@gmail.com and weerhe@hotmail.com

Uganda

Abstract

Benefits, procedures, and challenges of introducing Oral Assessment in English Language at UCE was studied with a view to helping UNEB and NCDC make an informed decision as to whether or not and how to introduce Oral Assessment in English Language at UCE. UNEB is expected to give a well-researched reply about the introduction of Oral Assessment in English Language to the demand driven request from the interested stakeholders. NCDC is hoped to get invaluable information that may help in evaluating the procedures of oral assessment proposed in the reform. The learners will be helped to come out well- rounded with all the four language skills of reading, writing, speaking and listening that will be of use in practical situations. The benefits of oral assessment found during the study were that oral assessment helps in the reduction of plagiarism, tests the candidate's ability to respond clearly, thinking on the spot and ability to handle questions and applications of theory to practice. Oral assessment also tests a candidate's broad understanding of the subject and relevance of learnt material in practical situations. Challenges of oral assessment were found to be in the areas of reliability and validity, examiner/examinee background characteristics, time management, large number of candidates coupled with high costs. A cross-sectional survey research design was used. The population consisted of students, teachers and subject experts of English Language, German, and French subjects. Data collection was done using questionnaires, focus group discussions and interview guides. Analysis was done qualitatively and quantitatively. Conclusion drawn is that oral assessment in English language is worthwhile introducing, despite the challenges cited. The possible implications suggested are: re-tooling of the teachers of English Language on how to handle oral assessment; introduction of classroom based assessment and computer- based program for oral assessment in English Language.

Key words: Oral Assessment; Practical skills; Procedures; Benefits; Challenges

Background

In the more recent times, the World Bank has produced a model for determining the quality of public examinations. One of the identified characteristics for assessing the quality of public examinations according to the World Bank (2000) is beneficial effect on classroom practice. Against this background, the examination system should promote good teaching and learning practices. In other words, an examination system should provide systematic feedback to teachers. Indicators of this characteristic are:

- Encouragement of developments of higher order thinking skills (HOTS) and not placing emphasis on recall of facts (Bloom's Taxonomy, cited in Nitko & Brookhart (2007).
- Promotion of development of practical skills, for example listening and speaking skills in languages and in the sciences (World Bank, 2000).

An examination body of good repute and quality, to a great extent, should satisfy these criteria. Uganda

National Examinations Board (UNEB) has significantly met these practices by assessing all the science subjects both through theory and practical examinations and some of the foreign languages orally, for example French and German languages.

In addition, contemporary trends in educational assessment demand that public examination bodies do not limit their testing activities to cognitive ability of the learners, but should equally cover other vital domains of learning such as affective and psychomotor domains. Moreover, emphasis is being shifted from theoretical accumulation of knowledge to performance-based learning in which students not only acquire knowledge, but equally gain requisite skills that are needed in the world of work/ work place. To effectively do this, oral assessment is a requirement. Joughin (2010, p.1) defines oral assessment as “any assessment of learning which is conducted by the spoken word”. This research tried to determine benefits that will be got, the procedures that can be adopted and challenges that may be faced if oral assessment in English Language is to be introduced at UCE by UNEB.

Statement of the problem

Oral language competence assessment plays an important role in language teaching and learning. According to World Bank (2000), one of the identified characteristics for assessing the quality of public examinations is the beneficial effect on classroom practice and in the world of work. One of the indicators is the promotion and development of practical skills like listening and speaking in the language subjects. Practical skills in this constantly changing environment are also essential in meeting the challenges of everyday life like communication. UNEB has significantly met this requisition by assessing the French and German languages orally but not English Language which is the official and language of instruction during the teaching and learning process. This study therefore examined the benefits to be achieved, the procedures to be followed in

conducting oral assessment, and challenges likely to be faced if oral assessment in English Language is to be introduced at Uganda Certificate of Education examinations.

Purpose

The study examined benefits that can be achieved, the procedures that can be adopted, and challenges that may be faced in the process of introducing oral assessment in English Language at UCE with a view of helping UNEB and NCDC and other examining bodies weigh whether or not to introduce oral assessment in English Language and other languages being examined.

Objectives

The specific objectives of the study were to:

1. Identify benefits of introducing oral assessment in English Language at UCE.
2. Determine the procedures that may be adopted if oral assessment in English Language is introduced at UCE.
3. Establish challenges that may be faced if oral assessment in English Language is introduced at UCE.

Literature review

Oral Assessment Benefits

According to Joughin (2008) oral assessment helps people to identify themselves with their words whereas the writing separates the knower from the known. The students own the words in oral assessment and present them in their own style. The second element in oral assessment is the passion and force with which students express their ideas in front of an audience, which may be an assessor, or a group of assessors. Penny (2008) in his research on the Unexpected Rewards of Qualitative research in assessment reports that one other benefit of the oral assessment is the provision of a good learning opportunity for the student. Scholten (2004) agree that the oral exams are a very useful method of teaching and testing. Oladokun (2009) also says that oral tests that

assess students' skills in spoken language is the best alternative to paper and pencil test. Oral assessment if conducted provides a basis for effective communication in the language which can be used in the world of work and elsewhere. Killen (2001) observes that oral assessment helps students to see the relevance of learned material in practical situations, for example participating in a debate, which may help one to become an active participant in local council debates, parliamentary debates etc. Katz & Stack (2004) state that the combination of linguistic and communicative competence measures how best a learner would achieve academically. Researchers at the Brown University (www.alliance.brown.edu, 2006) agree with Gracia (2002) who says that oral skills could aid comprehension of materials read. They further argue that as oral assessment tests an important transferable skill, it strengthens student employability. Cheng & Curtis (2004) observe that oral assessment can be used to correct or minimize social ills in the allocation of opportunities, as well as to upgrade the performance of academic institutions. Singh (2007) found that use of oral assessment for multicultural groups could provide students with an opportunity to improve their language abilities. He found that use of oral assessment of students can lead to tremendous benefits in terms of language, interpersonal relations and preparation for workplace.

Oral assessment as reported by Nelson (1993) also can produce the added benefit of improved relationship between students and the assessors. Unlike in the written exams when assessors mark scripts of candidates unknown to them, in the oral exams the assessors and the assessed have the opportunity to chat to each other face-to-face on an individual basis, something they do not get the opportunity to do when assessing written exams. As Nelson rightly say "the most unexpected prospect of holding oral exams is an improved relationship with the candidates as individuals. After all, how else would you ever find time to privately chat with an individual you assess?" (p.70). Oral assessment genuinely helps an

assessor to discriminate between a real bright candidate and a mild and moderate/ mediocre candidate. This is why Tzurriel (2000) refers to oral assessment as Dynamic/ Interactive Assessment. Joughin & Collom (2003) say that students will encounter situations which require proficiency in oral communication regularly in their future professional lives, for example job interviews, presentations at conferences or to co-workers. They also state that oral assessment can encourage deep approaches to learning. Students will learn differently for an oral examination, trying to find out what is essential to the course material and get a thorough understanding. Another benefit of oral assessment as they pointed out is the reduction of plagiarism. No Child Left Behind (NCLB) (2002) adds its voice by stating that oral assessment in English language can be an access to career prospects like international jobs and provision of a solid foundation for higher level courses such as GECA and AS levels and the International Baccalaureate Diploma as well as excellent preparation for employment.

Procedures of conducting oral assessment

The literature reviewed is from the various assessment bodies in various languages: West African Examinations Council (WAEC), Cambridge and UNEB. WAEC (2013) shows that the oral examination tests the receptive and productive abilities of candidates in the areas of reading, listening and recognition of different aspects of English speech. Cambridge (2012) assesses candidates' ability to use English as a medium of practical communication. Candidates must demonstrate ability in reading, listening and speaking. UNEB's experience in handling German oral assessment at UCE, UNEB (2008) includes a dictation of a passage in the process of the oral assessment. Candidates listen to conversations or statements each followed by a set of questions to be answered orally. At UACE, the examination consists of oral reading of the passage and verbal responses to the question set on the passage. The questions require analysis, comprehension and evaluation. In handling the oral examination of French at UCE (UNEB, 2008), the

candidates are expected to handle sentence stress, intonation, phrases and clarity, as well as vowel quality and fluency. The candidates should answer questions using a range of structures stimulated by various question structures of the language. The candidates should also be able to converse fluently in French. At UACE French, there is the listening comprehension and spoken expression tested during the oral assessment. The mode of administration is either group or individual.

The research tried to explain which of the procedures used by the different examining bodies can be adopted by UNEB

Oral Assessment Challenges

Oral examinations have been criticized by many authors for lack of reliability (Wood, 1991). According to Joughin (2010), reliability is concerned with how dependent a student's result could be on the case given to the student for examination, the level of difficulty of the follow up questions that are asked after presentation, who examines the student and whether an examiner's assessment could change over a period of time in examining a large number of students. The examiner in an oral examination participates actively in the examination process and their participation could introduce a bias (Joughin, 2008). Finch & Taeduck (2002), in their research on Oral Testing and Self-Assessment-The Way Forward, presents another challenge of oral testing as how to make the oral test a true assessment of spoken abilities rather than how well a student can read a passage or can produce well-memorized responses that have little meaning for him. Lazaraton (2002) (www.cambridge.com) reports that the OPI of the ACTFL, for instance, is criticized for being incapable of measuring oral proficiency needed and displayed in real life situations. Another important challenge that influences the oral assessment is the assessor's attitude towards students from other background. Lambert, Hodgson, Gardner, and Filenbaum (1960) first observed the Linguistic stereotype

hypothesis, which holds that even short samples of non-prestige varieties of speech are sufficient to trigger among listeners a cascade of negative evaluations of speakers. Many of those evaluations are quite extraneous to language behaviors such as origin, attractiveness, civility etc.

Rubin (2002) in his body of research reveals that not only does speech style affect judgments of the assessors, but the converse is also true: social expectations regarding the assessors also affect the student's judgments of their speech proficiency. When students are led to believe that they are listening to an assessor from a certain background, they rate the assessor's speech as more highly accented, and their listening comprehension measurably declines. Another challenge that affects oral assessment is that between written and oral assessments is the amount of time candidates have to answer questions (Utley, Mitchell & Phillips, 1983; Rust, 2001). In their review of oral/aural graded tests, they say that the time pressures in an oral examination are quite different from those in a written paper. There are usually time guidelines for an oral but no strict rule. This takes away from the student the pressure of managing time and spending the correct amount of time on each question. However, there are different time-related pressures in an oral exam. The time is managed by the assessor in an oral examination but not by the assessed. It is the assessor who decides when the student is ready to move on to the next question, and it is the assessor who sets the pace for the examination. Oladokun (2009) agrees with the above report saying that oral test is time consuming, especially, if interactive approach is adopted. Kofitise (2009) says that though assessors could be trained to provide reliable assessment, their ratings could be affected by the number of candidates taking the examination. Abioye (2010) rightly underlined the fact that it is difficult to ensure that all examiners are marking to the same standard when large numbers of candidates, and consequently assessors are involved. In the case of oral tests, large candidature presents peculiar challenges such

as adequacy of time since it commonly involves face-to-face interaction with the assessor which is an effective method of testing many candidates, numerous speech varieties, and idiosyncrasies and bias of many assessors. However, the idea of “large” is relative depending on the environment. Oral assessment is also noted to be very expensive by its very nature (Surface, Poncheri & Bhavsar, 2008). According to Ayesha, Alastaire and Leshia (1999), there appears to be large individual differences in confidence among students in oral examinations. The most confident students are better able to show the examiner what they know in an oral examination than the less confident students, whereas ideally in written examinations, one would not expect to see such large discrepancies.

It is obvious that these challenges are real and formidable, but should they stand in the way of the oral assessment in the English Language as there are immense benefits as shown in the literature review.

Methods/ Procedures

The study employed both an exploratory and descriptive study methods because the research was addressing a subject about which there are high levels of uncertainty, at the same time accurate facts were to be provided for valid representation.

The target population was all the teachers of English Language from the schools, the subject specialists of English Language, French and German from Uganda National Examinations Board (UNEB) and National Curriculum Development Centre (NCDC) besides the students. The schools were all sampled from all over the country to cater for a fair representation of all characteristics of the population.

Multi-stage, random and purposive sampling techniques were used in the study. The first stage was to group the districts according to the 16 zones (NAPE, 2010). One district was randomly selected from each of the 16 zones.

The random sampling was used to select the schools and students who participated in the study and purposive sampling was used to select the specialists of the English Language, French and German languages from UNEB and NCDC.

Ten English Language teachers were purposively selected from each of the sampled districts/division. All these were done to have a fair representation of the population. French and German students were purposively sampled from A’Level class from all over the country. This was on the assumption that these students at this point would have had enough experience that could help them contribute usefully to the topic of discussion. Using the number of candidates that registered for the 2015 examinations at A’Level in these subjects, 10% from each of the total number of candidates who registered were sampled (60 and 20 candidates of German and French respectively) participated in the study. One French specialist and one German specialist were purposively sampled from UNEB and NCDC respectively. The French and German teachers sampled participated in the study to give in their rich experiences in handling oral assessment. Altogether, there were 3474 participants in the study.

The study used focus-group discussion guide, questionnaires and interview schedule to gather data that explained the benefits and challenges of oral assessment and suggested procedures to be followed if oral assessment is to be introduced in English Language at UCE.

The focus group discussion was to gauge the feelings, opinions and thoughts of the students about the benefits and challenges of oral assessment (Linville, Lambert-Shute, Fruhauf & Piercy, 2003, p.211). To maximize the information that were gathered from the discussions, a semi-structured interview method was used. This allowed the researchers to ask questions from a prepared list and for the other part to be determined by the flow of the discussion or the responses of the participants. This

method helped in the exploration and expansion on comments made, issues raised and controversies that arose. The interview schedule, focus group discussion guide and the questionnaires were all circulated to the officers in the research department of UNEB to assess the validity before the final copies were used for data collection.

The validated focus-group discussion and interview checklists and questionnaires were piloted to assess the reliability of the documents. The reliability co-efficient of the structured documents for the participants were calculated using the Cronbach Alpha co-efficient and the results were found to be 0.89.

Documentary search for the procedures involved in the handling of oral assessment in English Language and the other languages of French and German were summarised by the researcher and presented to the language specialists. These summaries were studied and analyzed by the specialists and teachers of English Language sampled to find out which procedure(s) can be adopted for use by UNEB at UCE if oral assessment is to be introduced. What the majority agreed upon with sound basis was proposed for adoption.

The analysis was done qualitatively and quantitatively. Data collected from the questionnaires, interviews and focus group discussions documents were coded, frequencies and percentages developed and summarized to come up with the benefits and challenges. The data from the focus group discussion and the interview were grouped according to the different themes and percentage. The frequent benefits and challenges cited were grouped into themes and presented. The suggested procedures given by the specialists of English Language were taken and presented.

Findings/Results and Discussion

Focus group discussions, questionnaires, and interviews were used to gather information about the benefits that

may be achieved, procedures to be followed, and challenges that may be faced if oral assessment in English Language is to be introduced at UCE.

Focus Group Discussion

The focus group discussion held with the subject experts and the teachers after the review of the documents from the different examining bodies WAEC, (2013); Cambridge, (2012); UNEB, (2008) came up with the following proposed procedures to be followed by UNEB if oral assessment is introduced at UCE.

The learning outcomes to be tested that were suggested by the experts and the teachers were: pronunciation, fluency, conversation, discussion and maturity as done by Cambridge (2012).

The participants agreed during the focus group discussion that oral assessment in English Language should target learners right from Senior 1 up to Senior 4. The type of assessment to be offered was agreed to be both formative and summative. Formative assessment was suggested to be done from Senior 1-3 and during the first and second terms of Senior 4 and summative assessment to be done at the end of Senior 4 with the rest of the examinations.

Activities that were agreed upon by the participants during the focus group discussions to be assessed are: 1. Giving instructions, 2. Describing events, scenes, scenery, occasions etc., 3. Conversations, 4. Debates and discussions, 5. Rephrasing a speech (reporting a speech) events e.g. running commentaries etc.

The actual procedures agreed upon by the participants during the focus group discussions were as follows:

- Selection of appropriate resources for assessment, for example in Senior 1 oral tasks could be about familiar experience like describing an event. This task or exercise should be linked to the next level, for example a conversation developing into a dialogue or a debate.

- Assessment activity should be fitting for individual or group.
- The assessment should be administered by the teacher. Records are to be kept by the school and remitted to UNEB.
- Scoring will be either numerically or descriptive.
- The learners should be able to understand the scoring process.
- Scores/ descriptors obtained from S.1-S.3 could be converted to percentage (%) at S.4.
- Results can be reported separately as oral results or combined with written results.
- Alternatively, and most preferable, the results can be put separately to make it stand out as a competence (practical skills).

Teachers’ Questionnaire

A questionnaire was administered to teachers of English Language to find out the benefits and challenges of conducting oral assessment in English Language at UCE.

The findings from the teachers are shown in Table 1.

Table 1
Oral Assessment Benefits

| No. | Statement | Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|-----|---|----------------|---------------|-------------|-------------|-------------------|
| I | Assessment helps the candidates to acquire the skill of personal presentation | 104 (65.0%) | 40 (25.0%) | 0 (0%) | 0 (%) | 16 (10%) |
| Ii | Oral Assessment in English Language will provide a learning opportunity for the learners at this level | 96 (60%) | 40 (25.0%) | 8 (5.0%) | 16 (10%) | 0 (0%) |
| Iii | The theory learnt in English Language classes will relate to practice during Oral assessment | 112 (70%) | 32 (20%) | 16 (10%) | 0 (0%) | 0 (0%) |
| Iv | By teaching and testing oral comprehension, the students will acquire the linguistic and communication abilities in the Language | 128 (80%) | 24 (15%) | 8 (5%) | 0 (0%) | 0 (0%) |
| V | There will be a positive relationship between oral competence in English Language and academic achievement | 104 (65%) | 40 (20%) | 0 (%) | 24 (15%) | 0 (0%) |
| Vi | Oral assessment given to the students in English Language will give students an opportunity to improve their spoken language abilities in the subject | 152 (95%) | 0 (0%) | 8 (5%) | 0 (0%) | 0 (0%) |

| | | | | | | |
|------|---|-------------|--------------|-------------|-----------|-------------|
| Vii | Oral assessment will also improve on interpersonal relationship between the teachers and students of English Language | 72 (45%) | 56 (35%) | 16 (10%) | 0 (0%) | 16 (10%) |
| Viii | Oral assessment in English Language will help the students to prepare for career development | 96 (60%) | 40 (25%) | 0 (0%) | 0 (0%) | 24 (15%) |
| Ix | Oral assessment will reduce on examination malpractice (plagiarism) | 24 (15%) | 104 (65%) | 16 (10%) | 8 (5%) | 8 (5%) |

Majority (95%) of the respondents agreed that oral assessment will help the candidates to acquire the skills of personal presentation; 85% agreed that oral assessment will provide a learning opportunity for the learners at this level. This is in agreement with Killen (2001) & Joughin and Collom (2003) who observe that students see relevance of theory to practice through oral assessment; ninety percent agreed that theory learnt in the English Language classes relates to practice during oral assessment; Like Penny (2008), 95% of the respondents agreed that testing oral comprehension will encourage students to acquire the linguistic and communication abilities in the language; 85% agreed that there is a positive relationship between oral competence and academic achievement; overwhelming majority of 95% strongly agreed that oral assessment will give an opportunity to the students to improve their spoken language abilities in English Language. This is agreement with Singh (2007) who found out that oral assessment provides an opportunity to the learners to improve their language abilities. Agreeing with Nelson (1993), 80% of the teachers indicated that oral assessment improves on interpersonal relationship with the teachers and the learners; 85% supported the view that oral assessment in English Language will help the students to prepare for career development and finally 80% agreed that oral assessment can reduce on examination malpractice (plagiarism).

Table 2 shows the challenges associated with oral assessment.

Oral Assessment Challenges

| No. | Statement | Strongly Agree | Agree | Neutral | Disagree | Strongly Disagree |
|-----|---|----------------|----------|----------|----------|-------------------|
| i | Reliability of oral assessment (results from oral assessment) is a challenge | 24 (15%) | 56 (35%) | 40 (25%) | 40 (25%) | 0 (0%) |
| ii | Oral assessment is a true assessment of spoken abilities | 64 (40%) | 88 (55%) | 8 (5%) | 0 (0%) | 8 (5%) |
| iii | The student's background characteristics influence the assessor's (examiner's) attitude towards students from other backgrounds | 8 (5%) | 24 (15%) | 64 (40%) | 32 (20%) | 32 (20%) |
| iv | The examiner's background characteristics influence the assessor's (examiner's) attitude towards students from other backgrounds. | 0 (0%) | 24 (15%) | 72 (45%) | 32 (20%) | 32 (20%) |
| v | Time affects how candidates answer questions during the oral assessment | 16 (10%) | 24 (15%) | 8 (5%) | 80 (50%) | 32 (20%) |
| vi | The number of candidates adversely affect oral assessment | 160 (100%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) |
| Vi | Oral assessment is very costly | 144 (90%) | 16 (10%) | 0 (%) | 0 (0%) | 0 (0%) |

As shown in Table 2, the teachers (160) 100% ranked the number of candidates as the probable highest challenge to be faced followed by the high costs (144) 90% involved if oral assessment in English Language is introduced at UCE.

Interview

Interview was done with subject experts who included subject officers at UNEB and NCDC. What was brought out during the interviews were that the theory learnt during class can relate to practice during oral assessment. This is as found out by Singh (2007) that oral assessment provides students with an opportunity to improve their language abilities. It was also clearly brought out that oral

assessment gives the students an opportunity to improve their spoken language abilities.

On the other hand, the main concerns given were the high costs involved in the administration of oral exams. This is in agreement with the findings of Surface, Poncheri & Bhavsar (2008) who noted that oral assessment is very expensive by its nature. The examiners have to be facilitated to move and stay in the field visiting schools around the country, which requires a lot of money. It was also noted that there seemed to be lack of genuine (reliable) examiners because of the subjectivity of the oral assessment process. This is in agreement with the findings of Kotitse (2009), Abioye (2010) & wood (1991). Most outstandingly was the challenge of the high

number of candidates to be assessed orally. All the subject experts and teachers interviewed (166) 100% unanimously said that the number of students in the English Language will pose a great challenge in handling oral assessment.

Conclusion/ Recommendations/Implications

The benefits outweigh the challenges, hence it can be concluded that it is worthwhile having oral assessment in English Language.

As suggested by the subject experts, one way of overcoming the challenges of oral assessment without losing the benefits could be the use of a computer to present the questions to the candidates. Candidates could answer the questions on a computer and the computer could prompt the candidates according to their needs. This will save on time because many candidates will be simultaneously examined since the human element is removed.

There will be need for change in the format of assessing English Language. All the four skills of languages (listening, reading, speaking and writing) will have to be assessed unlike in the past where only reading and writing were assessed leaving out the listening and speaking skills.

The assessment of oral competence in English Language will bind with the assessment of functional skills in the current English Language Paper 1 and bring out full tests of competence in the subject.

The teachers of English Language will have to be re-tooled in the teaching and assessment of oral skills at UCE level.

References

Abioye, T. (2010). *Managing language testing in Nigerian large classes: Processes and prospects*. www.ccsenet.org/elt.

Aysha, A., Alastair, P., Leslie, R. (1999). *Assessing thinking and understanding: Can Oral assessments provide a clearer perspective?* A paper presented at the 8th International Conference, Edmonton Canada, July 1999.

Cheng, L. & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In Cheng, L., Watanabe, Y., & Curtis, A. (eds), *Washback in Language Testing: Research contexts and Methods*, pp. 3-17. Mahwah, NJ: Lawrence Erlbaum Associates.

Gracia G. E. (2002). *Oral English Proficiency: Issues and Questions*, In Reading instruction of English Language Learners.

Joughin, G. (2010). A short guide to oral assessment. Version 2. 17 February 2010, *CEDIR*, University of Wollonngong, Leeds, Metropolitan University, pp.1-22.

Joughin, G. (2008). *Student conception of oral assessment*. 32,323-336.

Joughin, G. and Collom G. (2003) Oral Assessment. The Higher Education Academy.

http://www.heacademy.ac.uk/resources/detail/resource_database/id433_oral_assessment. retrieved: August 30, 2011.

Joughin, G. (1999). Dimensions of oral assessment and students' approaches to learning. In Brown, S. & Glasmer, A. (eds). *Assessment matters*. Buckingham: the society for Research into Higher Education and Open University Press, 146-156.

Killen, R. (2001). *Productive teaching and learning*. Paper delivered at the eighth International; and Educational research Network conference on Learning, Spetses, Greece.

- Kofitse, A. (2008). *Bring Back Oral English: A Case for the Re-introduction of Oral English at the WASSCE Level*. A paper presented at the Annual conference of AEAA, Accra, Ghana.
- Kofitse, A. (2009). *Assessment of Oral English by the West African Examinations Council (WAEC): Challenges and Prospects*. A publication of the Association for Educational Assessment in Africa, Accra, Ghana.
- Lazaraton, A. (2002). A Qualitative Approach to the Validation of oral Language Tests. www.cambridge.org.
- Lineville, D., Lambert-Shute, J., Fruhauf, C. A., & Piercy, F. P. (2003). Using participatory focus groups of graduate students to improve academic departments: A case example. *The Qualitative Report*, 8 (2), 210-223. Retrieved from <http://www.nova.edu/ssss/QR/QR8-2lineville.pdf>
- No Child Left Behind (2002). *The No Child Left Behind Act of 2001*.
- Nelson, J. (1986). Implementing oral exams as part of the school exam system. In *New approaches in the language classroom: coping with change*. Proceedings of the National Modern Language convention, 2nd Dublin, Ireland: Dublin University Trinity College, 31 January-1 February, pp. 70-78.
- Nitko, A. J. & Brookhart, S. M. (2007). *Educational Assessment of Students*. Prentice Hall. New Jersey, Columbus, Ohio, 5th Edition.
- Rust, C. (2001). "A Briefing of Assessment of Large Groups". *LTSN Generic Centre Assessment Series* 12.
- UNEB (2008). Uganda Advanced Certificate of Education. *Regulations and syllabuses 2009-2013*. Published by the Uganda National Examinations Board.
- UNEB (2012). Uganda Certificate of Education. *Regulations and syllabuses 2012-2016*. Published by the Uganda National Examinations Board.
- Utley, D., Mitchell, R. & Phillips, J. A. (1983). Hear/Say: A review of oral/aural graded tests. *Schools Council Examinations Bulletin*, 44
- Wall, A. L. (2001). Evaluating an undergraduate unit using a focus group. *Quality Assurance in Education*, 9 (1), 23-31.
- World Bank (2008). Educational tests and measurements. *Assessing National Achievement levels in Education*. The International Bank for Reconstruction and Development. The World Bank, Washington, DC.

Considering Authenticity of Skills Assessments

Helyn Kim, PhD

hkim@brookings.edu and

Esther Care, Ph.D

ecare@brookings.edu

Contact Information: 1.202.238.3156

Abstract

Recently, there has been increased demand for skills that extend beyond traditional academic ones to include skills such as collaboration and critical thinking. However, for assessment purposes there is less knowledge about how to capture what students can do, than what they know. Considering authenticity of assessment tasks can provide a framework for how assessments can be adapted or created to foster learning and capture the processes that students will need to apply in their future. This paper examines the authenticity of a computer-based assessment of collaborative problem solving task and considers how it can be adapted for classroom purposes. Challenges to adapting and translating a computer-based task into one that could be implemented in the classroom, as well as implications for designing authentic assessments, are discussed.

Keywords: authenticity, collaborative problem solving, classroom assessment, computer-based assessment

Considering Authenticity of Skills Assessments

Over the last several decades, there have been calls from global organizations to re-think education beyond acquiring content and knowledge to include a wider range of skills for learning, work, and life. For instance, UNESCO proposed a vision of education as “learning to do, learning to be, learning to live together” (Delors et al., 1996), followed by Organisation for Economic Co-operation and Development’s (OECD) interest in focusing on key competencies that were categorized into three broad groups: using a wide range of tools for interacting effectively, engaging with others, and taking responsibility for managing their own lives (Rychen&Salganik, 2003). Similarly, the Assessment and Teaching of 21st Century Skills (ATC21S; Griffin, McGaw & Care, 2012) project proposed four components, including ways of thinking, ways of working, tools for working, and skills for living. In other

words, there is clear support and increased demand for skills that extend beyond traditional academic ones to include 21st century skills, such as collaboration and critical thinking (Care & Anderson, 2016; Care, Anderson, & Kim, 2016). Nevertheless, there is less clarity on how to develop a broad range of skills in students, and notably, how to assess these skills to capture the underlying cognitive and social *processes*, rather than just mastery of content knowledge.

Considering authenticity of assessment (Gulikers, Bastiaens, &Kirschner, 2004) - including whether the assessment task, the physical context, the social context, the assessment result, and the assessment criteria, resemble the complexity of real-life situations - can provide a framework for how assessments can be created in order to capture the processes that students will need to apply in the future (Ashford-Rowe, Herrington, & Brown, 2014; Mohamed &Lebar, 2017; Wiggins, 1993)

as well as provide information that can support skills development. The objectives of this paper are twofold: 1) examine a computer-based assessment of collaborative problem solving, namely PISA 2015 Collaborative Problem Solving (OECD, 2017), in terms of its authenticity; and 2) consider how the PISA collaborative problem solving assessment can be adapted for classroom purposes. Implications for using the framework to design authentic assessments to capture the underlying processes are discussed. This particular skill has been selected to demonstrate the approach, because it combines many of the subskills typically identified as important for students in classrooms globally (National Research Council, 2011).

Authenticity of Assessments

The traditional education assessment method of paper and pencil, while suitable for evaluating fact-based knowledge, is insufficient for measuring progress and mastery of 21st century skills. For example, a basic numeracy question, such as $15+10$, has a correct answer. However, inquiry-based questions, like “what is the problem to be solved and how can it be approached?” are much more difficult to assess using traditional methods because there are multiple ways to answer the question. Skills and processes that are essential in the real world are even more complex to assess because not only is a simple yes or no answer insufficient, what is required is demonstration of application of these to novel situations. For instance, collaborative problem solving requires integrating cognitive and social skills to navigate not only the problem at hand, but also interactions with others. These are the transferable processes that students need to succeed; and therefore, these are the processes that need to be captured for useful assessment which can inform learning and teaching goals.

One way of aligning assessment to current educational goals is to increase the authenticity of assessments. Authenticity is a concept used to describe assessments tasks that are analogous to real-world tasks, in which

students meaningfully and effectively apply specific knowledge, skills, processes, and competencies to solve problems (Frey, 2014; Gielen, Dochy, & Dierick, 2003; Gulikers et al., 2004; Mueller, 2005; Whitlock & Nanavati, 2013; Wiggins, 1993). In other words, how “authentic” an assessment is has to do with the degree to which it resembles the criterion situation along a number of dimensions and provides information about what students are able to do (Whitlock & Nanavati, 2013). This paper uses the five-dimensional framework for authentic assessment based on an extensive literature study by Gulikers and colleagues (2004). The five dimensions, upon which a task can vary in degree of authenticity, include the assessment task itself, the physical context, the social context, the assessment result or form, and the assessment criteria.

Assessment tasks that are considered to be authentic have students carry out activities and processes that would appear in real life applications. Although the assessment task may be authentic, the *physical context*, or where the student is assessed, may not necessarily be authentic, which could mean that students are not truly using their competencies as they would in real-life situations. An assessment that examines a student’s ability to change a tyre in a clean, auto mechanic’s garage with all of the “right” equipment may not necessarily resemble a situation that a student might encounter in the real-world—such as changing a tyre in the pouring rain, in the middle of a dark highway with cars speeding by, and with less equipment—which means that all of the competencies that are necessary for this real-life situation will not be assessed.

Along with the physical context, the *social context* affects the authenticity of an assessment. Authentic assessments should include the social processes that are associated with the equivalent real-life situations. This does not mean that all authentic assessments include a social element but that this should be part of the assessment if the actual situation involves some form of the social. In the case that the real-life situation is normally conducted

individually, then the assessment should also be completed by an individual (Gulikers et al., 2004). Once the assessment is completed, there is typically an *assessment result or form* that is characterized by a product or performance that is analogous to a real life output, and allows for the demonstration of the wide range of underlying competencies through combinations of multiple tasks and indicators, rather than a single test, to draw fair conclusions about the student's abilities (Darling-Hammond & Snyder, 2000). Students should be able to present and defend their work to show mastery of the competencies (Wiggins, 1989). These results are then evaluated against some forms of *assessment criteria*, which should be made explicit to students prior to the start of the assessment as a way to guide their learning.

Taken together, the five dimensions of authentic assessment provide guidelines for considering whether an assessment is aligned with what the student would encounter in the real world, in terms of the skills and knowledge necessary to complete the task, and within the same kind of physical and social contexts.

Method and Results

The two aims of the current paper are to examine the authenticity of a collaborative problem solving task from the PISA 2015 Collaborative Problem Solving assessment (OECD, 2017) and adapt the task for classroom use.

First, in order to examine degree of authenticity of assessment, the PISA 2015 Collaborative Problem Solving (OECD, 2017) is considered as an example. One particular unit, Class Logo, within PISA 2015 CPS assessment was examined and analyzed against the five dimensions of authentic assessments as defined by Gulikers and colleagues (2004). Table 1 provides a summary of how the Class Logo was viewed through the authenticity dimensions. Descriptions of the Class Logo unit, as well as the degree of authenticity of the tasks are provided in the following section.

Table 1. PISA 2015 Collaborative Problem Solving Assessment and Dimensions of Authenticity

| Dimensions of Authenticity¹ | Description¹ | PISA 2015 Collaborative Problem Solving Assessment² |
|---|---|---|
| Assessment task | Presents students with activities and tasks that are also carried out in the real world; resembles the criterion task with respect to the integration of knowledge, skills, and attitudes, its complexity and its ownership. Students should perceive the tasks as representative, relevant, and meaningful; and have ownership of the tasks. | Tasks require student and two agents (Sarah and Mark) to work together to produce a class logo for a sport event. The underlying processes of discovering, communicating, monitoring, repairing, and building a shared understanding are needed to solve the problem. Throughout the task, students are leading the group in reaching a common goal. |
| Physical or Virtual Context | Reflects the way knowledge, skills, and attitudes will be used in professional practice. Fidelity in terms of how closely a simulation imitates reality. Number and kinds of resources should also resemble those available in criterion situation. Amount of time should be more like real life rather than constrained. | Tasks are computer-based and allow the student to manipulate a limited number of variables within the problem space. Class Logo unit has a time constraint of 20 minutes. |
| Social Context | Social processes of the assessment resemble the social processes in an equivalent situation in reality. | Students interact with two computer agents that simulates various scenarios and is scaffolded such that if the 'correct' response is not given, the agent provides multiple prompts of which the student can choose to help direct the conversation. Students engage in social processes necessary to help the agents collaborate with each other to complete the task. |

¹Adapted from Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52, 67-86.

²Based on Class Logo unit (p. 64-81) in *PISA 2017 Collaborative Problem Solving Framework*. Retrieved from:

<http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>

| Dimensions of Authenticity¹ | Description¹ | PISA 2015 Collaborative Problem Solving Assessment² |
|---|--|---|
| Assessment Result | Result is related to the kind and amount of output of the assessment task, independent of the content of the assessment. Characterized by four elements: (1) quality product or performance that students can be asked to produce in real life; (2) demonstration that permits making valid inferences about the underlying competencies; (3) full array of tasks and multiple indicators of learning in order to come to fair conclusions; and (4) present their work to other people, to defend their work to ensure that their apparent mastery is genuine. | Scores are based on 26 multiple choice chat responses and are given a score ranging from 0-2 points for each item for skills, such as monitoring results of action, evaluating success in solving the problem, and communicating with team members about the actions to be performed. Students also receive scores on the number of attempts needed for classmates to rate the logo with 5 stars. Fewer attempts result in higher scores (0-2). |
| Criteria | Criteria (characteristics of the result that are valued) and standards (level of performance expected) should be set and made explicit and transparent to learners beforehand. Criteria should be related to a realistic outcome; criteria and standards should concern the development of relevant professional competencies and should be based on criteria used in real-life situation. | Items are scored according to levels of the competency, which can receive partial or no credit, and aggregated to provide an overall scale for CPS competency. This overall proficiency score is then mapped onto a four-level proficiency scale based on progressions. |

Second, in addition to examining the degree of authenticity of the PISA 2015 CPS assessment along the five dimensions of assessment task, physical context, social context, assessment result, and criteria, we considered how this collaborative problem solving assessment, specifically the tasks involved in the Class Logo unit, could be adapted for use in the classroom, regardless of whether the tasks themselves are considered to be high in authenticity. There are seven assessment tasks in the Class Logo unit. Each task, as well as the response options and scoring procedures, were examined and considered for feasibility in the classroom setting. The elements of the assessment tasks were kept as similar as possible to the tasks as designed by PISA 2015. For example, the wording of the problem, as well as the expected outcome of each task item is the same. However, adaptations were made to consider the classroom context. For instance, the original task has a student collaborating with two computer agents to produce a logo for a sport event that receives a five-star rating. The same goal and sub-tasks are in place for the classroom activity, but instead of computer agents, the student interacts with other students in the classroom. Figure 1 shows the adapted assessment tasks.

The Degree of Authenticity: PISA 2015 Collaborative Problem Solving

Collaborative problem solving (CPS) is a critical skill for the 21st century workforce, and according to PISA 2015, is “the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution” (OECD, 2017, p. 6). PISA defines three core collaborative problem solving competencies: 1) establishing and maintaining shared understanding; 2) taking appropriate action to solve the problem; and 3) establishing and maintaining team organization (OECD, 2017, p. 9). These major competencies are crossed with found individual problem solving processes (exploring and understanding;

representing and formulating; planning and executing; and monitoring and reflecting) that results in 12 specific skills “associated with actions, communications, cognitive and social processes, and strategies that define what it means for the student to be proficient” in collaborative problem solving (Graesser et al., 2018). This definition of CPS is useful in thinking about the structure and context of the assessment, as well as the processes that are being assessed. The PISA 2015 CPS assessment is discussed in relation to the five dimensions of authenticity (Gulikers et al., 2004).

The PISA 2015 CPS assessment assesses an individual’s ability to solve problems in a collaborative context. (OECD, 2017; Rosen & Foltz, 2014). The PISA 2015 CPS embeds the assessment of collaborative problem solving in a well-scaffolded environment, where the student collaborates with a fully controlled computer agent, rather than working with other students. One example is a unit called Class Logo. Here the student “collaborates” with Sarah and Mark (computer agents) to create a five-star rated logo for a sporting event (OECD, 2017). The student is tasked with leading the group, and Sarah and Mark have to draw the logo.

There are a total of seven *assessment tasks* within the Class Logo unit. The student participates in initial

discussions with the agents on how the logo should be designed, encounters challenges in collaborating with Mark and Sarah during the process of producing drafts of the logos, and provides feedback on the tasks. There is therefore ownership of the task, as the student participates in the process of reaching the goal. However, the task items are a series of multiple choice questions, from the initial introduction of oneself to providing feedback to the agents. None of the task items requires the student to generate his own response, raising the question as to whether the underlying cognitive and social processes are at play. Nevertheless, the response choices throughout the tasks can provide some insight into the proficiency level of the student (Graesser et al., 2018). For instance, if a student tends to answer

questions but does not ask questions that may mean that the student meets minimal standards of proficiency, whereas a student who tends to ask questions and initiates requests may be above the standard level of proficiency. In addition, given that the student is tasked with leading the group, rather than participating in producing the logo alongside Mark and Sarah, whether the task resembles the complexity of a real-life collaborative problem solving task or even engaging in collaborative problem solving, as defined by PISA, is at question.

The *physical context* of the task takes place within a computer-based program, rather than taking place within the classroom, despite the fact that the task is categorized as an “in school” context task. There are some issues related to fidelity and the presentation of material and resources that are available within the problem space. Although there is no time limit to explore the task space, a five-star rated class logo has to be designed within 20 minutes and is limited to only five trials, which may be unrealistic compared to what the student may encounter in the real world.

Regarding the *social context*, a student “interacts” with two computer agents that simulate various scenarios. Mark is considered to be the collaboratively-oriented agent, where he supports and praises the student, builds consensus, and provides ideas. On the other hand, Sarah

is designed to act as the low collaboratively-oriented agent, who interrupts and disagrees with the group members, makes negative comments, and does not follow the agreed upon plans. In the process, the student actually selects from multiple choice options. The student is scaffolded such that if the ‘correct’ response is not given, the agent provides the response re-worded - which provides the facility for the student to continue. Although the social processes that exist in the real world are at play during the task, such as establishing a shared understanding of the problem, repairing misunderstandings, and consensus building, these processes are being scaffolded by the computer program,

since the student is interacting with agents, rather than actual people.

The *assessment result* is based on multiple choice artificial chat responses. The results are aggregated to provide an overall scale for CPS competency. This overall proficiency score is then mapped onto *standards*, which for this assessment is a proficiency scale based on where the student is on a progression or scale of CPS. There are three levels of proficiency: below standard of proficiency; at standard of proficiency; and above standard of proficiency (Graesser et al., 2018). These proficiency levels are not explicitly made known to students before beginning the task; however, the descriptions of the levels of proficiency provide clear standards and expectations that could reasonably be based on real-life situations.

Authentic Assessment of Collaborative Problem Solving in the Classroom

The adaptation of the Class Logo unit for the classroom (Figure 1), was designed to parallel the original PISA 2015 task as much as possible (OECD, 2017). Despite similarities, the adapted version deviated from the original assessment in two ways due to feasibility issues in the classroom. First, the level of scaffolding provided in the computer-based PISA assessment is not possible because the student is interacting with other students, rather than within a controlled, structured system where the other group members’ abilities influence what goes on

in the group. In the computerized version, the two agents’ responses and behaviors are scripted and therefore invariable within response, so that certain behaviors are elicited by the student; however, this is not possible within the classroom where students interact with each other dynamically and in real-time. Second, in the original PISA task, the students were given 20 minutes; however, this time frame is not feasible for students to accomplish all of the components of the

task in real time. Therefore, this task would need to be adapted either by extending the time frame, reducing the number of logos to create, and/or simplifying the scoring process.

Albeit these limitations, adaptation for the classroom increases the authenticity of the assessment, especially in the dimensions of the physical and social contexts.



Figure 1. Adapted Class Logo Task for Classroom Use

Class Logo

Your school is holding a sports competition, and your class has been asked to help with the preparations. You will get into teams of three for this task. Your team must design a logo to be used on posters advertising the event, and some of your classmates outside of your team will rate your designs. Your goal is to achieve a 5 star rating for your design. You will have 5 attempts to achieve this 5-star rating.

Follow the steps below:

- Step 1: Everyone in the group describes their abilities in logo design. Based on this, discuss and determine which 2 team members will draw and which team member will lead the group.
- Step 2: Read the background information about the sports competition below; team members should come to a shared understanding of the problem.

Sports competition information
When: Summer **Where:** Park **What:** Running, Soccer, Tennis
Logo criteria: Colorful, simple, not used before
Previous logos:  

- Step 3: Begin discussing a plan of how to design a logo (e.g., leader provides suggestions on the symbols and colors to be used). Each group member should share their own point of view and provide reasoning for their ideas and thinking.
- Step 4: Each designer will draw a logo on their own to present to the group.
- Step 5: The team will need to come to a consensus about which logo will be sent to classmates for feedback. If there is some disagreement, the team leader should urge designers to provide reasoning, so that a consensus can be reached prior to submitting the logo to be rated by the class.
- Step 6: Once a rating and feedback has been received, the team leader should review with the rest of the group members and elicit responses regarding their interpretation of the feedback, as well as repair any misunderstandings, before deciding if the team will choose to edit the logo that they have already received feedback on or if the team will choose the other logo that hasn't received any feedback to submit. Also, the team will need to reach a consensus on which elements of the logo should be adapted.
- Step 7: Team leader should communicate with the team members about the actions to be performed before submitting again for feedback.
- Step 8: Repeat steps 6 and 7 until a five-star rating has been achieved. There are only five opportunities to receive a five-star rating.

Note. Adapted from Class Logo unit (p. 64-81) in *PISA 2017 Collaborative Problem Solving Framework*. Retrieved from:

<http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>

Discussion

Over the years, there has been strong support for the importance of developing a broad range of skills, beyond traditional academic subjects. But, how to assess the processes involved in these skills remains elusive. As the goals of education change, assessment paradigms must also “shift from conventional assessment to authentic assessment” (Koh, Tan & Ng, 2012, p. 137), in order to capture the intricacies of real-world scenarios and develop the skills needed to succeed in school and in life. The purpose of this paper was to examine and describe an existing computer-based assessment of a 21st century skill, specifically PISA 2015 Collaborative Problem Solving (CPS), along five dimensions of authenticity (Gulikers et al., 2004) and adapt it for the classroom.

The PISA 2015 CPS assessment took the approach that collaborative problem solving can be seen as a construct comprised of two major components - a cognitive, problem solving component, and a social, collaborative component. There was variability in the degree of authenticity across the five dimensions, which may be attributable to the goal of the assessment—to measure individual abilities within collaborative contexts, rather than capturing collaborative problem solving competencies more dynamically. One thing to note is that the five dimensional framework of authenticity (Gulikers et al., 2004) discussed in this paper is just one way of examining authenticity of assessments. Therefore, these five dimensions should not be seen as absolute, but rather one framework that can help structure the design of an assessment.

Challenges to adapting the task.

As demonstrated during the adaptation process, there were issues that arose with translating a computer-based

task into one that could be implemented in the classroom. Some of these challenges are due to the nature of the task itself, while others have to do more broadly with the issue of authenticity. For example, one challenge has to do with the degree to which success of an individual student is

dependent on the other students who are in the group. This may complicate the assessment of the student’s level of CPS because it may be confounded by either the task at hand or the group in which they were placed. Moreover, no two groups will have the same set of exchanges, skills sets, or conversations, adding to the complications regarding classroom-based authentic assessments. Therefore, standardized processes of scoring and cross student comparisons become issues. In the computer-based CPS assessment, scores are automatically calculated based on the multiple choice responses; it is difficult to adapt this system for the classroom. Although a list of the processes that should be captured can be provided, how to monitor each student’s collaborative problem solving processes in real time in the classroom poses another set of challenges. PISA 2015 defines CPS as establishing and maintaining shared understanding, taking appropriate actions to solve the problem, and establishing and maintaining team organization. However, without documenting all conversation, it may be difficult to capture the presence or absence of these processes.

Computer-based assessments can provide varied and complex stimulus scenarios for students to engage with. It may be that adaptations of these assessments for the classroom present impractical burdens on the teacher. The computer-based assessments can both create obstacles for students to overcome through the task, in order to elicit certain processes, as well as scaffolds for students unable

to move forward in the task. Emulating this degree of complexity in a controlled manner in the classroom,

where teachers need to deal with many students and of different abilities, may be impractical in many classrooms.

Ways to address issues and increase authenticity.

Along with specific modifications of the task, examining the degrees in the five dimensions of authenticity within the PISA 2015 CPS assessment task provides insights as to what needs to be considered when developing authentic assessments more broadly. First, the educational goal matters—if the goal is to create students who are successful professionals, increasing the authenticity to better reflect real-life situations can be valuable, and not only for student learning. Authentic assessment practices in teacher education can affect how teachers teach; therefore, authentic assessments can act as a tool in teacher education programs to prepare novice teachers (Darling-Hammond & Snyder, 2000).

Although not discussed in relation to the five dimensions, the educational or proficiency level of students should be taken into consideration when designing assessments, particularly in the classroom, in order to be able to capture a wide range of abilities. In order to do so, a strong understanding of how a particular construct develops from less complex to more sophisticated forms, as well as the underlying components and processes, is essential. Moreover, the focus should be on the process, rather than the product or outcome. In the example task above, although the end goal is to create a logo, the students are engaging in the cognitive and social processes throughout the entire task, and how they engage in these processes is what needs to be assessed, in order to determine whether students are able to collaboratively solve problems.

By adapting a computer-based assessment of CPS for the classroom, it is clear that there are advantages and disadvantages with regards to authenticity for each medium. The computer-based assessment allows for progress monitoring and immediate feedback, as well as

a more controlled medium with built-in scaffolding and as a result, the scoring process was standardized. While the classroom-based assessment cannot feasibly incorporate

these elements, it offers other advantages that are not available within the computer-based system, such as the authenticity that may better mirror the skills and processes that students will need to succeed in the real-world.

References

Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education, 39*, 205-222.

Care, E., & Anderson, K. (2016, May). *How education systems approach breadth of skills*. Washington, D.C.: Center for Universal Education, Brookings Institution.

Care, E., Anderson, K., & Kim, H. (2016, September). *Visualizing the breadth of skills movement across education systems*. Washington, D.C.: Center for Universal Education, Brookings Institution.

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education, 16*, 13-37.

Frey, B. B. (2014). *Modern classroom assessment*. Thousand Oaks, CA: SAGE Publications.

Delors et al. (1996). *Treasure within: Report to UNESCO of the International Commission on Education for the Twenty-first Century*. Paris, France: United Nations.

Gielen, S., Dochy, F., & Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post-, and true assessment effects. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of*

- assessment: *In search of qualities and standards* (pp. 37-54). Springer Netherlands.
- Graesser, A. R., Foltz, P. W., Rosen, Y., Shaffer, D. W., Forsyth, C., & Germany, M. (2018). Challenges of assessing collaborative problem solving. In E. Care, P. Griffin, & M. Wilson (Eds.), *Assessment and Teaching of 21st Century Skills: Research and Applications* (pp. 75-91). Springer.
- Griffin, P., McGaw, B. & Care, E., (Eds.) (2012). *Assessment and Teaching of 21st Century Skills*, Dordrecht: Springer.
- Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52, 67-86.
- Koh, J. H., Tan, C., & Ng, P. T. (2012). Creating thinking schools through authentic assessment: The case in Singapore. *Educational Assessment Evaluation and Accreditation*, 24, 135-149.
- Mohamed, R., & Lebar, O. (2017). Authentic assessment in assessing higher order thinking skills. *International Journal of Academic Research in Business and Social Sciences*, 7, 466-476.
- Mueller, J. (2005). The authentic assessment toolbox: Enhancing student learning through online faculty development. *Journal of Online Learning and Teaching*, 1, 1-7.
- National Research Council. (2011). *Assessing 21st century skills*. Washington, DC: National Academies Press.
- OECD. (2017). *PISA 2017 Collaborative Problem Solving Framework*. Retrieved from: <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>
- Rosen, Y., & Foltz, P. W. (2014). Assessing collaborative problem solving through automated technologies. *Research and Practice in Technology Enhanced Learning*, 9, 389-410.
- Rychen, D. S., & Salganik, L. H. (eds.) (2003). *The Definition and Selection of Key Competencies (DeSeCo): Key competencies for a successful life and a well-functioning society*. Gottingen, Germany: Hogrefe & Huber.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessments. *Phi Delta Kappan*, 70, 703-713.
- Wiggins, G. (1993). *Assessing student performance—Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass.
- Whitlock, B., & Nanavati, J. (2013). A systematic approach to performative and authentic assessment. *Reference Services Review*, 41, 32-48.

Challenges of Setting and Conducting Practical Skills Assessment in Musical Arts Education in Zambia

Boscow Mubita

bmubita@exams-counci.org.zm/kumbonja63@gmail.com

Zambia

ABSTRACT

Musical arts have been assessed for many years in Zambian schools and colleges of education by the Examinations Council of Zambia. In spite of this, the setting and conducting of practical skills assessment in musical arts are still not without challenges especially in the area of practical skills that express African sensibilities, bearing in mind that Musical Arts Education falls in the category of cultural subjects. This paper reports on an investigation whose purpose was to explore the teaching approaches and models of assessment used in practical skills [listening, composing and performing] in schools and teacher education institutions, and the competence of the teachers and lecturers in assessing practical skills from the perspective of indigenous musical arts. The study involved two teacher education institutions and four schools. It qualitatively investigated the orientation of teachers of musical arts with regard to Eurocentrism and Afrocentrism. The paper discusses the results on the teaching of practical skills, the Eurocentric and Afrocentric inclinations of the musical arts teachers and lecturers, and looks at the impact of such dispositions with regard to their competency as assessors of indigenous musical skills. It also touches on the availability of the indigenous musical materials in the teaching and assessment of practical skills in schools and teacher education institutions. It highlights the pros and cons of objective and subjective assessment and the employment of community based connoisseurs in the quest to enhance the validity of formative assessment of practical skills in indigenous musical arts. Challenges in the assessment of the concept of timbre through the sole use of either Western or African musical instruments are commented on in the light of the diversity of candidates' musical experiences in the urban and rural environments.

Key words: Assessment, Practical skills, Eurocentric and Afrocentric musical arts, connoisseurs, concept of timbre

Introduction

Assessment as an activity of evaluating learners' achievements has remained indispensable to the teaching and learning processes. Like in any other learning area, it has been used over time to ascertain learners' progress in the practical skills in Musical Arts Education. Essentially, musical arts are listening, composing and performing [CDC, 2014]. The Musical Arts Education Assessment Schemes, which are an interpretation of the Musical Arts Education syllabi based on the revised curriculum, emphasize, at both the Junior Secondary and Senior Secondary school levels, the assessment of the three

practical areas: listening, composing and performing [ECZ 2014, 2016].

At junior level, listening is assessed through an audio compact disc. Candidates are also required to compose a piece of music which they may choose to write down or record on an audio compact disc or perform. The composition theme is based on a cross cutting issue like gender-based violence. In addition to this, candidates are expected to sight-read a phrase, sing a song, play a musical instrument or perform a dance.

At senior level, besides the identification and transcription of musical elements, candidates are required to analyse a Western art and an African indigenous or contemporary music aurally perceived after having been given music scores to study. The paper on composing demands the application of compositional and arranging skills. A project is also given for candidates to compose a piece of music of a specific length which should be written down or recorded on an audio compact disc. The performing paper includes sight-reading as a compulsory test, and the playing of an instrument, singing or dancing. Apart from sight-reading which is exclusively an individual candidate's assessment, the rest are done as solo as well as group performances.

However, to have a comprehensive picture with regard to assessment of practical skills in musical arts in Zambia, the teaching approaches employed should also be appreciated. To contextualize the setting and conducting of assessment of practical skills in musical arts, the national educational aspirations should be brought in the fore. Therefore for the setting and assessment of practical skills in musical arts to be of relevance to the country, it goes without saying that they should be done in a way that focuses on achieving the national goals of the education system. Regarding the arts, the Ministry of Education has expressed the following:

'The performing and creative arts have a twofold role to play in the education of young (*people*). According them (*the arts*) some importance in the school curriculum promotes wider knowledge of deeper appreciation for Zambia's rich cultural heritage and thereby contribute to the preservation and development of this heritage' [MoE, 1996:38-39].

To consolidate this aspiration, the Ministry of Education has set for itself a serious undertaking to produce

'a learner capable of appreciating Zambia's ethnic cultures, customs and traditions, and upholding national pride, sovereignty,...' [MoE, 1996:5].

The assessment of practical skills in musical arts should therefore be evaluated against the task specified by the Ministry of Education in the education policy, namely to increase

'access to education and life skills training' [MoE, 1996:5].

Tertiary education for music teachers in Zambia has been observed over the years to emphasize the Eurocentric conceptualization of musical elements. The Afrocentric drive has been there but to a lesser degree. Therefore, music education as conducted in schools, and not necessarily as written down at policy level, has perpetuated the Western representations of music concepts in terms of content, method, philosophy, psychology and outcome [Mwesa in Herbst, 2005]. This disposition trickled down into schools where pupils are systematically exposed to Western music and not to their own community music. This has led to the situation where musical skills in Western music have been associated with educational prestige in some sections of Zambian society, especially in the urban areas. In this case, some learners have opted, when they get to school, to learn some music from the West and consequently shun their own indigenous musical arts. This undermines the promotion of cultural values that are conveyed by Zambia's cultural practices as suggested in the education policy.

Zambia as a nation is characterised by ethno-cultural pluralism which, from the angle of ethnomusicology, expresses musical heterogeneity. This means that learners are in contact with various types and styles of music found in their location, different from the others. It is also true that the learner, especially in the rural areas, experiences some form of musical dichotomy in terms of school and community music. Kabede [1995:13] brings

out what can be appreciated as an advantage to learners in the urban areas, where they experience more or less the same

‘popular or contemporary music or urban music which blends well with interethnic and international style’.

In addition, while the learners in the urban set-up have an opportunity to listen to more of Western music and practice more on Western musical instruments whereas the learners in the rural areas experience, to a larger degree, the sounds of the traditional musical instruments in their location on which they play music. In this sonic milieu, it is not easy to set a listening examination, especially on elements like *timbre* [tone colour] for the whole nation without running the risk of disadvantaging some candidates. Commenting on this, Mubita... [*et al*] in Herbst [2005:176] observed that music should not be set in a way that disadvantages the candidates in the rural areas. There should be a balance between Western and indigenous music. It stands out that the assessment of listening skills in the testing of *timbre* cannot be done in a homogenous way, bearing in mind the sundry nature of traditional musical instruments in the communities.

In its context, the Education Curriculum Framework encourages subjects that provide hands-on experience to the learner [2013: iv]. Musical Arts Education fits well in this setting. This suggests that the models of assessment in the musical arts should contribute to a systematic creation of a real-life environment which should ultimately contribute to the exertion of educational influence on the learner, creating a mind-set that will enable the learner to be creative and productive in later life, and eke out a livelihood through musical entrepreneurship and productivity [CDC, 2013: viii, x]. On this aspect, the position of the Ministry of Education is very clear. The purpose of education, which includes musical arts, is to improve human lives through the broadening of learner capabilities which can ultimately be put to profitable use in the economic, social, cultural,

scientific and political fields [MoE, 1996:91,95]. From this view point, it becomes evident that the assessment of listening skills in musical arts in Zambia still has challenges in satisfying its educational mandate of equipping the learner with wide-ranging and realistic listening experiences. The assessment of aural skills therefore needs to be contextualized.

The Eurocentric approach to the teaching of musical arts does not rhyme with the Afrocentric version in more ways than one. Whereas the philosophy of musical arts education from the African view point still stresses practical-oriented learning with the holistic or *Gestalt* pedagogic approach, the Western has tended to be atomistic; learning one element at a time which is the abstraction of what is to be taught [Dargie, 1998:116]. However, it has the advantage of the use of computer technology at the centre [Mapoma in van Niekerk, 2001]. Dargie further asserts that in the African holistic approach the learning of music hinges on the ability to listen with feelings that are rippled throughout the whole body. This listening skill is more profound than that where music rudiments are perceived unconnectedly. In both approaches there is educational value. However, the challenge teachers of music face is their lack of confidence in balancing aspects of the two approaches in meeting the educational aspirations of the country as stated in the policy.

In the area of performance, to make the assessment of performing skills effective, singing, playing a musical instrument and dancing should be done in real-life situation. Learners must be given an opportunity to perform together with reputable musicians in the school during specially organised events. The learners can also give educational concerts to the school as well as the community [Myers in Taylor & Gregory, 2000:291]. In the process of this wholesome interaction, to refine the focus of assessment so that it is rendered valid, the following questions can be used: What skills are necessary for success on a *particular musical instrument*?

What is the most efficient means of acquiring those skills? What are the efficient and effective techniques in teaching the skills? [McPherson in Taylor & Gregory, 2000: 260].

Statement of the Problem

Despite the fact that musical arts have been assessed for many years through formative assessment in schools and colleges of education and through summative assessment by the Examinations Council of Zambia, the setting and conducting of practical skills assessment still have challenges.

An aspect of the listening skills assessment does not reflect much of real-life scenario. It still focuses on single elements and ignores the holistic presentation of music as it occurs in the community. The diversity in sonic environments in the various regions of the country that learners are naturally exposed to, makes it difficult to set inclusive items in the assessment of *timbre* without disadvantaging some candidates.

Although composing music is a very profitable skill, the assessment of which should give learners an opportunity to even gain experience in music recording skills, the financial cost that goes with the process poses a hurdle for many schools. For this reason the project in composition only ends at writing down the music. At junior level, the challenge is lessened by the fact that it is marked locally at the examination centre. Therefore, candidates have an option of bringing to life their compositions by performing them. The challenge is more at senior level where it is marked centrally.

The assessment of performance in the areas of sight-reading, singing, playing a musical instrument and dancing is conducted at the examination centres. Marks are then processed accordingly. The challenges here are that should a candidate express dissatisfaction over the results in this aspect, the examining board has nothing to fall back on as evidence.

The school music, which is more Eurocentric, differs in character from the music the learner experiences in the community all because teachers have a bias against indigenous and African contemporary music. In this respect, there is still a challenge of contextualizing musical arts education to assign relevance to the consequent setting and assessment of practical skills.

Teachers of music who are more oriented in Western art music and less so in African indigenous musical arts lack the confidence to assess practical skills that express African sensibilities. This has been a long standing challenge in both schools and colleges of education.

Purpose of the Study

The purpose of this study was to explore the teaching approaches and models of assessment used in practical skills [listening, composing and performing] in schools and teacher education institutions, and the confidence of the teachers and lecturers in assessing practical skills from the perspective of indigenous musical arts.

Objectives of the Study

The study had the following objectives:

1. To explore the music genres used in the schools and colleges in the teaching of practical skills.
2. To evaluate the confidence of the teachers and lecturers in assessing indigenous musical arts skills.
3. To explore the modes of assessment teachers and lecturers use in practical skills in musical arts education.

Research Questions

1. Which genres of music are used in teaching of practical skills in schools and colleges of education?
2. How musically equipped are the teachers and lecturers in handling assessment of practical skills in indigenous musical arts?

3. What are the modes of assessment teachers and lecturers use to assess practical skills?

Method

The study employed a survey research design. A questionnaire was administered in six institutions of learning located in Lusaka and Chongwe districts.

Out of a population of eleven government secondary schools offering Musical Arts Education in Lusaka district, four were sampled, representing 36 percent of the population. Two government teacher education institutions in Lusaka and Chongwe districts were sampled out of the population of three, representing 67 percent of the population.

All the nine (9) teachers and eight (8) lecturers in Musical Arts Education in the sampled institutions were respondents. The questionnaire was administered to the seventeen (17) of them. However, only twelve (12) responded which translates into 70.5 percent.

The instrument had 14 items which reduced to 11 after a validity scrutiny. The instrument comprised five parts. Part A was for respondent’s personal data, Part B was on the teaching of practical skills in musical arts, Part C was on the music teacher’s/lecturer’s musical bias in terms of Western art and African indigenous music, Part D was on the assessment of practical skills in musical arts and Part E addressed the interventions employed to ensure quality of instruction and assessment of practical skills in indigenous musical arts.

Results and Discussion

**PART B: TEACHING OF PRACTICAL SKILLS
(How often musical practical skills were taught in a week in schools)**

Figure 1(a) below indicates how often practical skills were taught in schools involved in the study.

- Responses from teachers

(The figures on top of the bars represent the number teachers)

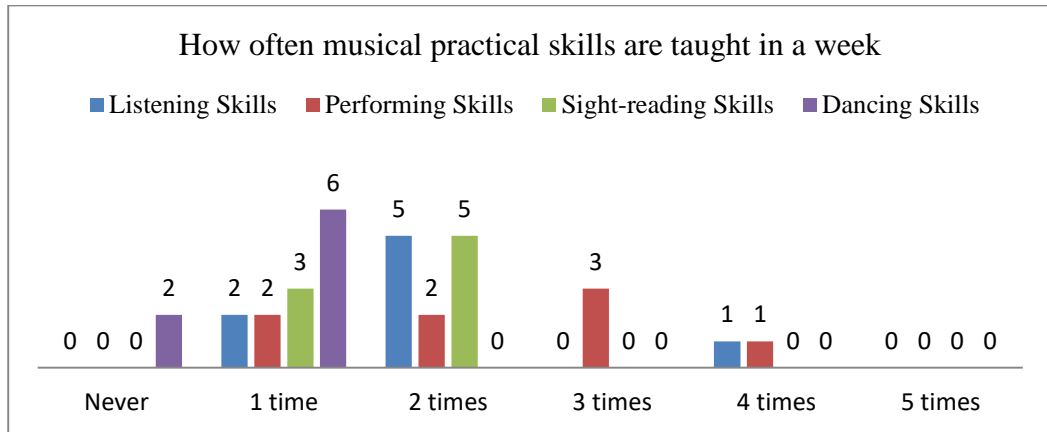


Figure 1 (a): How often the practical skills were taught in a week in schools

The data in figure 1 (a) above shows that only 12.5 percent of the respondent teachers taught listening skills four times a week. The highest number of teachers at 62.5 percent taught listening skills twice a week. Music is essentially sound. It is important that teachers of music teach music through music. They should not talk about music to the learners; they should make music speak for itself. Through ear training, learners develop analytical listening skills [Hoffer, 1967:21] which are indispensable in the life of a musician.

Performance is distributed across the week with 25 percent of the teachers conducting lessons once, another 25 percent twice, 37.5 percent three times and 12.5 percent once. Performing skills should be paid attention to because they constitute the main arena where the composer meets the audience. Music performance is a very high level thinking skill [Killian, 1995]. During performance a lot of decisions are made by the performer sometimes in split seconds in the process of bringing to life the composed music.

Sight-reading skill was taught only once a week by 37.5 percent and twice 62.5 percent of the respondent teachers. Sight-reading is a very important skill for both

instrumentalists and vocalists. It links well with the listening skill. Skilled sight-readers hear the tones in their minds before they play or sing them. Sight-reading develops other musical skills through inner-aural perception. It is a skill that should not be ignored in musical arts education [Brassine *et al* in van Niekerk, 1998:61].

Dancing is the least taught skill in schools yet it is assessed by the national examining board both at junior and secondary levels. 25 percent do not teach dancing skills at all whereas 75 percent, teach it only once. The teachers have left the training of dancing to the ‘indigenous school’ [*the community*] which makes up the cultural background. It could be that they do not know how to methodically go about it because the featured dances are prominently Afrocentric. Dancing has a lot of health benefits. It can help people stay fit in more ways than one.

Figure 1 (b) below indicates how often practical skills were taught in a week in teacher education institutions covered in the study.

- Responses from lecturers.

(The figures on top of the bars represent the number of lecturers)

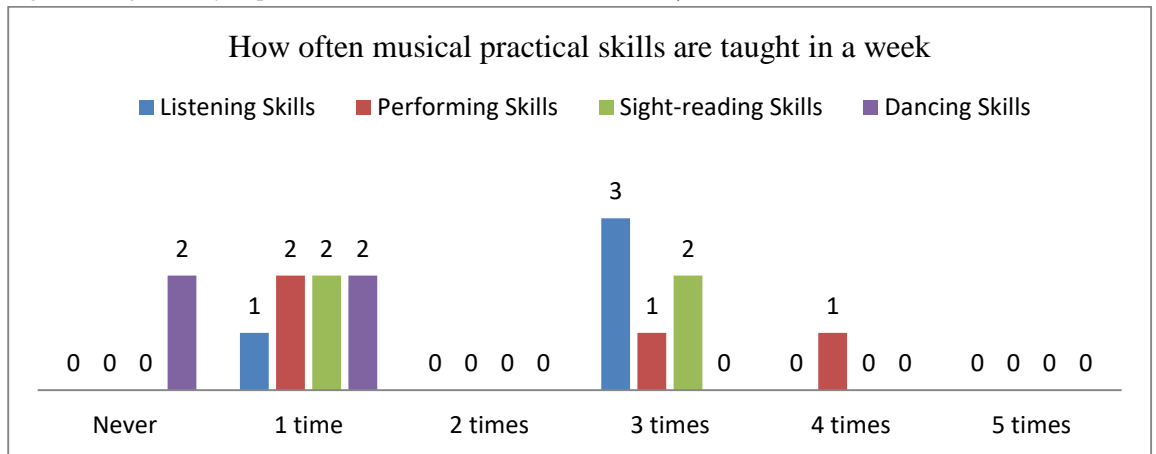


Figure 1 (b): How often the practical skills were taught in a week in teacher education institutions

At tertiary level listening activities were taught thrice by 75 percent of the respondent lecturers. Here is where listening techniques should be taught convincingly because the learners are going to be teachers. The methods applied should be valid. Ear training should be conducted with a focus of making the learners gain the ability to hear what they read [Brassine *et al* in van Niekerk, 1998:65].

Like at the school level, dancing skills are not paid attention to at tertiary. This is what trickles down into the schools. When the students graduate and get deployed in schools, they do not pay particular attention to dancing. The study also enquired on whether learners had exposure to music recording studios. This is a very important

component in preparing the learners for musical engagements in real life. Considering the central role audio recording technology plays in the music market today, the assessment of the composition project involves an option of candidates participating in the recording of their work. Despite the well appreciated necessity for learners to get inducted in audio recording basics, financial realities in the schools have kept the opportunities from being realised. However there were reports of one or two parents who assisted their children to access the audio recording studio for their composition project at junior level at some school in Lusaka. This was at a school that was not in the study. The results on this enquiry were as expressed below:

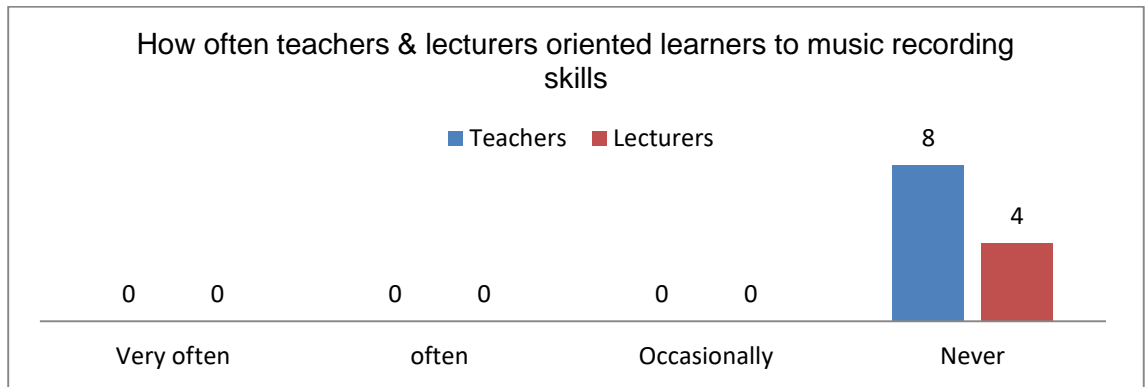


Figure 2: How often teachers/lecturers oriented learners to music recording skills

In the institutions covered in the survey, learners were not introduced to music recording skills. The use of the music recording studio has come up in musical arts as a means of equipping with special skills. It has been justified by the booming music recording business in the community. Learners are expected to be taught how to use, firstly, the music writing software and the affiliated functions of creating audio or MIDI files for example, and secondly, the basic steps in the process of recording. The hindrance may have been the lack of finances for the purchase of such software. Thankfully, the Ministry of Education

was, at the time of this study, in the process of distributing materials to schools which included computers. It will be upon the school administrators to purchase the relevant music software. Should it be found to be beyond the financial reach of the school, the administrators can seek to partner with parents or community members. The Ministry of Education supports such partnerships [MoE, 1996].

In some study that involved similar matters of cost sharing, one parent felt that if schools worked with the community, the financial burden of buying musical equipment would be made light. He also observed that a

school that had musical arts could not fail to raise money through musical concerts [Mubita...[et al] in van Niekerk, 2005:175].

At one of the tertiary institutions in the study, a senior lecturer mentioned that plans were underway for the institution to start teaching music technology which will necessitate the putting up of a music recording studio. He expressed hope that when the first cohort of teacher trainees graduate from the training in such technology, the music recording studios will be available in schools

to cover the training and assessment of music products like audio compact discs.

PART B: TEACHING PRACTICAL SKILLS

(Music genre used in the teaching of practical skills)

To find out the inclination of instruction in musical arts, the survey instrument had a question on the genre of music used. Figures on top of the bars indicate the number of teachers who used a particular music genre for a particular practical skill.

- (A respondent could choose more than one option.)
- Response from teachers

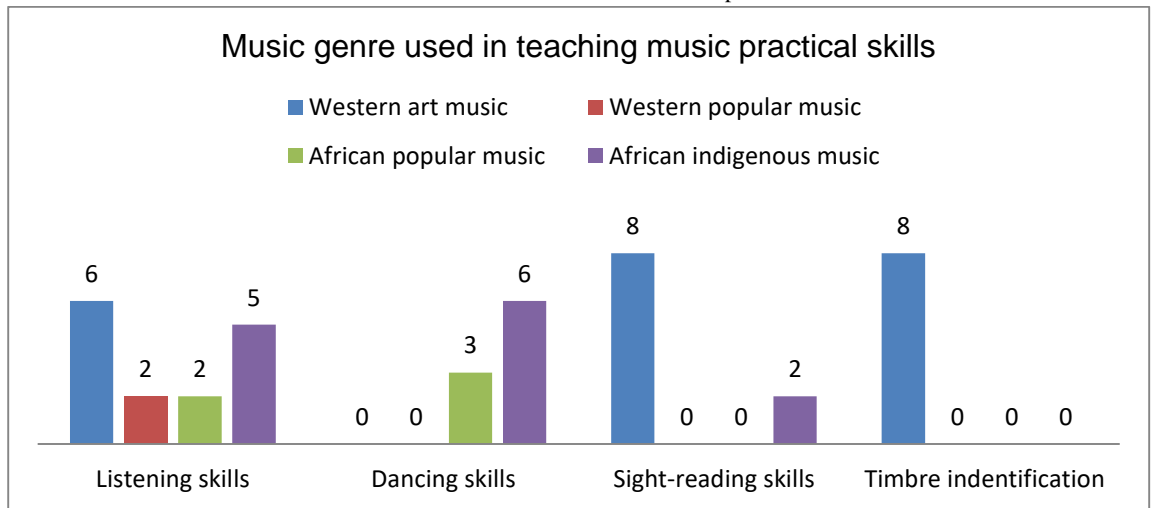


Figure 3 (a): Music genres used in the teaching of practical skills

As to which music genre was used in the teaching of practical skills in schools, figure 3 (a) above shows that teachers used all of the indicated genres in the survey instrument in the teaching of listening skills. For the art of listening, 25 percent of the teachers did not use Western art music, the rest did. 62.5 percent used African indigenous music. This means there was a big intersection of those who used both the African and the Western genres.

In the training of dancing skills, African music, in the categories of popular and indigenous, was used. As indicated above in figure 1 (a), dancing was the least taught in schools at the rate of once a week in the case of 75 percent of the respondent teachers and *not at all* in the case of 25 percent. This suggests that it was those same teachers who also used African music in the teaching of dancing skills in figure 3 (a). Some respondents claimed to have been using African indigenous in listening, dancing and sight-reading skills. It was also indicated that African popular music was used in the teaching of listening and dancing skills.

In the teaching of sight-reading skills 100 percent of the teachers used Western art music with only 25 percent using African indigenous music. When it comes to *timbre* [tone colour], we find that all the teachers used Western art music. It means that the *timbre* the learners are trained to identify was that of the Western musical instruments some of which may not be available even in the urban schools, let alone in the rural areas. This trend sets a lot of things upside-down. The natural way should have shown more use of the music from the learner's surroundings and less from the foreign environments. In

the teaching of musical arts education, it is meaningful to start from the music that the learners are familiar with; something from their cultural environment before proceeding to other types of music coming from other cultures [DoE, SA, 1995:1].

The same enquiry was given to lecturers in the two teacher education institutions. The responses were as indicated in figure 3 (b) below:

- (A respondent could choose more than one option.)
- Responses from lecturers

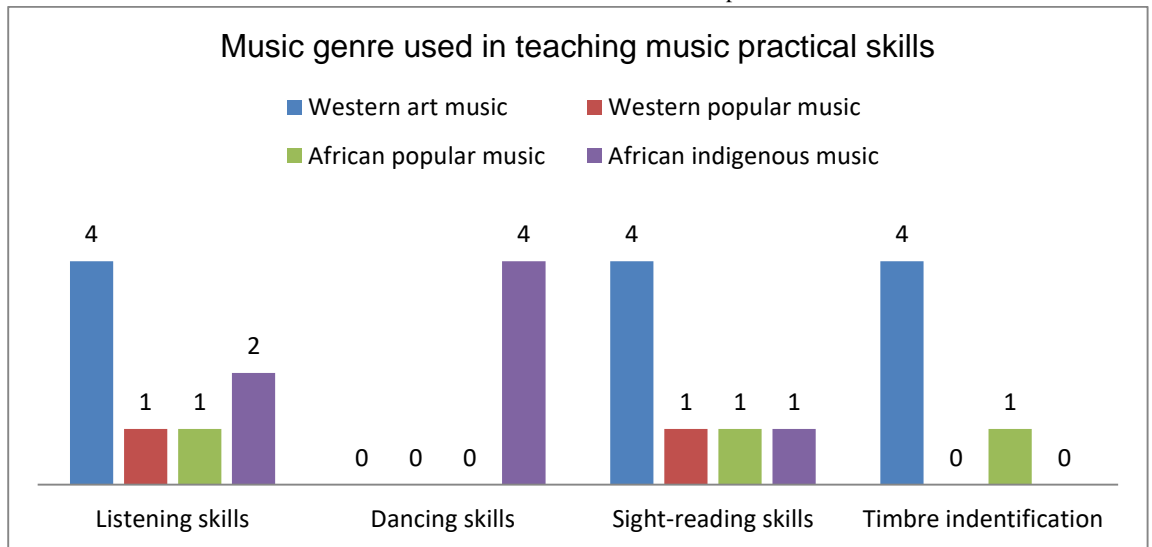


Figure 3 (b): Music genres used in the teaching of practical skills

The picture in the teacher education institutions is more like that in the schools. It is interesting to note that here again all the reflected genres in the survey instrument were used in ear training. The prominence of Western art music in the teaching of sight-reading and *timbre* identification skills is unchallenged by any other. 25 percent of the respondent lecturers used African popular music in *timbre* identification. This genre sometimes

includes local instruments. To some noticeable extent, it involves musical instruments that are used both in the urban and in rural areas, taking the example of African drums and shakers or rattles.

In the area of sight-reading, like in listening skills, only 25 percent of the lecturers used all the genres.

In teacher education institutions, like in schools, no one used Western art or Western popular music in the

teaching of dancing. When it comes to dancing, only African indigenous music was used.

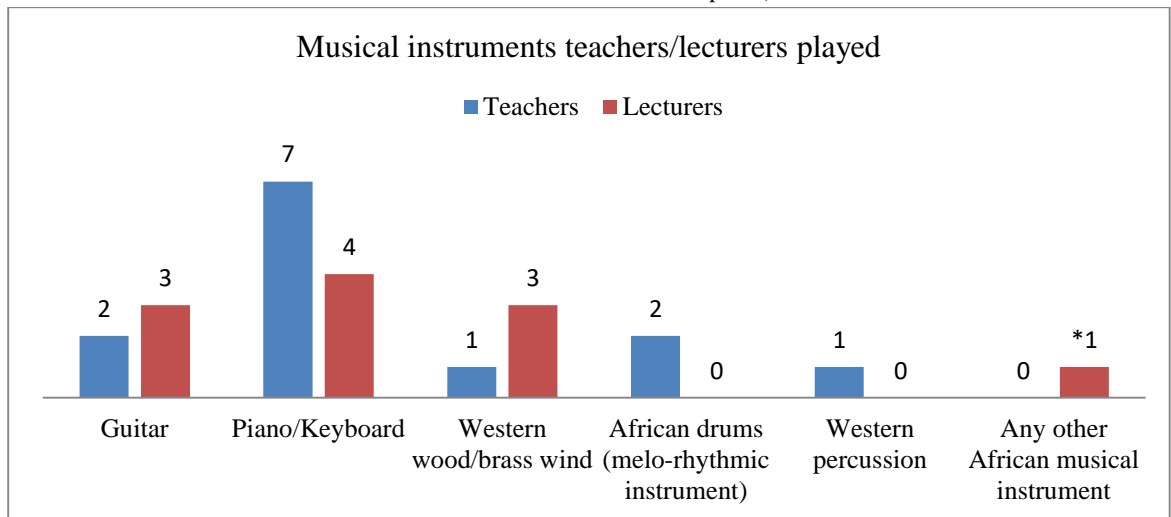
Findings from another study in the psychological dimension of music indicate that by listening to the music of their culture, people learn intuitively at a very tender age to appreciate musical structures [Herbst in van Niekerk, 2001:63].

PART C: MUSICAL BIAS OF TEACHERS AND LECTURERS

On the question of musical instruments the teachers/lecturers played, the responses were as indicated in figure 4 below. This was intended to investigate the Eurocentric and Afrocentric biases of the teachers and lecturers. The understanding is that teachers can give tuition only on the instrument they play. If it is on a Western musical instrument, then the practical skills will be pro-Western and vice versa.

Figure 4 shows the responses from teachers and lecturers.

- (A respondent could choose more than one option.)



*Mbira from Zimbabwe.

Figure 4: Musical instrument teachers/lecturers played

All teachers played musical instruments of Western origin. It was only 25 percent of them who also played the African drum. The African drum is not classified as a percussion instrument in the African musical context. It serves as a melo-rhythmic instrument for it can make a noticeable melodic contour unit combined with some rhythmic structure [Nzewi, 1997].

As stated above, the assumption here is that teachers and lecturers can only teach skills on the instrument they can

play satisfactorily. We can guess that the scenario in the schools and teacher education institutions in the study is that learners also have the same inclination toward the use of Western musical instruments. This gives a different picture from the findings of a study done in East Africa by Agak [in van Niekerk, 2001:42] where some secondary schools in Uganda taught the Tube fiddle [Endingidi] and the Thumb piano [Mbira, Zansa, Adongo] to the learners. Indigenous songs were also taught and training in dances and indigenous musical instrumental playing covering the aspects of theory and practice was given.

This brings us to a point where we should know which aspect of the Musical Arts Education syllabus [CDC, 2014) the teachers and lecturers were confident in. The Musical Arts Education syllabus, in line with the curriculum requirement, covers not only Western music but also world musics, African indigenous and the Zambian contemporary music. Previously, the school syllabus did not expressly indicate these areas in the music course.

Often, lack of teaching materials is cited as a justification for the teachers' failure to teach some parts of the syllabus. However it is surprising to find such situations in our schools in this computer age when Internet serves

as a port to many sites where materials on a wide range of subjects can be found. It has been noticed from other fora, like the subject association conferences, that there is a tendency in most teachers in practical subjects to exclude themselves from the list of potential text book writers. This has undesirably affected the availability of teaching materials.

In regard to this teachers and lecturers gave responses as shown below:

- (A respondent could choose more than one option.)

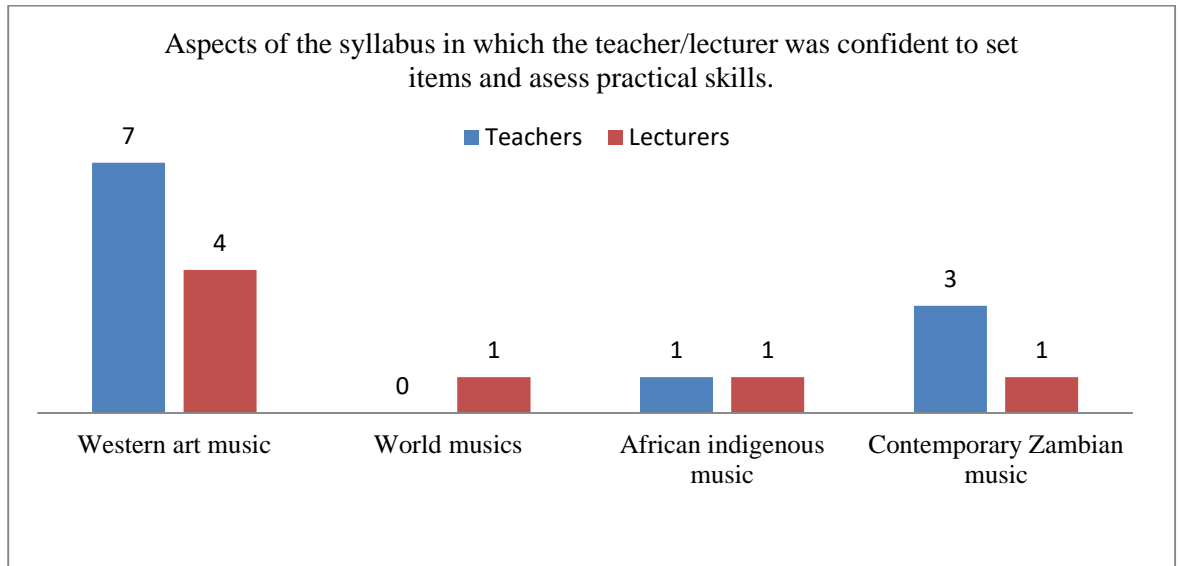


Figure 5: Aspect of the Musical Arts Education syllabus in which the teacher/lecturer was confident to set items and assess practical skills

In the study, only 17 percent of teachers/lecturers indicated that they were confident to handle the aspects of the syllabus that dealt with African indigenous music. 33 percent expressed confidence in handling Contemporary Zambian music. Other studies in African

schools and institutions of higher education showed that there were very few lecturers who could competently teach the theoretical content of African music [Nzewi in van Niekerk, 1998:465], let alone the playing of the African musical instruments.

PART D: ASSESSMENT OF PRACTICAL SKILLS

The study looked at the model of assessment the teachers/lecturers used in assessing music performance. This was the crux of the matter. The approach in teaching suggested the model of assessment. The assessment should point to the desired practical skills in a way that is relevant to the field as portrayed in real-life. The argument here is that education may be likened to dietetic practices. Much as the general purpose of eating is to keep healthy; there is diet for the wellness of the heart, for weight loss, for sound eyesight and so forth. Although education may be for the general intellectual growth of the learners, it can deliberately target the improvement of specific understanding and skills. The choice of the

assessment model for the deliberately chosen skill will, in this case, be determined by its validity; how well it measures the intended skills.

Practical skills in musical arts should therefore be assessed in a manner that focuses on the desired skills for the learners to develop the musicianship that is applicable in the social and economic arenas of their society.

Four models of assessment were put in the survey instrument for the respondents to indicate which of them they used in the assessment of practical skills in musical arts. The results were as indicated below:

- (A respondent could choose more than one option.)

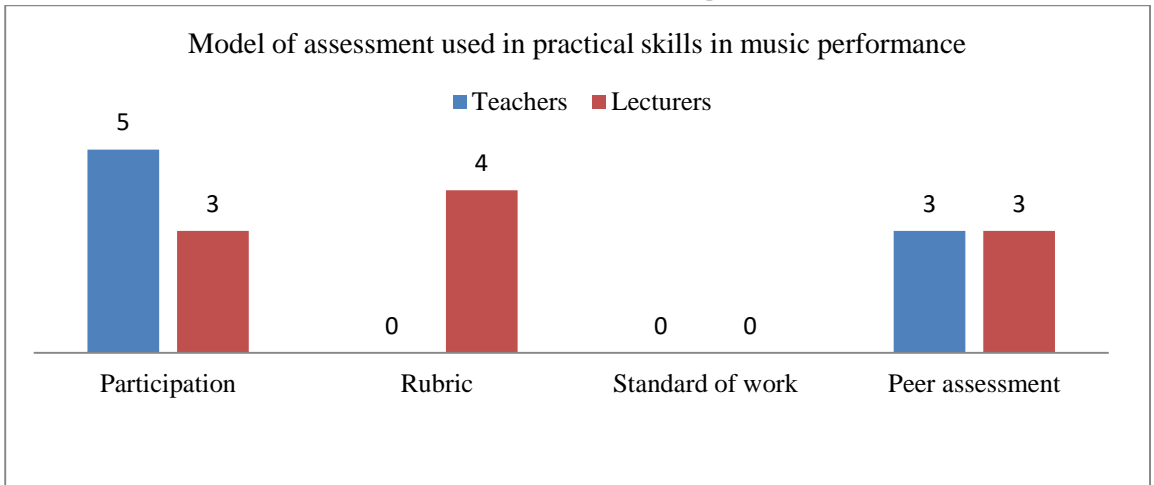


Figure 6: Models of assessment used in assessing practical skills in music performance

The participation model suggests assessment of an individual in a group performance. Rubric may be applied to both individual and group performance. Standard of work suggests more of solo performance assessment. Peer assessment can be either for solo or group performances.

The assessment of music performance in a group brings in an educationally rich collaborative aspect. Music performance as a collaborative activity brings out the value of teamwork and fellowship as it improves interaction between performers who are involved in a mutual task. It focuses on social values and life skills. However, the challenge of this model of assessing music performance is the demand it exerts on the assessor. To ensure that group performance assessment is effective,

attention must be paid to the contribution of each individual in the group [Killian, 1995].

The information in *figure 6* indicates that **participation** and **peer assessments** were the common models of performance assessment in the schools and teacher education institutions covered in the study. Participation model of assessment entails that marks are given to learners on the basis of participation. It is taken for granted that those that participate learn something and deepen their understanding and improve their practical skill. The theory behind it is that when in the environment where music is happening, a person, who may lacking in one aspect or the other, will learn through *vicarious conditioning* by observing others who are better in singing, playing musical instruments or in dancing [Mwesa in Herbst, 2005: 186].

The results indicate that the rubric model of assessment was used only in the teacher education institutions. This is the same model that the national examining board uses for summative assessment. A rubric is a form of a matrix indicating specific ability or set standard for a specific mark. It is also called a grading scheme.

Peer assessment is the model of assessment involving learners. They give feedback to their fellow learners. The assessment is based on the agreed upon standards. This model of assessment encourages the acquisition of practical skills and gives confidence to the learners, especially when their observations are reinforced by the instructor’s remarks. Peer assessment encourages self-assessment. With self-assessment, we can suppose that a learner’s performance skills can improve influenced by some elements of intrinsic and extrinsic incentives.

With regard to holistic and atomistic approaches, the results were indicated hereunder:

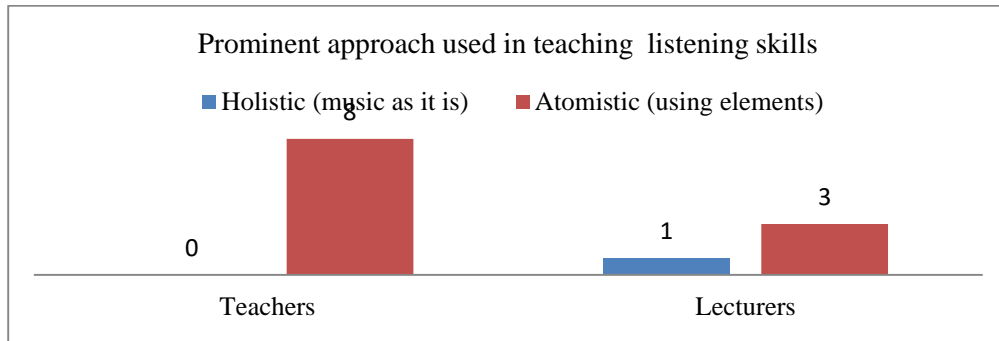


Figure 7: The prominent approach used in assessing listening skills

It can be noticed here that the teaching of aural skills is done through what is referred to as ‘abstraction of what is to be taught’ [Dargie in van Niekerk, 1998:116]. This approach is good but not comprehensively so. A more comprehensive approach, which is typical to Africa, is that which is *Gestalt* in nature. Herbst [in van Niekerk, 2001: 60] observes that

‘African composers take advantage of the psycho-acoustical fact that the human mind is predisposed to group objects of similar or equal qualities in the process of *Gestalt* formation’.

If we are to equip the learners with important listening skills in musical arts we need to emphasize the holistic approach in assessment. When learners get used to hearing isolated elements, they may find it difficult to

decipher the music they encounter in real life situations. This may delay their stability in the business of making music for a livelihood should vocation take them that route.

Nzewi [2005: vii] makes an observation that

‘The indigenous “classroom” [the *community*] provides a holistic approach to life-long learning which is contextualized within daily socio-cultural living. Indigenous African musical arts philosophies, structures and practices contain the knowledge and materials for culturally relevant, theoretical, creative and practical musical arts education and practice’.

It follows therefore that if assessment of musical practical skills is not valid in the sense of targeting the relevant skills, those which would be readily applicable in the community music arena, the musical arts will be rendered ineffective with regards to improving lives.

On finding out how learners responded to the marks given to them using the adopted assessment models, teachers and lecturers gave the responses as indicated below:

- (A respondent could choose more than one option.)

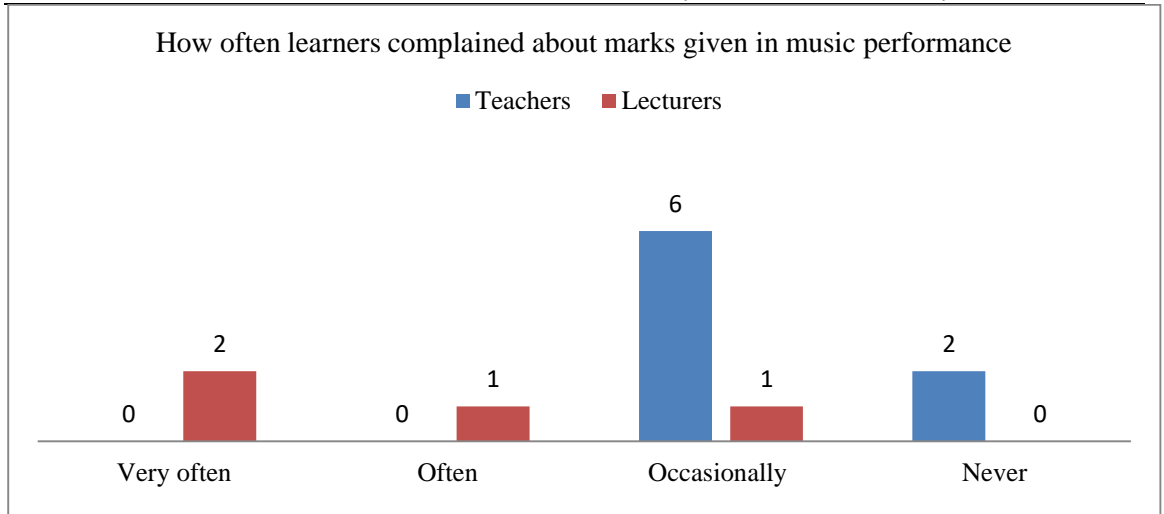


Figure 8: How often learners complained about marks given in performance

The assessment that takes place in the process of learning is referred to as formative because it provides information on the learner's improvement. Where formative assessment is applied correctly, it gives feedback in real time. However the information in *figure 8* above shows that in the study, of all the respondents only 16 percent never experienced a grievance from learners on how they would have marked their performance. 58 percent had occasionally received the evidence that learners accepted the marks they got with reservations. 25 percent received complaints at the frequency of often to very often, and

this happened exclusively in teacher education institutions.

This shows that the assessment of musical performances as it is in the institutions of learning needs to be looked at. Measures should be taken to improve the models. If, for example, the same rubric model which is used in the formative assessment is the same that is used in the summative assessment, we can expect that if the failing candidates got to know their component marks for performance, some would express misgivings and launch an appeal to the examining board.

It was therefore necessary to find out how the teachers and lecturers rated their assessment pertaining to being objective or subjective. Following were their responses:

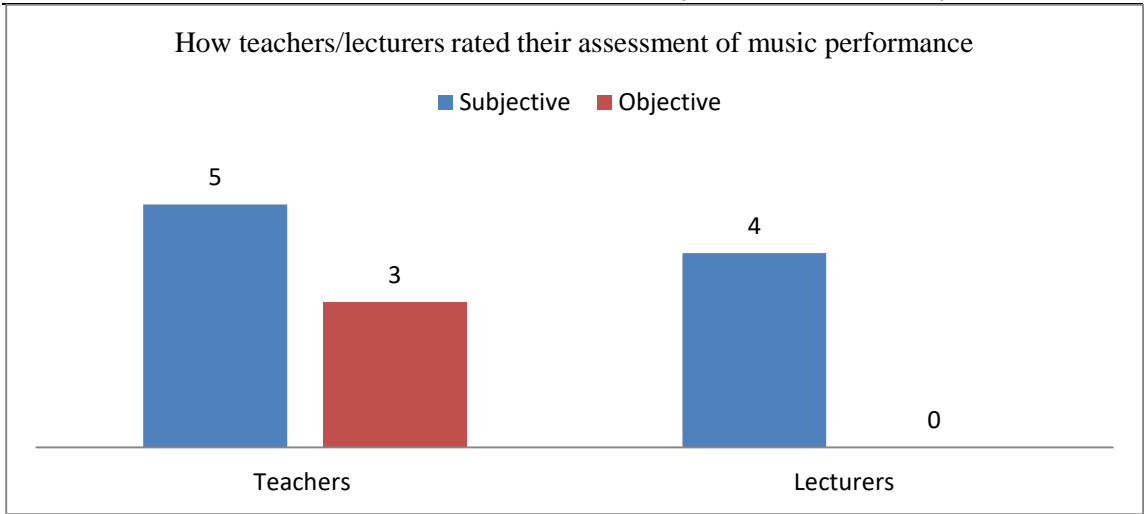


Figure 9: How the teachers/ lecturers rated their assessment of musical performance in terms of being subjective or objective.

Responding to the item that asked whether the teachers and lecturers assessed objectively or subjectively, 25 percent of the respondents indicated that they rated the performances objectively while 75 percent indicated that they did it subjectively. In a short interview, the researcher had with the teachers, those who claimed to mark objectively justified their stance. They said following the assessment tool prevents subjectivity.

All the lecturers in the study indicated that their evaluation of practical performance in musical arts was subjective. In a short interview, one of the lecturers said that it was very difficult to assess musical performances objectively. On the one hand, it is true that there is personal involvement in the assessment of musical arts;

as an art, music appeals to the emotional or affective centre of an engaged viewer or listener. The appreciation that comes forth is therefore subjective. On the other hand, he also observed that since music is, besides being an art, a science as well, it is the scientific tasks that can be objective.

This leads us to ask the following questions: If it is difficult to objectively judge musical performances, can we ever expect reliability in this aspect of Musical Arts Education? Can there be any comparability? When we ignore the emotional appeal of music performance, will we not be eluding the real-life justification of musical arts?

When asked which component of the musical arts syllabus the teachers and lecturers had enough teaching materials in, the responses were given as shown below:

- Figures on top of the bars represent the number of teachers/lecturers

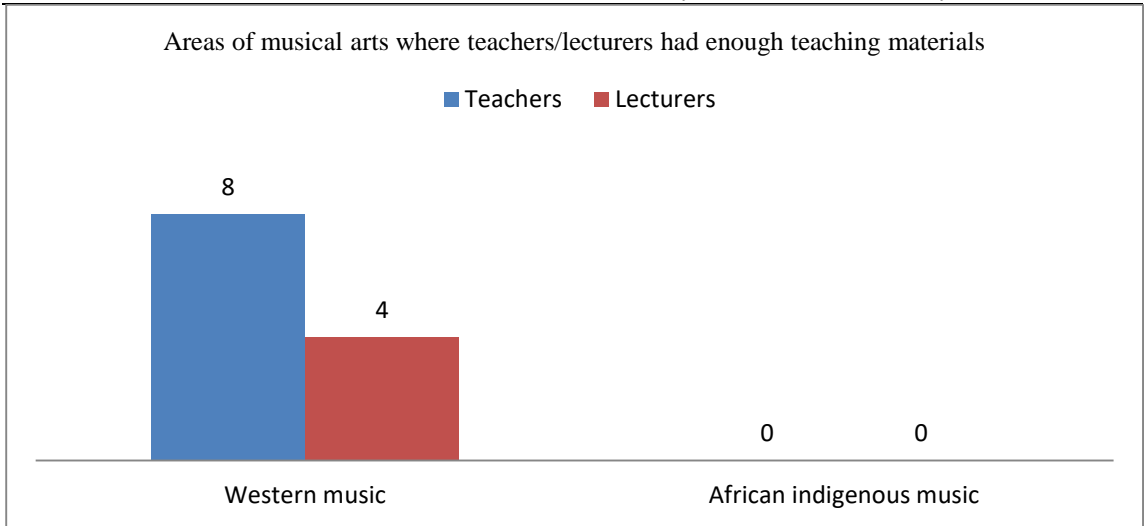


Figure 10: Aspect of the musical arts where teachers/lecturers had enough teaching materials

It is surprising to find that all the teachers and lecturers in the study indicated that they did not have enough teaching materials in African indigenous music. The researcher feels that the materials to teach Western music cannot be more than those to teach African indigenous music. The community music is a sure source of such materials, all in public domain. Folksongs and the contemporary music are all local materials for teaching musical arts. We may not be able to see them because we are biased towards Western music.

PART E: INTERVENTIONS EMPLOYED TO ENSURE QUALITY OF INSTRUCTION AND ASSESSMENT OF PRACTICAL SKILLS IN INDIGENOUS MUSICAL ARTS

The study also considered finding out whether the schools and teacher education institutions used community based experts (connoisseurs) who are commended as experienced people in musical arts. Respondents were asked whether they used connoisseurs both in instruction and in assessment. The responses were as shown below:

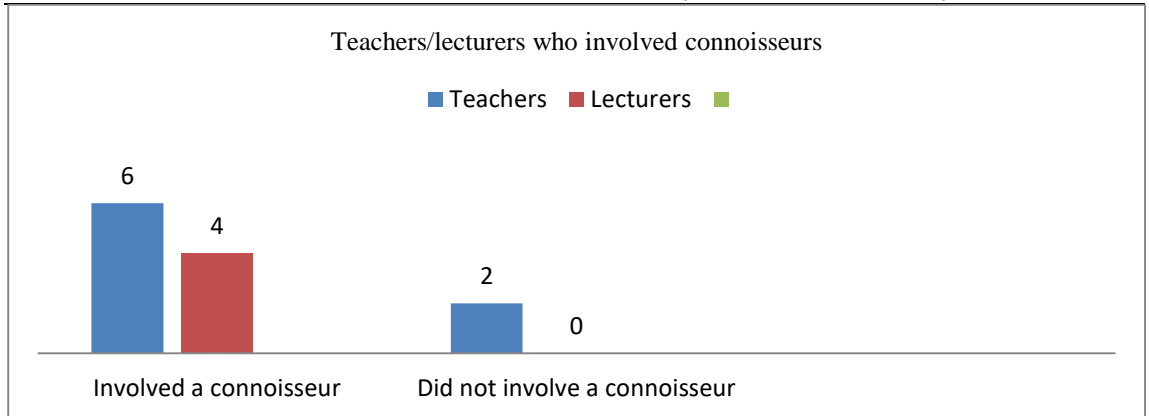


Figure 11: Whether teachers/ lecturers involved a connoisseur to complement their teaching and assessment of practical skills

It has been seen that all the teachers and lecturers in the study were more skilled in Western than in African indigenous music. We have even seen that they feel more confident to assess Western musical skills as opposed to the African indigenous musical skills. It is for this reason that they have indicated, as is expressed in *figure 11*, that they used *connoisseurs* of music from the community to help out. But these connoisseurs are not in the picture when it comes to the national summative examination.

It is not only in Zambia where such experts are used. The *connoisseurship* model of assessing practical skills is used in other places as well the example of which is the United Kingdom where it has been in application for a long time especially in the arts [Robbins, 2007]. In line with this view, Miya in Oehrle [2003:3] says

‘No one can claim to know everything about music in Africa. African cultures are so diverse and so is their music.’

It is therefore justified to consult an expert in the community where the class teacher is not confident. In

addition to this the Department of Education in South Africa stated that

‘Because music arises from and functions within a rich variety of life experiences...it follows that, when attempting a music practice with which the teacher is not acquainted, a person with more knowledge of the subject should be consulted ...’ [DoE, RSA, 1995].

The credibility and reliability of the *connoisseurship* model of assessment depends on whether the *connoisseur*, as an expert, demonstrates repeatability of judgement. In other fields it will depend

‘on how well the link is between the assessor and the society of practitioners of that particular subject’ [Robbins, 2007].

Conclusion and Recommendations

Conclusion

From the findings presented above, the following conclusions are made:

Firstly, the teachers and lecturers are biased towards Western music. Therefore when setting musical items they target practical skills and assessment approaches that are appreciated from the Eurocentric perspective.

Secondly, teachers and lecturers lack confidence to teach aspects that deal with the theory and practical skills in African indigenous musical arts.

Thirdly, assessment of *timbre*, when conducted with a Eurocentric focus may disadvantage learners in the rural areas.

Fourthly, the teaching and assessment of skills in music technology and the affiliated techniques in music recording are still farfetched.

Lastly, learners in teacher education institutions often complain about the marks they are given during the formative assessment in musical performances. There is therefore a possibility for a candidate to launch an appeal and request for a remark of the musical performance done during the summative assessment.

Recommendations

The paper recommends the following interventions:

1. The subject association concerned with musical arts should organise workshops for the re-orientation of teachers and lecturers in the teaching and assessment of African music in general and the required practical skills, in particular.
2. At these workshops, techniques in the assessment of aural skills should be re-evaluated.
3. School administration should engage teachers of music and other stakeholders in addressing the issue of music technology – focusing on music recording skills.
4. The examining board should identify and make a repertoire of dances and indigenous musical instruments from the regions on which the setting and assessment will be based. This will create a common ground and mitigate the

undesirable effects of musical diversity of the country on the national examinations.

5. Audio-visual technology should be used in the assessment of practical skills in music performance especially in summative assessment. Individual and group performances should be recorded and the product kept by the examining board. In case of an appeal for a remark, the examining authority will have something to fall back on.
6. The involvement of *connoisseurs* in musical arts should be regulated and legitimized through a process of accreditation. A recognised assessment board can accredit them and this will upgrade their status to a level where not only the schools can employ them on local arrangements, but even the examining board could use them in summative assessment in national examinations.

References

- Curriculum Development Centre [2013]. *Zambia Education Curriculum Framework*. Ministry of Education, Science, Vocational Training
Curriculum Development Centre [2014]. *Musical Arts Education Syllabus, Grade 10 to 12*. Ministry of Education, Science, Vocational Training and Early Education, Lusaka: Zambia.
- Department of Education, South Africa. [1995]. *Supplementary Material, Interim Core Syllabus 1994, Class Music Grade 1 to Standard 10*.
- Examinations Council of Zambia [2015]. *Junior Secondary School Leaving Examination Assessment Schemes*. Oxford: Bookworld Africa limited.
- Examinations Council of Zambia [2016]. *School Certificate Examination Syllabi*. Oxford: Bookworld Africa limited.
- Herbst, A. [ed]. [2005] *Emerging Solutions for Musical Arts Education in Africa*. Selected articles prepared

- in advance for the PASMAE conference and commissioned research-based chapters on the collective voice of conference participants. 5 to 11 July, 2003, Kisumu: Kenya
- Hoffer, C.R. [1967]. *The Understanding of Music*. Belmont: Wadsworth Publishing Company Inc.
- Kebede, A. [1995]. *Roots of Black Music: The Vocal, Instrumental, and Dance Heritage of Africa and Black America*. New Jersey: Africa World Press
- Killian, J.N. [1995]. *Assessment in Music Performing Groups: Some Practical Suggestions*. Paper prepared for presentation at Suncoast Music Education Forum, University of South Florida Tampa, 12 to 14 October, Florida: USA.
- Ministry of Education [1996]. *Educating our Future: Policy on Education*. Government Printers, Lusaka: Zambia.
- Nzewi, M. [1997]. *African Music: Theoretic Content and Creative Continuum*. Oldershausen: Institut für Didaktik Populärer Musik.
- Nzewi, M. [2005]. *Learning the Musical Arts in Contemporary Africa Informed by Indigenous Knowledge Systems Volume 1*. Ciimda Series: Compress www.compress.co.za
- Oerhle, E. [2003]. *The Talking Drum*, News Letter Issue No.20. Network for promoting intercultural education through music. PASMAE.
- Robbins, J.H. [2007]. *Application of Process Control for the Maintenance of Standards and the Quality Assurance of Results in a Connoisseurship Model of Assessment*. Paper presented at the 33rd IAEA Conference, 16 to 21 September, Azerbaijan: Baku.
- Taylor, M. & Gregory, B. [eds]. [2000]. *Music of the Spheres*. Conference Proceedings for International Society for Music Education, 24th ISME World Conference, 17-22 July, Edmonton: Canada.
- Van Niekerk, C. [ed]. [1998]. *Ubuntu: Music Education for a Humane Society*. Conference Proceedings for International Society for Music Education, 23rd ISME World Conference, 19-25 July, Pretoria: South Africa.
- Van Niekerk, C. [ed]. [2001]. *Selected Conference Proceedings from the Pan African Society for Musical Arts Education PASMAE Conference*, 21-25 August, Lusaka: Zambia.

Investigating the Appropriateness of the use of Oral Communication Competencies as a Practical Component in the Assessment of Language: The Case of Ordinary Level Special Bilingual Education (SBE) French in Cameroon.

Monono Ekema Humphrey (PhD)*

hemonono54@yahoo.com

John Teneng Awa (PhD)*

tndipingwi@yahoo.com

Kenneth NguFoncha (PhD)*

kfoncha@yahoo.com

***Cameroon GCE Board**

Oral assessment (*Viva voce*) measures students' knowledge at any level, in probing students' level of understanding. Concerns about validity, reliability and fairness, however, have been raised as to the appropriateness of using oral examinations as a practical component, especially in the assessment of language in high stakes testing. In the Cameroon General Certificate of Education (GCE) examinations, these concerns are exacerbated by the extraneous variables of lack of anonymity, bias, examiners' competencies and the absence of a moderation and quality control framework. There are genuine worries about the 25% weighting assigned to the oral component of this examination when the validity and reliability of the items are not ascertained. This study investigated the extent to which performances in the oral assessment in Special Bilingual Education (SBE) French are noticeably different from their performances in written assessment. Scores for all the 792 candidates for the Ordinary Level Paper 1 Multiple Choice Questions (MCQ), Paper 2 (Essay) and Paper 3 (Oral) examination of the 2016 GCE in Special Bilingual Education French were entered in Microsoft Office Excel Version 2010 and analyzed using SPSS 20.0 (IBM, 2012). The variables were first of all explored using Boxplot so as to identify outliers and assess their validity. Given that the three variables were scored in different scales Multiple Choice Questions (MCQ)- 50, Essay- 160 and Oral- 20), they were all harmonized on a 20-point scale. The Cronbach Alpha's reliability coefficient was very small (-0.085), indicating that the internal consistency assumption was violated. This is a clear indication that the trend of scores within cases and across test types is chaotically distributed or does not follow any reasonable pattern. The intra-class correlation coefficient that produces measures of consistency or agreement of values within cases was almost null ($R=0.027$, $P=0.905$) and by so doing, strengthening the violation of the internal consistency assumption. The study recommends that a rigorous framework that precisely defines the construct to be measured, the marking and grading criteria be put in place to moderate the inflated scores. Furthermore, the 25% weighting of the oral component of the SBE French be reduced to 10% in keeping with all practical examinations at the Board.

Key Words: Appropriateness, Practical Component, Oral Communication Competences.

Introduction

Oral assessment appraises students' capacity to think on their feet, (Ahmed, Pollitt and Rose, 1999). It is employed

to test students' understanding at any level, particularly in investigating students' levels of understanding and in assessing that understanding in the context of its application. Whereas conceptual and procedural knowledge can be appraised through various methods, oral assessment may be preferred when there is a need to ensure that the responses belong to the student. Arguments for oral communication had been advanced since the period of Classical Greece. Plato, 424-348, argued for the pre-eminence of the spoken word saying "*writing would destroy the need for memory*".

In recent times Ong (2002) supports the case for oral communication when he states that in oral cultures, people identify themselves with their words, whereas writing has the effect of "*separating the knower from the known*". Ong continues that the spoken word is combative and polemical. Kehn (2001) opines that one of the strengths of oral assessment "*is the ability to distinguish superficial from real knowledge through in-depth questioning*".

Background to the Study

The merger between Anglophone and Francophone Cameroon demanded the policy of bilingualism to be instituted. At its beginning, it was understood that bilingualism would initiate a new curriculum organization which will offer solutions to the multifaceted process of social change. The strengthening of bilingualism, it was thought, would be critical for national unity and the integration of Cameroon.

However, the term bilingualism in the Cameroonian context has been difficult to define, describe and delimit. Hoffman (1991) asserts that it may be up to everyone (or community) at least, to choose the definition of bilingualism that best suits his or her purpose. Ayafor (2002) contends that from government's point of view, language policy is limited to the use of English and

French within government domains and formal transactions within private sector domains. This exclusive language management strategy has been summarily named 'Official bilingualism'.

Chumbow (1980) had earlier contradicted this when he stated "Yet, it may not be viewed in linguistic principles as a linguistic policy in the real sense of the term as it lacks basic clear-cut linguistic definition, description, nor a systematic framework for its implementation. Tadjadjeu (1983) asserts that lack of adequate linguistic prescriptions and frame of action leaves the policy no chance to develop as a language policy and it would not function in a socio-linguistic approach. There is lack of political courage and genuine political commitment in defining a consistent language policy and putting in place a systematic framework for its implementation.

Bobda and Tiomajou (1995) have indicated "... due to the absence of clearly defined objectives, the policy regarding bilingualism in Cameroon has remained over the years vague with a political rather than a linguistic goal." This position introduces the argument that bilingualism in Cameroon is not a language policy, thus putting its role as an integrative policy to question. It is a political instrument used to coerce and hold two separate political entities together for the political purpose of creating a unitary state.

In other words, French and English are only an instrumental medium of reaching and making re-unification and the unitary state work at any price (Chumbow, 1980). This explains why government lacks the political courage to come up with a policy formulation, as it might lead to ethnic or regional conflicts which might eventually spark national disintegration. Most government policy on the teaching and learning of English and French have been presented in vague, ambiguous and general terms to make non-compliance easy. They are usually presented as a mere statement of intent, without any provision for implementation. "It may well be articulated in

pronouncements and policy documents, but remain on paper as a manifesto or its implementation may be aborted by elite self-interest” (Bamgbose, 2000).

Tchombe (2000) in appraising the teaching of French to Anglophones acknowledged the “absence of equity in

focus.” The preparation of teachers to teach English to Francophones is well organized, even though in the field they meet with resistance. This is not the case of the teaching of French to Anglophones, who are very willing to learn French in order to survive.

The overpowering role of Pidgin English as a linguistic bridge between the two linguistic communities, both in private and official domains, has weakened the desire for Anglophones to learn French. Added to this, is “Franglais” a metropolitan hybrid code from French and English.

With the advent of globalisation and the ICT culture and the domineering role of English as a world language, students are beginning to have a passive attitude to the significance of French. This passive attitude encourages a negative attitude to French and eventually has significant implications on students’ performance.

Special Bilingual Education French (SBE) in the Cameroon General Certificate of Education (GCE) Examinations

In an effort to strengthen bilingualism, Special Bilingual Education French (SBE) was coined by officials of the Ministry of Secondary Education in Cameroon, adding to the already existing Ordinary Level French. Schools and candidates offering SBE are designated by the Ministry of Secondary Education. According to the officials, successful candidates in this subject will be able to:

- Communicate with ease using the skills acquired through this language.
- Apply knowledge, skills and understanding of social norms to face ordinary and complex life challenges.

- Acquire a physical, sporting and artistic culture through critical and responsible behaviour.

The Special Bilingual Education (SBE) French examination consists of three parts: Part 1 is Multiple Choice Questions (MCQ) weighted 25%; Part 2 is

Composition Writing weighted 50%, while Part 3 which is the oral examination (practical) is weighted 25%.

Arguments for/against the use of Oral Assessment

Adherents to oral assessment argue that it tests the limits of a candidate’s knowledge and understanding. This applies to both exceptionally capable candidates and candidates whose weaknesses might otherwise not be clearly exposed in the written component alone. Mackenzie, (2000) opines that in an Occupational Therapy Course, vivas are used both to develop and assess essential practice skills, including solving clinical problems, defending professional decisions, articulating rationales for interventions, developing and communicating innovative ideas-through being assessed and participating as assessors during presentations.

Joughin (2008) gave several reasons for using oral assessment:

- It is the best way to assess particular learning outcomes or abilities. Oral assessment provides insight into students’ cognitive processes. It permits judgment about students’ interpersonal and intra-personal competences which may not be evidenced in other forms of assessment.
- It allows probing of the depth and extent of students’ knowledge. It allows follow-up questions that can be used to determine the limits of what the students know unlike written examinations.
- It reflects the world of practice. Most fields of practice are dominated by talking rather than writing. Professional practices such as law, teaching and nursing are dictated more by talking.
- It improves the quality of student learning since students who anticipate being asked questions that

they cannot predict conclude that the best way to handle this situation is to develop a thorough understanding of what they are studying.

- It suits some students. Some students express themselves better orally than in writing.
- Unclear or ambiguous questions can be re-expressed or immediately clarified.
- It guarantees that the work is the student's own. Joughin (2008) further argues that many students have expressed a strong association with the words they use in oral presentation: "I own the words I speak more than I own the words that I write"

Joughin (2008) equally highlights the difficulties of Oral Assessment as:

- Undue anxiety.
- Hearing or speech difficulties.
- Oral assessment can be time-consuming particularly with large classes.
- Lack of anonymity. In several cases, examiners inevitably know whom they are examining.
- Worries are sometimes expressed that examiners may be influenced by students' gender, ethnicity and socio-economic background.
- Best practice requires a record of the assessment for future reference in case of appeal. Making and storing audio or video recording can be a daunting task.
- Examiners can mistake a student's articulateness for knowledge.

The Problem

There are genuine worries about the 25% weighting assigned to the oral component of this examination given that the validity and reliability of the items is not ascertained. These concerns are exacerbated by the extraneous variables of lack of anonymity, bias, assessors' competencies and the absence of a moderation and/or quality control framework.

Research Question

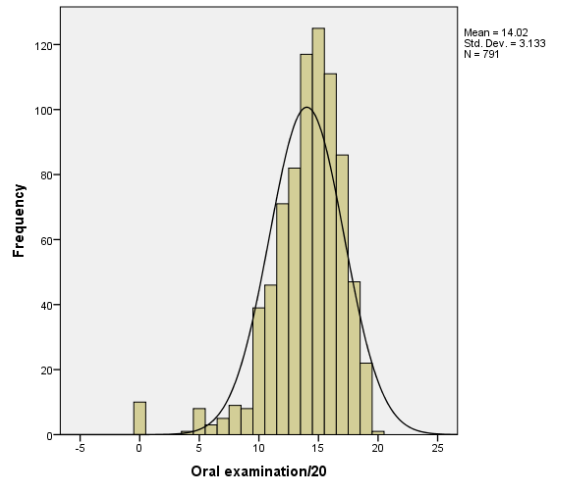
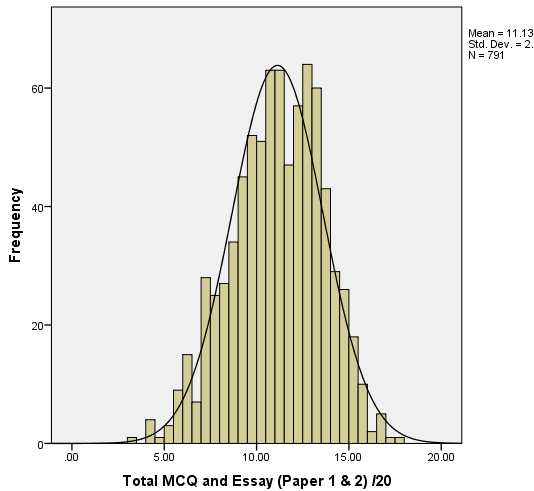
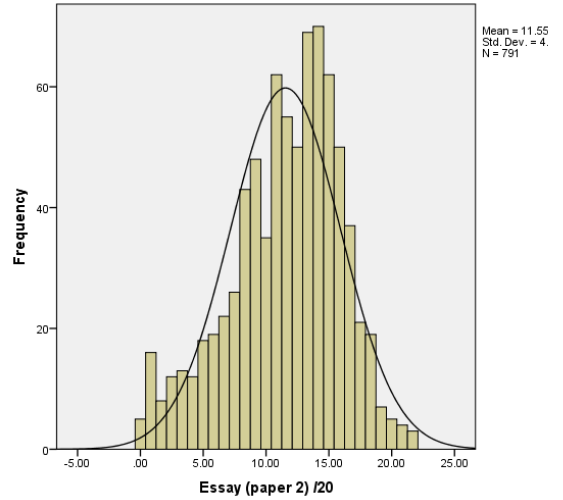
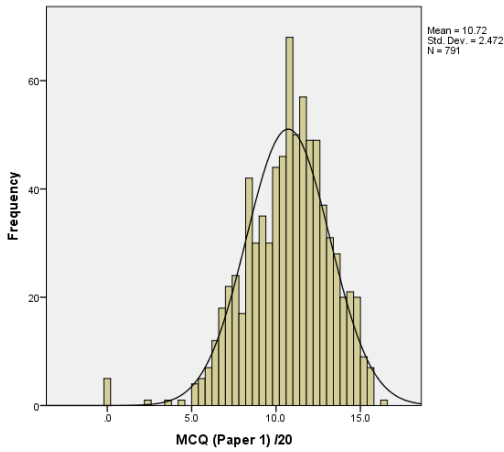
What is the effect of the 25% weighting of the oral component of the Special Bilingual Education (SBE) French on the overall performance of candidates on the subject?

Methods of Data Analysis

Scores for Paper 1 Multiple Choice Questions (MCQ), Paper 2 (Essay) and Paper 3 (Oral) examination for the General Certificate of Education (GCE) Special Bilingual Education French were entered in Microsoft Office Excel Version 2010 and analyzed using SPSS 20.0 (IBM, 2012). The variables were first of all explored using Boxplot as to identify outliers (bad data points) and assessing their validity. Given that the three variables were scored in different scales (Multiple Choice Questions (MCQ) - 50, Essay- 160 and Oral- 20), they were all harmonized on a 20-point scale. After the data cleaning the data for impurities and harmonizing scores of the papers to a 20-point scale, the variables were then all described using measures of central tendency such (mean and median) and measures of dispersion (Minimum, Maximum, Standard Deviation and Skewness).

The variables were screened for normality and the normality assumption did not fail (Kolmogorov Smirnov and Shapiro Wilk ($P > 0.05$) and this was supported by the fact that the mean and the median were very close and the distributions somehow followed a normal shape. The Paired Samples Test was used to assess the significance of deviation between the written and oral examinations.

Description of student performances



| | MCQ (paper 1)/20 | Essay (paper 2)/20 | MCQ and essay (Paper 1 & 2)/20 | Oral examination/20 |
|----------------|------------------|--------------------|--------------------------------|---------------------|
| Mean | 10.72 | 8.66 | 9.69 | 14.02 |
| Median | 10.80 | 9.13 | 9.84 | 14.00 |
| Minimum | .0 | .00 | 3.20 | 0 |
| Maximum | 16.4 | 16.50 | 14.41 | 20 |
| Std. Deviation | 2.47 | 3.30 | 1.89 | 3.13 |
| Skewness | -.64 | -.52 | -.50 | -1.49 |

Figure 1: Description of Student Performances

Figure 1 describes student performances. The distribution of student performances on the MCQ, essay and oral examinations approximately follow a normal distribution as they follow a Gaussian curve with little deviation and this justifies the use of parametric tests in group comparison and correlation. This is supported by the means and median values which are very close to each other. In MCQ, the mean value was 10.72 and the median 10.80, for the essay the mean was 8.66 and the median 9.13 while in the oral, the mean value was 14.02 and the median 14.00. The average score increased significantly from 8.66 for the essay, 10.72 for the MCQ to 14.02 at the oral. The distributions were generally negative or left skewed, indicating that the asymmetric tails extended toward more negative values with relatively more people scoring above the cohort average in the three tests. The value of Skewness was highest for the oral examination (-1.49) as compared to -0.50 for MCQ and Essay, therefore implying that students really performed poorer at the written examination (MCQ and Essay) as compared to the oral examination, and this is supported by the mean differences presented on Table 1 (PST: $P < 0.05$).

Table 1: Assessing the Significance of Deviation between Written and Oral Examinations

| | | Paired Differences | t | df | P-values |
|--------|--|--------------------|---------|-----|----------|
| | | Mean | | | |
| Pair 1 | MCQ (Paper 1) 20 - Oral examination/20 | -3.3021 | -23.393 | 790 | 0.000 |
| Pair 2 | Essay (Paper 2) /20 - Oral examination/20 | -5.35910 | -32.478 | 790 | 0.000 |
| Pair 3 | MCQ and essay (Paper 1 & 2)/20 - Oral examination/20 | -4.33046 | -33.165 | 790 | 0.000 |

Deviation in percentage

Table 2: Distribution of Deviation in Percentage

| Stats | Difference between oral and MCQ (Paper 1) in percentage | Difference between oral and essay (Paper 2) in percentage | Difference between oral and Paper 1 & 2 (MCQ and essay) in percentage |
|----------------|---|---|---|
| Geometric Mean | 15.85 | 15.5443 | 13.6150 |
| Median | 19.00 | 18.3333 | 17.4167 |
| Minimum | 1 | .83 | .08 |
| Maximum | 90 | 84.17 | 65.25 |
| Std. Deviation | 14.499 | 18.87807 | 14.36665 |

The average deviations of scores were 15.85% between oral and MCQ (Paper 1), 15.54% between oral and essay (Paper 2) and 13.61% between oral and MCQ and essay

(Paper 1 & 2) implying that candidates performed significantly higher in the oral examination than in the other tests.

Assessing consistency and reliability

Table 3: Reliability statistics

| Cronbach's Alpha ^a | N of Items |
|-------------------------------|------------|
| -0.085 | 3 |

The Cronbach's Alpha reliability coefficient was very small (-0.085), indicating that the internal consistency assumption was violated. The negative value is due to a negative average covariance among items, therefore violating the reliability model assumption. This is a clear indication that the trend of scores within cases and across test types is chaotically distributed or does not follow any reasonable pattern. This is supported by the Tukey's test of additivity that indicates no multiplicative interaction ($P=0.986$, >0.05). The intra-class correlation coefficient that produces measures of consistency or agreement of values within cases was almost null ($R=0.027$, $P=0.905$) and by so doing, strengthened the violation of the internal consistency assumption.

The correlation table below equally indicates no correlation between the written examination and the oral examination, with correlation coefficients all very close to 0 ($P>0.05$) as indicated in Table 4.

Table 4: Relationship between written and oral examinations

| | Pearson Correlation | Oral examination |
|-----------------------------|---------------------|------------------|
| MCQ (Paper 1) | R | 0.011 |
| | P-value | 0.764 |
| | N | 791 |
| Essay (Paper 2) | R | -0.041 |
| | P-value | 0.249 |
| | N | 791 |
| MCQ and essay (Paper 1 & 2) | R | -0.007 |
| | P-value | 0.835 |
| | N | 791 |

Nature of change

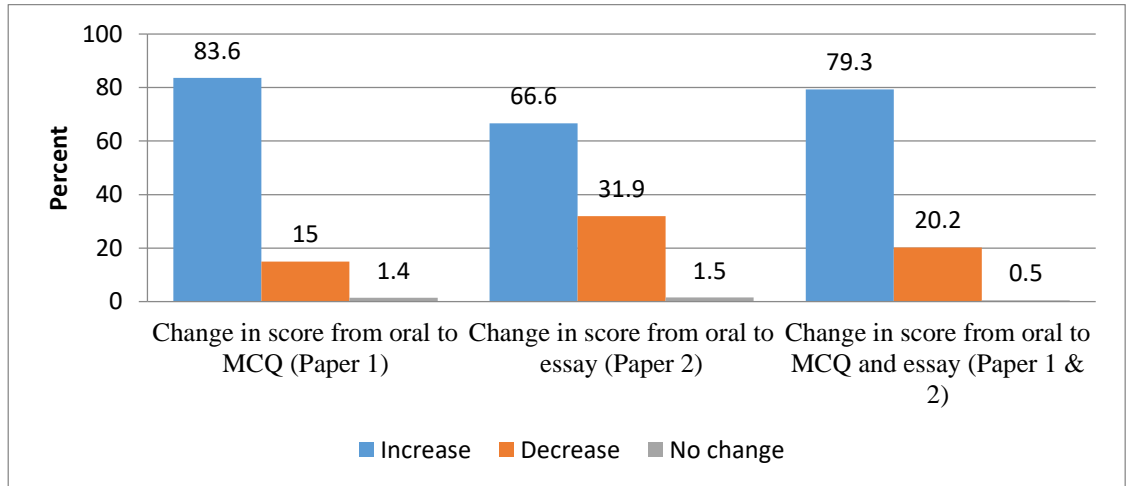


Figure 2: Distribution of nature of change between written and oral examinations

Dominantly, candidate scores increased during the oral examination with proportions of 83.60%, 66.50% and 79.30% for the Multiple Choice Questions (MCQ), Essay and the MCQ & Essay combined.

Conclusion

The concerns of fairness, validity and reliability in this examination have been confirmed by the analysis. There was a clear indication that the trend of scores within cases

and across test types was chaotically distributed or did not follow any reasonable pattern.

Recommendations

- The Cameroon General Certificate of Education (GCE) Board should develop and implement a rigorous framework that will define the construct to be measured, the marking and grading criteria.
- The 25% weighting ascribed to the oral component of SBE French examination be reduced to 10% to align with all practical examinations at the Cameroon GCE Board.
- To curb the element of bias, the Board should ensure that teachers charged with the conduct of the oral component of the SBE French examination should not be practicing French teachers of the division where they are assigned for the exercise. This should be in response to Joughin's worries about examiners' being influenced by the student's gender, ethnicity and socio-economic background to inflate the scores.
- The personnel involved with the conduct of the oral component of the SBE French examination should undergo professional development courses, during which the implications of the extraneous variables likely to affect the validity and reliability of the scores would be explained to them.

REFERENCES

Ayafor, I.M. (2002a). "Trilingualism within Official Bilingualism: The acquisition and use of English in Cameroon." In *Third Language Acquisition and Trilingualism: Proceedings of a*

Conference. FreskeAkademy: Marc Hooghiemstra.

Bobda, S.A. & Tiomajou, D. (1995). Integrating ESL and EFL : The Cameroon Experience. In the British Council, Senegal. *Across the West African divide. Proceedings of the West African English conference, Mbour, Senegal, 12-14 Dec.; 60-83.*

Chumbow, B.S. (1980). "Language and Language Policy in Cameroon" in Kofele-Kale, N. (ed.). *An Experiment in Nation-building: The Bilingual Cameroon Republic since Reunification*. 281-311. Boulder, co: Westview Press.

Hoffmann, C. (1991). *Introduction to Bilingualism*. London: Longman.

Joughin, G. (2008). Oral assessment from the learner's perspective. Saarbrücken: VDM.

Kehm, B. (2001) Oral examinations at German universities. *Assessment in Education* 8, 25-31.

Mackenzie, L. (2000). Occupational therapy students as peer assessors in viva examinations. *Assessment and Evaluation in Higher Education* 25(2), 135-147.

Ong, W. (2002). Orality and literacy: The technologizing of the word (2nd edn). New York: Routledge.

Tadadjeu, M. (1985) "For a Policy of Cameroonian Linguistic Integration: Extensive Trilingualism; The Cultural identity of Cameroon.

Authentic Assessment of Candidates with Profound Hearing Impairment in Practical Subjects at Ordinary Level: Is Zimbabwe getting it Right?

Mwale Frederick*¹

frederick_mwale@rocketmail.com

Zimbabwe School Examinations Council

Abstract

Teachers of students with Profound Hearing Impairment in Zimbabwe believe that these students perform well in practical subjects. Consequently, the three main special schools that enrol children with deafness in Zimbabwe have adopted a curriculum that is practical oriented. Despite these efforts, students with profound hearing impairment continue to fail or pass very normally. This research sought to establish authentic methods of performance measurement in practical subjects that mitigate the plight and needs of candidates with profound hearing impairment. The research focussed on Art, Woodwork, Fashion and Fabrics, Food and Nutrition, and Metalwork. The objective was to come up with assessment models that empower candidates with profound hearing impairment. The researchers collected marks for the school based assessment from two main special schools that enrol candidates with hearing impairment in Zimbabwe. Two sets of comparisons were made, one on raw marks and the other, between raw marks and weighted marks of the practical components. An analysis of students' artefacts was also done to ascertain how they compared with performance in the respective theory components. The researchers held focus group discussions and interviews with teachers and subject managers to solicit their opinion on how authentic practical subjects should be assessed. The findings were that candidates with hearing impairment performed best in the practical tests and worst in the theory components. Language was identified as the cause of low passes in the theory component. All subjects reflected that the high weighting in theory components affected the final mark in all the practical subjects. The researchers recommended that the coursework should be presented as a folio a collection of artefacts. A higher weighting should be recorded in practical test and the folio. The Theory components should then fetch the least weighting. The Theory components could take the form of a dialogue where candidates are asked and respond to questions in sign language to a note taker or scribe under video coverage.

Key words: *authentic assessment, quality life, folios, profound hearing impairment, Sign Language, scribe*

Background to the Study

The education of students with Hearing Impairment (HI) at Ordinary Level in Zimbabwe has failed to yield the intended objective of qualifying these students for a quality life after school. High dropouts are observed at grade seven, leaving very few students proceeding to secondary level. Those who drop out are relegated to the streets where they either beg or become vendors. Students with HI who proceed to secondary school level are faced with curriculum and assessment challenges. A straight

jacket assessment procedure has militated against successful completion of secondary school by students with HI. Over the years, high failure rate has been recorded in secondary schools that enrol these students. Teachers of students with HI in Zimbabwe blame the academic oriented curriculum for the failure rate.

In their quest to pursue a user friendly curriculum, schools that enrol students with HI in Zimbabwe have adopted a curriculum that is practical oriented. The stance has been influenced by the general belief among teachers

of students with Profound HI in Zimbabwe that these students perform better in practical skills than in any other component. The belief is anchored on the compensatory principle which encourages sharpening of the remaining senses and development of functional skills in students with HI for independent and quality life. The recommendation that every secondary school learner must do at least one practical subject and the subsequent circular which turned the recommendation into policy were a welcome development in special schools to cement the teachers' beliefs. (Nziramasanga Commission, 1999 and Secretary Circular Minute number 2 of 2006).

Despite these efforts, students with profound Hearing Impairment have either continued to fail or pass very normally. Questions then are asked: Are students with hearing impairment really gifted in practical skills? Are teaching methodologies appropriate to the learning of practical skills by students with HI? Lastly, how appropriate are the assessment procedures used by the Zimbabwe School Examinations Council (ZIMSEC) in measuring these candidates' performance in practical subjects?

This research sought to establish authentic methods of performance measurement that mitigate the plight and needs of candidates with profound HI in practical subjects. The research focussed on Art, Fashion and Fabrics, Food and Nutrition and Woodwork.

Statement of the Problem

Special schools adopted a curriculum that is practical oriented in a bid to mitigate the plight of students with HI. However very few students are proceeding for tertiary education and there are no students for practical subjects at Advanced Level. This could be evidence of poor performance in examinations at Ordinary level. The researcher therefore sought to interrogate Zimbabwe School Examinations Council's assessment procedures

that are meant to measure performance of candidates with HI in these subjects at Ordinary level.

Objective of the Study

The objectives of the study are to

- interrogate the appropriateness of ZIMSEC's assessment procedures in measuring performance of candidates with profound HI in practical subjects.
- suggest an assessment model that accurately measures practical subjects for candidates with profound hearing impairment at ordinary level.

Research Questions

- How is ZIMSEC assessing candidates with Profound HI in practical subjects at ordinary level?
- How authentic are ZIMSEC's assessment procedures in measuring performance of candidates with Profound Hearing Impairment in practical skills?
- What is the most appropriate assessment model that can measure accurately the performance of candidates with HI in practical subjects?

Significance of the Study

The results of the study were meant to reorient examinations bodies to come up with assessment models that accurately measure performance of candidates with Profound Hearing Impairment in practical subjects. The assessment model should narrow the gap between candidates with Hearing Impairment's true ability and observed scores thereby reducing the error score. The results should also influence practitioners and policy makers in their perception and treatment of practical subjects in schools and the field of work.

Literature Review

Conceptual Framework

This research was grounded on the **Rights Model** and the researchers were inspired by universal values which are informed by global experiences and knowledge and guided by the United Nations Convention on the Rights of People with Disabilities (The African Report on Children with Disabilities, 2014). People with disabilities, according to the convention, have a right to education and ultimately to fair and flexible assessment. Equity and equalisation of opportunities should be applied to education and assessment **of all, and for all** students for quality life. Reasonable adjustments, necessary and appropriate modifications in assessment are a corollary of the philosophy of equalisation of opportunities. Special considerations, adjustments and modifications should not be taken as acts of charity; rather they should be treated as fundamental freedoms and human rights issues of all students including those with Hearing Impairment. If assessment is a right, it should not impose a disproportionate or undue burden on a particular section of students while considered fair in other circles. Assessment under the rights model must be barrier free, valid in terms of satisfying particular assessment objectives, meeting the socioeconomic demands, creating a citizen who is productive and reliable. This perspective is in line with the philosophy of Vocationalism which is an educational philosophy or pedagogy claiming that the content of the curriculum should be governed by its occupational or industrial utility and marketability as human capital. Appropriate curriculum (that which is practical oriented) and an accurate performance measurement procedure definitely produce a citizen who is a good fit in his/her society in all respects.

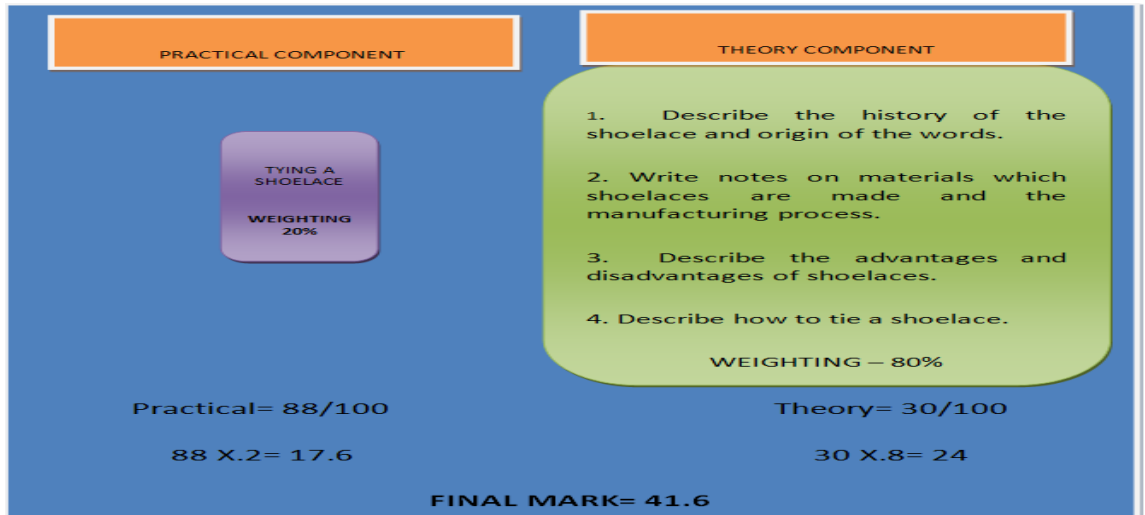
Philosophy influencing practical skills for The Deaf
Students with Profound Hearing Impairment need to be provided with opportunities for career education which include academic instruction, daily living skills, community experiences and vocational experiences.

Anamoah Committee Report on Education (2011) promulgates that education should now focus on inculcating in students skills and appreciation of the use of hands as well as the mind to make students creative and production oriented. This philosophy is inclined to the Vocationalist perspective which advocates that schools should be designed to develop people who are able to meet the demands that society places on its members (ACTE, 2010). They argue that theory is important but at some point you have to “get your hands dirty” to develop the physical control you need to do things and mastership. Practical skills help one to solve problems and create new things. Mapolisa and Tshabalala (2013) concurred by saying that technical vocational subjects are meant to give candidates real-life skills and enhance transition from school to work for many students. They observed that technical vocational subjects offered in schools need to be retained and in fact increased. In the same vein, education of the deaf should focus on making them self-supporting and wean them from being liabilities; help them lead a “normal” life, compare favorably to other human beings and contribute economically to the community in which they live. The application of the knowledge and skills gained through practical subjects or vocational education by deaf students is to help them become more independent through self employment or employment in industry.

Assessment of practical skills

Thompson (1979) concluded that although practical work was an essential component of courses, there was a persistent suspicion among examiners about the reliability of the results for practical assessments. The scholar observed reluctance on the part of the examiners to make the weighting of the practical work too large even though it was considered important. Thompson further hinted that assessment of what candidates can do is at least important if not more important than assessment of what he/she knows. This however is contrary to the belief widely held which assumes a strong relationship between the knowledge about the task and the actual performance.

Consequently theory has been put on the pedestal even in the assessment of candidates with deafness in practical subjects. See the illustration below.



Galway in Harden and Cairncross (2006) however pointed out that there is a poor correlation between a musician's theoretical knowledge of music and his ability to play the instrument and that many of the greatest pieces of music would not have been created if the composers had been concerned with the theory rather than the practice of composition.

Harden and Cairncross (2006) observed that the assessment of practical skills is often neglected and the contributing factor is the unsatisfactory nature of the assessment instruments commonly used. The two scholars propounded the objective structured practical examination which proffers a practical, reliable and valid alternative. The approach is meant to separate assessment of process and product through observation of performance and assessment of end result. They argue it is critical to provide adequate sampling of skills and content to be tested. They encourage an analytic and objective approach to assessment. City and Guilds (2003) suggest that assessors should be clear of learning and assessment objectives; on what has to be seen, produced, observed, and explained and under what conditions. Those assessing practical skills must be scrupulously clear about the criteria being used, and must be confident that any other assessor should come to the same evaluation given the same evidence, (Brown and Pickford, 2006). They compel assessors to write learning outcomes that lend themselves easily to practical assessment so that both students and fellow markers understand what is required. Brown and Pickford (2006) warned that assessors must make sure they are measuring exactly what they intend to measure. Measuring learning requires one to find a way 'not only to get at and assess the process but also to enable students to articulate their understanding of what they have achieved and how', (Kleiman, 2008).

City and Guilds (2003) advocated for assessing competences over time and observed that this has an advantage of having both a high face and predictive validity. The fact that the candidate has always been seen doing the task to the correct standard becomes highly

likely that the candidate will continue to do so in future. Brown and Pickford (2006) concurred to observing and reviewing performance on several occasions to assure consistency and added that keeping good systematic records for quality assurance and clarifying benchmarks and standards with fellow assessors is of great importance. City and Guilds (2003) advocated that assessment of practical skills should take place in the same location as the teaching and the assessors are likely to be the candidates' teachers or instructors. They also suggested that the teaching centres must have specially developed facilities that can be used to enable assessment of practical skills in an environment that is as close to work as possible. Nichol (2008) concurred when he argued that practical skills should always be assessed in a context as close as possible to that in which they will be used. This scenario demands that teachers, head of departments, internal and external assessors be properly trained and qualified. It is the researcher's conviction that once assessment of practical skills is done right, in accordance with what the above scholars suggested, there will not be any need for any special arrangements for candidates with profound hearing impairment.

Methodology

The researchers collected marks on the school-based assessment from two main special schools that enrol candidates with hearing impairment in Zimbabwe. A total of 25 ordinary level students' scores in practical and theory components were collected. A comparison was made on the two sets of raw marks and weighted marks. Hypothesis testing of the two means ($H_0: \mu_t = \mu_p$) of the raw data was done. μ_t and μ_p are the means of the theory and practical respectively. An analysis of students' artefacts was also done to ascertain how they compared with performance in the respective theory components. The researcher held focus group discussions and interviews with teachers and subject managers to solicit their opinion on how authentic practical subjects should be assessed.

DATA PRESENTATION AND ANALYSIS

Table 1 Comparison of theory and practical means

| | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|---|--------------------|----------------|-----------------|---|--------|--------|----|-----------------|
| | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | Lower | Upper | | | |
| woodwork practical mark - woodwork theory mark | 24.250 | 14.607 | 5.164 | 12.038 | 36.462 | 4.696 | 7 | .002 |
| art practical mark - art theory mark | 48.714 | 14.605 | 3.903 | 40.282 | 57.147 | 12.480 | 13 | .000 |
| fashion and fabrics practical mark - fashion and fabric theory mark | 27.636 | 12.902 | 3.890 | 18.969 | 36.304 | 7.104 | 10 | .000 |
| food and nutrition practical mark - food and nutrition theory mark | 19.143 | 14.826 | 3.963 | 10.582 | 27.703 | 4.831 | 13 | .000 |

Observable from **table 1** is that p values are in each case less than 0.05 implying that there was a statistically significant difference between the practical and theory means. The fact that the means are positive indicates that practical components registered a higher mean and students with profound HI scored very highly in the practical components than in the theory ones. The belief among teachers that students with HI perform better in practical subjects was authenticated.

1. H₀: $\mu_{wp} = \mu_{wt}$
H₁: $\mu_{wp} \neq \mu_{wt}$
2. H₀: $\mu_{ap} = \mu_{at}$
H₁: $\mu_{ap} \neq \mu_{at}$
3. H₀: $\mu_{fp} = \mu_{ft}$
H₁: $\mu_{fp} \neq \mu_{ft}$
4. H₀: $\mu_{fnp} = \mu_{fnt}$
H₁: $\mu_{fnp} \neq \mu_{fnt}$

The results of the Focus Group Discussion were presented under the sub headings below.

ZIMSEC assessment procedures in practical subjects at ordinary level

The participants revealed that all the practical subjects selected for the purposes of this research except Food and Nutrition have three papers which include coursework, practical and theory examinations. In Woodwork, Metalwork and Fashion and Fabrics, the teacher assisted course work has 20% weighting. The course work in Fashion and Fabrics comprises one or two garments sewn to the candidate's size. The coursework in Woodwork and Metalwork is concerned with one artefact and a folio. The folio or log book contains the situation for the design, design brief, investigation, solution, realisation and evaluation. Teachers and assessors agreed that the folio stages are concerned with defining the problem,

describing the materials used, highlighting improvements made on the artefact and also the weaknesses and strengths of the artefact. The course work is done in seven months. It was also revealed during the discussion that the maximum mark for the artefact alone is 20 and the other aspects of the folio fetch a maximum of 80 to make a total of 100 in both Woodwork and Metalwork. In a practical examination, candidates are expected to produce part of a construction or garment. The discussion revealed that in each case it is the end product that is assessed and not the process. The practical and theory examinations have equal weights, 50% each in food and Nutrition and 40% each in the other three subjects. The weighting shows that although practical work was an essential component of courses, there was a persistent suspicion among examiners about the reliability of the results for practical assessments (Thompson, 1979). The scholar observed reluctance on the part of the examiners to make the weighting of the practical work too large even though it was considered important. Teachers teach to examination specification hence it may not surprise anyone that candidates were doing one artifact for the whole two year course, and this was inadequate to qualify an individual as proficient in the respective subject.

The authenticity of ZIMSEC's assessment procedures at ordinary level

Participants unanimously agreed that the assessment of candidates with profound HI was not authentic since there was no procedure tailor made to meet the assessment needs of candidates with HI. The discussion highlighted that students with HI excel in the practical component to an extent that even the Ministry of Primary and Secondary Education collects the artefacts to exhibit during the highly esteemed Trade Fair Show. The distinction work that is observed in the artefacts is affected by the theory component which is rich in the English language. Students with HI have a deficiency in English consequently they experience difficulties in accessing the theory component. The grades are always lower than their actual performance. This was also

evident in the marks collected for the various tests in **Tables 1.**

Graduates with profound HI are not selected for tertiary education since there is high competition with the main stream candidates who may be strong in theory, yet not so strong in the practical component. The intended purpose that the knowledge and skills gained through practical subjects or vocational education by deaf students could help them become more independent through employment in industry or self employment is defeated. The philosophy behind the inclusion of practical subjects in the curriculum is skilling the students so that they are employable or self reliant, therefore if those with the skills are not rewarded accordingly because they failed theory then the assessment objectives need to be realigned.

The discussion revealed that candidates with HI are expected to respond in writing and even with the correct knowledge but they fail to express themselves. ZIMSEC assessors, on the other hand do not understand the language and hence the marking prejudices the candidates. The Focus Group Discussion also revealed that the coursework part which requires the description and evaluation of the work done also grossly affected performance of candidates with HI in that component. It should be noted then that the assessment procedure that may be deemed fair by the mainstream is actually handicapping students with HI because of language.

Suggestions on appropriate assessment procedures for candidates with HI in practical subjects at Ordinary level

Teachers of students with profound HI suggested that ZIMSEC could modify the language of instruction. Words that are also found in sign language vocabulary must be used in every instance. The glossary of examination terms must be that which is understood by candidates with HI. The group discussion also suggested that paper weighting should be increased or raised in the

practical component. Sign language should be used in the theory component. If candidates with HI expressed their responses in writing then there has to be transcribers who should re-write the script in the expected language and both scripts are forwarded for marking.

Conclusion and Recommendations

From the data that was collected and presented, it was evident that assessment procedures for candidates with HI were failing to accurately measure their performance mostly due to the language used in the theory and coursework components. The assessment instruments were designed by the hearing and with the hearing students in mind consequently the candidates with HI who were distinct in practical components were not awarded their actual grades due to the theory component whose weighting was higher than it should be. Candidates who have been rewarded by the assessment procedures are those who perform well in the theory components yet, as Galway in Harden and Cairncross (1980) pointed out, there is a poor correlation between theory and practical mastery. Education in such a scenario would have failed candidates with HI because according to Mapolisa and Tshabalala (2013), application of the knowledge and skills gained through practical subjects or vocational education by deaf students is meant to help them become more independent through self employment or employment in industry. The results of the research explain why people with HI had failed to benefit from the practical oriented curriculum despite how they have excelled in the practical components. Quality life for people with HI remains a dream until the assessors begin to value practical components and assess them in a manner that is accessible to them.

The researcher recommended that:

- Assessment boards should value the practical components and increase the weighting accordingly. A 30% or less weighting for a theory component suffices because the current economic order in Africa needs people who

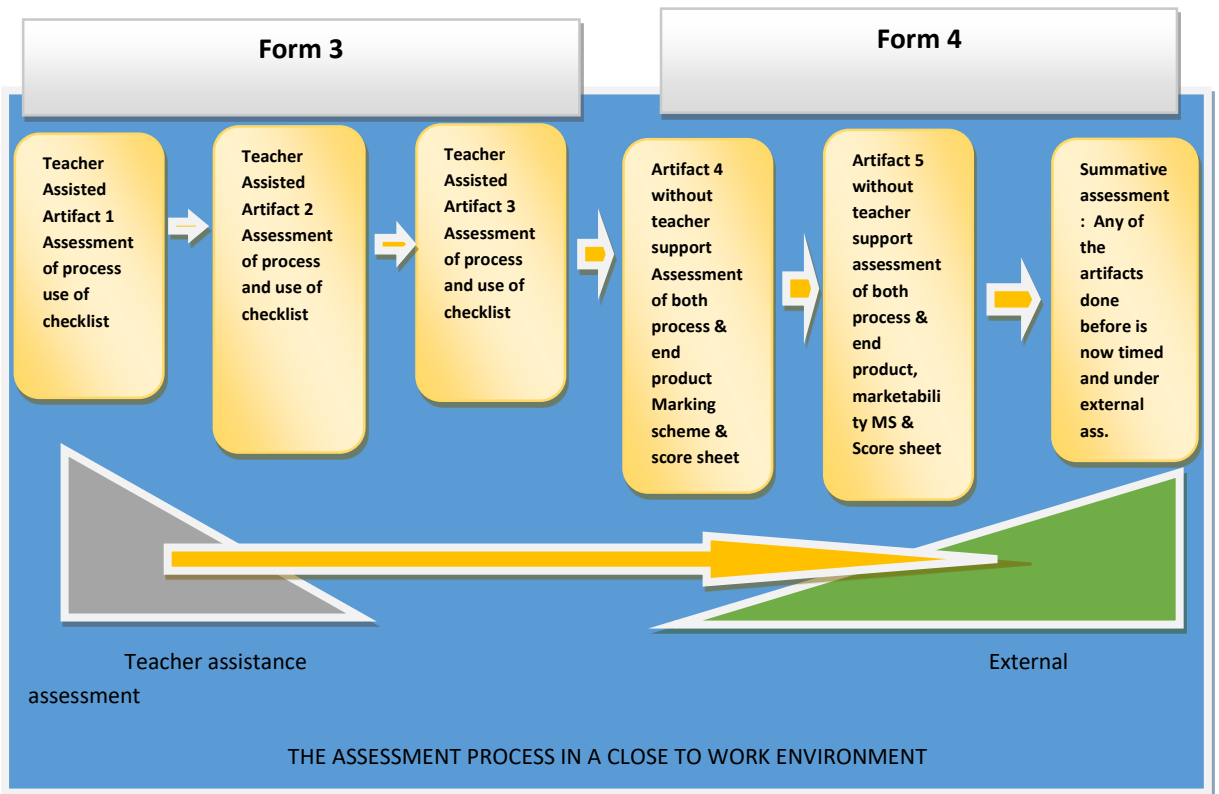
produce things and not those who merely explain and evaluate.

- Assessment boards should embark on rigorous training of assessors to enable them to handle work of candidates with HI in a more accurate manner. Where possible teachers of these candidates must be part of the assessing teams because they have the language and skills that are a prerequisite in the assessment of these candidates.
- Sign Language can never be replaced in the life of THE DEAF and it should be the language of instruction and assessment. Examination boards should allow candidates with HI to sign their responses to a scribe or note taker in a theory component under either a video or an

external assessor- let the examination take the form of an interview that qualifies an individual for work or further studies.

- Examination boards should move swiftly to embrace school-based assessment and give it the prominence it deserves. This will work well for students with disabilities in general in that they will be assessed by people who are empathetic.
- The Ministries responsible for Primary and Secondary Education should inject adequate funds in schools so that sheltered workshops are designed and practical subjects are learnt and assessed in an environment that is near to the normal workplace as possible.

The researcher also recommended a model of assessment shown below:



References

- ACPF, (2014) The African Report on Children with Disabilities Promising Starts and Persisting Challenges. Central Printing Press. Addis Ababa.
- ACTE, (2010). Up to the Challenge: The Role of Career and Technical Education and 21st Century Skills in College and Career Readiness. www.pzl.org/storage/documents/CTE_oct2010.pdf
- Anamoah Committee Report on Education (2011) Meeting the Challenges of Education in the Twenty-First Century. <http://ihdl.handle.net/123456789/728>
- Brown, S and Pickford, R (2006) Assessing and Practice: Key guides for Effective Teaching in Higher Education. Routledge, Amazon.
- City and Guilds, (2003) A Guide to Assessment of Practical Skills in International Vocational Qualifications, www.cityandguilds.com
- Harden, R. M. and Cairncross, R. G., (1980) Studies in Higher Education Volume 5, Issue 2, 1980
- Kleiman, P (2008). Assessment Criteria for Practical Skills: The Times Higher Education 27 March 2008. www.timeshighereducation.co.uk/401208.article
- Mapolisa, T and Tshabalala, T (2013) Attitudes of Teachers Towards The Policy of Teaching Practical/Vocational Subjects, International Journal of Asian Social Science. Journal
- homepage: <http://www.aessweb.com/journal-detail.php?id=5007>
- Nichol, D (2008) Assessment Criteria for Practical Skills: The Times Higher Education 27 March 2008. www.timeshighereducation.co.uk/401208.article
- Nziramasanga Commission, (1999) Report of the Presidential Commission of Inquiry into Education and Training Government Printers, Zimbabwe. Secretary Circular Minute number 2 of 2006, Zimbabwe
- Thompson (1979) Assessment of Practical Skills: The Objective Structured Practical Examination (OSPE). Routledge.

SUB THEME E: INTEGRATING SOFT SKILLS ASSESSMENT IN PUBLIC EXAMINATIONS

Perceptions on Integration of Soft Skills Assessment in Public Examinations: The Case of Lesotho Basic Education

*Ms Ramokoena Camilla 'Matseko**

Examinations Council of Lesotho

favour.ramokoena@gmail.com

Abstract

This paper reports on the views of some educators in Lesotho, on the teachability and assessment of specific aspects of soft skills. The approach taken was qualitative and a focus group discussion of six participants was employed. The group discussions focused on the teachability and assessment of the “4 C’s” of Creativity, Critical thinking, Communication and Collaboration. The discussions revealed that the afore mentioned soft skills could be taught while assessment of the same skills seem to pose a major challenge to both teachers and assessment bodies. Challenges relating to the competencies of teachers to effectively teach and assess the soft skills were highlighted as being major. One of the policy implications therefore, was the need to capacitate teachers on how to integrate assessment of soft skills during instruction as part of Assessment For Learning.

Key words: Soft skills, deep learning, surface learning, hard skills.

Introduction

The test for education at the start of the 21st century is to train students for life-long learning equipping them with a repertoire of skills such as communication skills, critical thinking skills, problem solving skills, creative skills and others. Globally, the gap between the skills and capabilities of graduates and the demands of the world of work seems to be growing wider and wider. It is evident that fulfilling complex demands and tasks requires not only knowledge and skills but also involves strategies and routines needed to apply the knowledge and skills, as well as appropriate emotions and attitudes, and effective management of these components. Thus, the notion of competencies encompasses cognitive but also motivational, ethical, social, and behavioural components. The crux of soft skills therefore, is the need to integrate, synthesize and creatively apply content knowledge in novel situations. Consequently, soft skills assessments must systematically ask students to apply

content knowledge to critical thinking, problem solving, and analytical tasks throughout their education so that we can help them hone this ability and come to understand that successful learning is as much about the process as it is about facts and figures.

The need for a paradigm shift in assessment with more emphasis on soft skills cannot be over emphasized. The acquisition of soft skills demands students to engage in deep learning as opposed to surface learning. Deep learning can be viewed as a process of examining new facts and ideas critically, tying them to existing cognitive structures and making numerous links between ideas.

Students therefore, demonstrate deep learning by showing an increase in understanding of a subject involving the grasp of core principles; an ability to apply newly understood principles in a variety of different contexts and situations and long-lasting personal change.

The opposite of deep learning is surface learning and this involves accepting new facts and ideas uncritically and attempting to store them as isolated, unconnected items.

In an attempt to reform the education system in Lesotho, a policy framework for curriculum and assessment was developed and published in 2009. The policy framework stipulates that basic education aims at developing functional and permanent literacy and numeracy as well as creative and critical thinking for effective living and life-long learning. It also aims at equipping learners with knowledge, attitudes, and skills which enable them to respond to socio-economic and technological changes (MOET, 2009). The core competencies earmarked in the policy include collaboration, cooperation, and effective and functional communication.

The need to improve acquisition of soft skills as highlighted in the policy on Curriculum and Assessment in Lesotho formed the rationale behind this investigation. There seems to be no record in the country about research studies on the teaching and assessment of soft skills in Lesotho Basic Education

Some Related Literature

What are considered as Soft Skills?

Soft skills is a sociological term relating to a person's Emotional Intelligence Quotient (EQ), the cluster of personality traits, social graces, communication, language, personal habits, friendliness and optimism that characterize relationships with other people (Aworanti, 2012). Examples of Soft Skills include;

- Communication
- Critical and structured thinking
- Problem solving
- Creativity

- Teamwork capability
- Negotiating skills
- Self-management
- Time management

- Conflict management
- Cultural awareness and others.

The 21st century skills are usually categorized under cognitive, interpersonal, and intrapersonal. However, interpersonal skills represent different abilities such as knowledge of social customs and the capacity to solve problems associated with social expectations and interactions. Non-cognitive skills are those attitudes, behaviours, and strategies which facilitate success in school and workplace, such as motivation, perseverance, and self-control while hard skills are related to technical aspects of performing a job (Aworanti, 2012). These skills usually require the acquisition of knowledge, are primarily cognitive in nature, and are influenced by an individual's intelligence quotient score.

Blaga (2014) makes a distinction between what constitutes competency (cognitive) as opposed to the dimensions of behavior (non-cognitive). According to Blaga (2014), competency encompasses aptitude, knowledge and skill while behavior entails beliefs, attitudes and action.

Taking creativity and communication as examples, it is very important to consider some literature on these aspects. Treffinger (2002) talks about the four categories of personal creativity characteristics as depicted on figure 1.

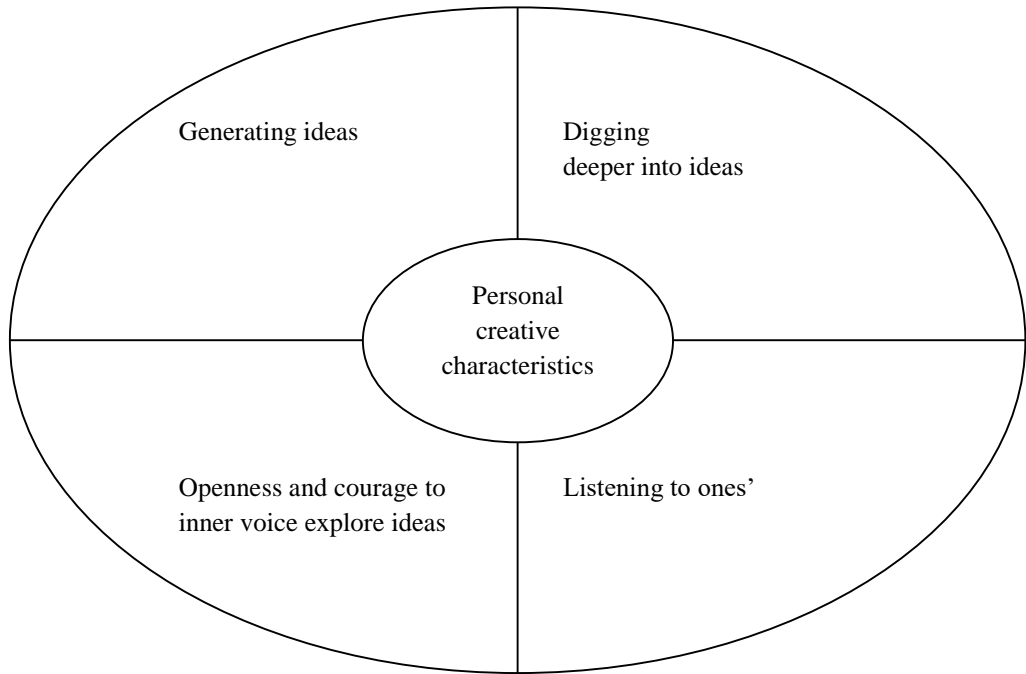


Figure 1: Four categories of personal creativity characteristics

The analysis of the categories indicate that the four categories listed do not necessarily follow any chronological order but of most significance is mind engagement which can be clearly deduced. Barbot (2011), perceives creativity as being a function of conative factors. Barbot (2011), outlines some factors as core factors in creativity. The core factors include perseverance, tolerance for ambiguity, openness to new experiences and others.

Communication is defined by Hardani (2012) as a meaningful act of exchanging information between two or more parties. Hardani (2012) explains the nature of communication in terms of assertive, aggressive and non-assertive.

Effective communication can be inhibited or advanced by a number of things which includes social norms, culture and language. Lack of common vocabulary can also result in limited range of emotions which can be clearly displayed. Even though two of the four aspects of soft skills have been used as examples here, the assessment of all the four and their acquisition can be placed on the same continuum.

Contextualising Assessment and Learning Hard and Soft Skills

The State of Victoria (2004) defines learning as an active process in which students develop new ideas and concepts based upon their existing knowledge. In experiential learning, learning is about meaningful experiences that result in a change in individuals'

knowledge and behaviour. According to Crick, Broadfoot and Claxton (2004) what is learned counts as knowledge or skill, which can take the form of the ability to do something which could not be done before; or a new understanding about the world; or something of spiritual, emotional, or aesthetic significance.

Vermunt and Vermetten (2004) identify three components of learning as being: cognitive, affective and regulation. Cognitive processing activities refer to those thinking activities that students use to process subject matter. Affective activities involve emotions that arise during learning and lead to affective states that may positively, neutrally, or negatively influence the progress of a learning process. Regulation activities, on the other hand, steer the cognitive and affective activities and therefore, indirectly lead to learning outcomes (Vermunt&Vermetten 2004).

According to the State of Victoria (2004) the product of learning includes knowledge and intellectual skills, attitudes and emotional responses, social behaviour and movement skills. It continues that through our interactions with our unique personal environments we learn to think, act and feel in ways that contribute to our individual human identities. The State of Victoria (2004) added through learning we recreate ourselves, we become able to do something we were never able to do and through learning we extend our capacity to create and to be part of the generative process of life.

Constructivism and Soft Skills Development

Biggs (2003) regards constructivism as a philosophy of learning founded on the premise that, by reflecting on their experiences, students construct their own understanding of the world they live in. Moreover, Biggs contends that assessment should be made part of the learning process, ensuring that it provides students with information on the quality of their learning.

The role of assessment in Constructivism is further highlighted by Geysers (2004) in that assessment can foster learning when the mode of assessment, rather than the objectives of the curriculum content determines the nature of the understanding the student derives from the taught content. Learning should be lively and productive; students are not to behave like inactive receivers of information, but should be seen as active participants through interaction with the environment and through the reorganization of their own mental structures. The interrelationship of pedagogy, student assessment and curriculum content is vital.

Greece (2002) further contends that in Constructivism, the student constructs his/her own conceptualisations and finds his/her own solutions to problems, thereby mastering autonomy and independence. The context within which learning takes place, the students' beliefs and attitudes are crucial and so are the students' experiences. Constructivism also tallies with experiential learning. In principle, experiential learning involves personal involvement, student initiation, evaluation by the student and it has a profound effect on the student (Greece 2002).

Having looked at the constructivists approach and its focus, the task of aligning teaching, learning and assessment cannot be overlooked.

Constructive alignment

As stated by Seigel (2004) a good teaching system aligns teaching method and assessment to the learning activities stated in the objectives to support appropriate student learning. Thus Seigel (2004) perceived assessment as the "senior partner in learning and teaching", but an inaccurate assessment could counteract everything. In addition, Jervis and Jervis (2005) stipulated that in constructing aligned teaching, first it is necessary to specify the desired level or levels of understanding of the content in question. The model in Figure 2 shows how

assessment could be aligned to other aspects of teaching and learning (Brown, 2001).

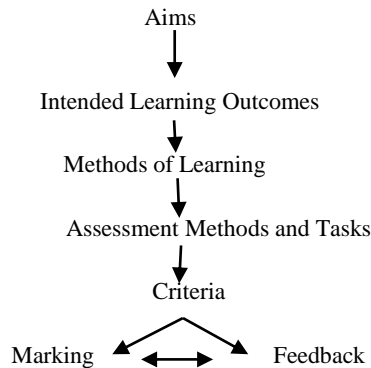


Figure 2: Model of Constructive Alignment.

Rust (2002) postulates that a shift in thinking about the process of instruction design has led to the following three-stage model: the first step is to identify clear learning outcomes; the second is to design appropriate assessment tasks that will directly assess whether each of the learning outcomes has been met; and the third step is to design appropriate learning opportunities for the students to get them to a point where they can successfully undertake the assessment tasks.

However, Rust (2002) argued that there is a lag between changes in teaching methods and changes in assessment as well as patchy changes in assessment. Furthermore, Brown (2001: 4) stated that effective assessment methods and tasks should be related to the learning outcomes and the methods of learning.

According to Biggs (2003) constructive alignment enables the students to do the real work with the teacher simply acting as a broker between the students and the learning environment that supports the appropriate learning activities. In order to help students to pace their learning and to engage seriously with the material from the start, there is a need to build in regular assessment tasks (Rust, 2002). Biggs goes on to indicate that,

education is about conceptual change and not simply the acquisition of information. The argument here is that attempts to create a need to learn by the use of ill-conceived and urgent assessment are counter productive, with the result being that students merely respond to assessment tasks without engaging in the tasks in any depth. Hence defeats the intentions to develop the soft skills.

Methodology

The approach taken was qualitative and a focus group discussion of six participants was employed. The six participants were purposively selected. The group discussion focused on the teachability and assessment of the “4 C’s” thus; Creativity, Critical thinking, Communication and Collaboration. The “4 Cs” provided main themes used to guide the discussion. Note-based analysis was carried out on the notes from the focus group, the debriefing session, and the summary comments from the moderator.

Findings

The group discussed a range of experiences that they felt impeded effective teaching and assessment of soft skills in basic education. One of the challenges of traditional

pedagogy is that of using quizzes, exams, and assignments to assess individual performance. These quizzes or exams cannot accurately measure interpersonal and leadership skills (Fadel, 2012). There is also little in terms of teacher expertise in combining knowledge and skills in a coherent ensemble, with guiding materials, and assessments. The inexperience of teachers coupled with the curriculum that is overburdened with content, makes it much harder for students to acquire (and teachers to teach) skills fostering creativity and that which establish lifelong learning habits through deep learning (Fadel, 2012).

The group itemized some of the key challenges in assessing interpersonal skills which includes;

- the accuracy with which interpersonal expertise can be measured.
- the significance of context in which assessments of interpersonal skills are administered, the group noted the complexity of trying to take the context into account in assessment
- acquisition of soft skills involves some interaction with other people, hence individualized testing makes it almost impossible to standardize the assessment procedures.

Perception on the integration of Soft Skills in Public Examinations

It was evident during the group discussion, that integration of soft skills in public examinations poses a major challenge to the assessors. One of the challenges had to do with the different codes of morality depending on the indoctrination of learners. The many factors which come into play while setting or attempting to enforce certain moral codes were seen to create a huge dilemma in developing valid, reliable and fair assessment criteria for certain aspects of soft skills. Issues around culture, religion, age, gender, language competence and others

were also regarded as major determining factors in demonstrating the acquisition of soft skills.

Discussions

It is evident that skills are not stable characteristics (traits), but rather the demonstration of an appropriate performance in particular contextual/situational condition, despite the fact that this performance is only carried based on the prior existence and combination of personal and contextual resources. This purports that skills acquisition, development and expression (or inhibition) depends at all times both on personal characteristics, such as contextual or situational characteristics, and on the dynamic interaction between the two (Kechagias, 2011).

According to Kechagias (2011), fulfilling complex demands and tasks requires not only knowledge and skills but also involves strategies and routines needed to apply the knowledge and skills, as well as appropriate emotions and attitudes, and effective management of these components. Thus, the notion of competencies encompasses cognitive but also motivational, ethical, social, and behavioural components. Concerns have been raised about the distortion of the curriculum as a result of over-concentration on the areas of the curriculum subject to high stake testing. Teachers are inclined to rely on drill, and may expect students to engage in surface learning strategies (such as memorizing, rehearsing, and rote learning) as opposed to deep learning. This in itself inhibits the acquisition of soft skills.

One other major dilemma relates to the existing definitions and taxonomies of interpersonal skills which were developed in the context of face to face interactions, while new technologies tend to foster interactions that do not occur face to face. In this era of advancements in technology, the task of building interpersonal skills in students therefore becomes quite a mammoth one. The escalating use of cell phones for communication either with phone calls or by text messaging is having a very

negative toll on the interpersonal relationships, both at personal level and also at professional level. Communication in the 21st century is such that people communicate with their hands instead of their mouths whether by emails, texting, social networking, instant messaging, or Skype. It has become a common phenomenon for people to spend days or months without coming face-to-face with another person, while still remaining connected with the rest of the world. People seem to spend copious hours on the internet resulting in what is newly termed “internet addiction disorder”.

The personal touch brought about by a smile or handshake is gradually becoming a thing of the past. People are slowly losing the importance of non-verbal communication such as gestures, body language and facial expression in a conversation. Text messages or emails can be interpreted differently as they lack human emotions. Most people nowadays no longer talk to each other instead they prefer texting messages to each other even if they are in the same room.

Despite the challenges surrounding the teaching and assessment of soft skills, it is evident that as the world continues to evolve, acquisition of these skills cannot be left to chance.

Conclusion and Recommendations

The group had some consensus that higher-order skills (“21st Century Skills”), such as the “4 C’s” of Creativity, Critical thinking, Communication, Collaboration, and others are essential for absorbing knowledge as well as for work performance (Fadel, 2012). There was also an agreement amongst the group on what the soft skills are, and how teaching methods by way of authentic tasks can affect acquisition of soft skills, but there were concerns raised regarding limited time available during the school year, given the overloaded curriculum (Fadel, 2012). The group also acknowledges the need for re-definition of communication in the 21st century.

Consequent upon the findings and conclusion in the study, the following recommendations were made:

1. Assessment of soft skills should form an integral part of continuous assessment at school level.
2. There is need for interventions that would be geared towards improving effective teaching and assessment of soft skills at the basic education level.
3. Teachers’ in-service training on the teaching and assessment of soft skills should be made a priority on the Ministry of Education and Training agenda.
4. The Early Childhood Care and Development (ECCD) curriculum should integrate the teaching of soft skills.
5. Checklists and rubrics should be used to ensure both reliability and fairness in assessing soft skills.

In the absence of a grading process that is reliable and fair in which the students know the grading criteria, what counts, and by how much, the integration of soft skills in public examinations was regarded as farfetched. The recommendation is to capacitate teachers to enable them to effectively integrate assessment of soft skills as an integral part of assessment *For* and *As* learning.

References

- Aworanti, A. (2012). Integration of “Soft Skills” Assessment into public Examining in Technical and Vocational Education (TVE). National Business and Technical Examinations Board. Presented During the 31st Annual Conference of Association of Educational Assessment

- in Africa Held at Botswana. Retrieved from www.nabtebnigeria.org/.../2012/.../INTEGRATION-OF-“SOFT-SKILLS...
- Barbot, B. (2011). Assessing Creativity in the classroom-Bentham open Retrieved
- Biggs, J. (2003). *Teaching for quality learning at university*. 2nd edition. Buckingham, UK: Open University Press.
- Blaga, A. (2014). What is the difference between competencies and behaviors when establishing performance criteria? Retrieved from www.performancemagazine.org
- Brown, G. (2001). *Assessment: A Guide for Lecturers*. UK. Learning and Teaching Support Network. (LTSN) Retrieved from <http://www.ltsn.ac.uk>
- Crick, R.; Broadfoot, P. & Claxton, G. 2004. Developing an Effective Lifelong Learning Inventory: the ELLI Project. *Assessment in Education: principles, policy & practice* 11(3): 247-272.
- Fadel, C. (2012). *What should students learn in 21st century?* Center for Curriculum Redesign. Retrieved from <http://oecdeducationtoday.blogspot.com/2012/05/what-should-students-learn-in-21st.html>
- Geysler, H. (2004). Learning from assessment. In S. Gravett & H. Geysler (Eds). *Teaching and Learning in Higher Education*. 90-109. Pretoria: Van Schaik.
- Greece, D.T. (2002). *Constructivist Learning*. Retrieved from <http://www3.telus.net/linguisticsissues/constructivist.html>
- Hardani, W. (2013). The Effectiveness of Communications Hierarchical Organizational Structure. Retrieved from www.ijssh.org
- Jervis, L.M. and Jervis, L. (2005). *What is the Constructivism in Constructive Alignment?* Retrieved http://www.bioscience.heacademy.ac.uk/journal/vol6/Beej_6_5.htm
- Kechagias, K. (2011). Teaching and Assessing Soft Skills - MASS Project Retrieved http://mass.educational-innovation.org/attachments/396_MASS%20wp4%20final%20report%20part-1.pdf
- Ministry of Education and Training. (2009). *Curriculum and Assessment Policy Framework*. The Age Multimedia Publishers: Maseru
- Rust, C. (2002). The impact of assessment on student learning. *Active Learning in Higher Education* 3(2): 145-158.
- Seigel, S. (2004). *Constructive Alignment*. Retrieved <http://www.resources.scalingtheheights.com/Constructive%20Alignment%20and%20Biggs.htm>
- State of Victoria. Department of Education and Training. (2004). *Research on Human learning*. State of Victoria: Melbourne.
- Treffinger, D.J. (2002). *Assessing Creativity-A Guide for Educators*. Retrieved from www.files.eric.ed.gov
- Vermunt, J. and Vermetten Y. (2004). Patterns in Student Learning: Relationships Between Learning Strategies, Conceptions of Learning, and Learning Orientations. *Educational Psychology Review* 16(4): 359-383.

Integrating the 21st Century Skills into the West African Senior School Certificate Examination (WASSCE): Challenges and Opportunities

Olanrewaju Ojo Emupenne

The West African Examinations Council, Nigeria
emupenne@gmail.com; **Skype:** lanre.emupenne
Tel: +2348051100115; 8029043290

ABSTRACT

The dynamism of education cannot be overstressed. It is common knowledge that students' success in school, work and life generally depends on their ability to think outside the box and this is the focus of the 21st Century Education. It is against this background that this study, through quantitative research method, investigated the extent to which the 21st Century skills have been integrated into the teaching, learning and assessment regimes in senior secondary schools in Nigeria. The respondents who comprised 88, 82 and 88 students from urban, semi-urban and rural areas respectively were randomly selected to make a total sample of 258 from Nigeria's six geopolitical zones. The instrument for data collection, a questionnaire made up of 31 items, was administered to them. The study was methodically guided by three research questions viz., to what extent are SS students familiar with the dynamics of 21st Century Skills?; To what extent are SS students engaged in social networking and online resources as a form of interaction? How would the proper grasp of the fundamentals of 21st century skills by SS students serve as causative factor for the assessment of these skills by WAEC? Portions of the data obtained were descriptively analysed using SPSS and Microsoft Excel. A larger percentage of the respondents, more than 30.0% in each case, have never made use of the 21st century thinking and working skills. Similarly, while most of the students, about 56.6%, own cell phones, only few of them, less than 33.3% in each case, own smart phones, iPad/iPhone, tablet/netbook/kindle/nook/e-reader or laptop computer. More so, a larger percentage of the students, more than 40.3% in each case, have an active account with popular online social media through which they engage in social networking. In sum, it was discovered that 21st century skills have not been properly integrated into the Nigerian education system and that the adverse effect of this is enormous. It is based on these findings that the study endeavours to underscore the need for a policy roadmap by hinting at strategic framework for understanding and addressing the 'teaching-learning-assessment of 21st Century skills' question in the Nigerian educational system. Additionally, further research on the role of teachers in the assessment of the 21st century skills is recommended.

Keywords: Assessment, Education, Schools, Skills, Teaching

1. INTRODUCTION

The term '21st century skills' has been conceptualized to mean 'a broad set of knowledge, skills, work habits, and character traits that are believed...to be critically important to success in today's world, particularly in collegiate programs and contemporary careers and workplaces' (Glossary of Education Reform cited in Adlon, 2015). These skills of which we speak can be applied in all academic subject areas, and in all educational, career, and civic settings throughout a student's life.

In the educational discourse across the globe, 21st century skills have gained currency and have been noted to encompass wide-ranging skills that are not easy to define. More so, the term is often used interchangeably with other terms such as 'applied skills', 'cross-curricular skills', 'cross-disciplinary skills', 'interdisciplinary skills', 'transferable skills', 'transversal skills', 'non-cognitive skills', and 'soft skills' (Great School Partnership, 2014; Rajendran, 2015).

It is further noted in edglossary.org that 'while specific skills deemed to be 21st century skills may be defined, categorized, and determined differently from person to person, place to place, or school to school, the term does reflect a general consensus'. The Assessment and Teaching of 21st Century Skills (ATC 21) Organization has offered a framework for organizing different types of 21st century skills (Binkley et al., 2010). This framework includes four classes of skills as follows: ways of thinking (Creative Skills, Critical Thinking Skills, Problem-solving Skills, Decision Making Skills and Learning Skills); ways of working (Communication Skills, Collaboration Skills, Team-working Skills, Networking Skills and Ethics and Etiquette Skills); tools for working (Information, Media and Technology Skills and Information Literacy Skills); skills for living in the world (Civic Literacy/Citizenship, Life and Career Skills, Social Responsibility and Entrepreneurship Skills).

(1) THE CRITIQUE OF 21ST CENTURY SKILLS

It follows the Ecclesiastics philosophical underpinnings that 'the thing that hath been, it is that which shall be; and that which is done is that which shall be done: *and there is no new thing under the sun*' (KJV). A critical analysis of this verse of the Holy Scripture shows that '21st century skills' concept (with the exception of some of the Digital Literacy skills) is an adaptation of some educational postulations from past centuries. To corroborate this assertion, some educationists have made reference to the thoughts of Confucius and Montaigne.

Trilling and Fadel (2009) aver that Confucius recognized the need for learning by doing, quoted as: "*I hear and I forget, I see and I remember, I do and I understand*". Michel de Montaigne said "*Rather a mind well-shaped than well-full*". Additionally, prior to the dawn of the new millennium, some universities had a way of rating their students' performances. For instance, the University of Harvard in the United States of America developed a Students' Rating Scale several years back before the dawn of 21st Century. The rating scale includes:

- i. Creativity, original thought;
- ii. Motivation;
- iii. Independence, initiative;
- iv. Academic achievement;
- v. Effective class discussion;
- vi. Disciplined work habits;
- vii. Potential for growth; and
- viii. Personal qualities and character.

These are some of the items that have been incorporated into 21st Century skills.

(2) ASSESSMENT OF 21ST CENTURY SKILLS BY WAEC

The critique above notwithstanding, 21st century skills are very germane to the success of students in schools and at workplaces after graduation. These skills need to be

taught and assessed. Examination bodies such as the West African Examinations Council (WAEC) have a great role to play with respect to the assessment of 21st century skills. WAEC as a regional examination body conducts the West African Senior School Certificate Examination (WASSCE) in Anglophone West African countries, namely Ghana, Liberia, Nigeria, Sierra Leone and the Gambia as well as in some schools in a few Francophone countries in the sub region that use English as medium of instruction and examination.

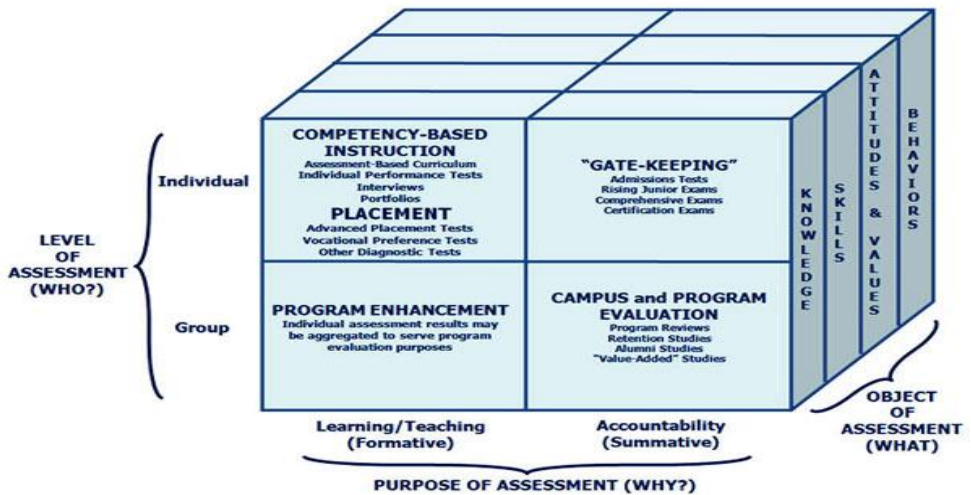
The sole aim of WASSCE is the assessment of final year students in senior secondary/high schools.

The utmost significance of assessment to learning cannot be overstated considering the fact that it is ‘the process of gathering and discussing information from multiple and

diverse sources in order to develop a deep understanding of what students know, understand, and can do with their knowledge as a result of their educational experiences’ (Huba and Freed, 2000). In corroborating Huba and Freed’s view, Erwin (1991:15) opines that ‘assessment is the systematic basis for making inferences about the learning and development of students. It is the process of defining, selecting, designing, collecting, analyzing, interpreting, and using information to increase students’ learning and development’.

Different approaches to assessment procedures of test development, construction, and administration have been developed by scholars and examination bodies. These are tacitly captured by Patrick Terenzini’s taxonomy as represented thus:

Figure 1: Assessment Taxonomy



Source: Patrick Terenzini, 1989.

It should be borne in mind that assessment of 21st century skills incorporates both formative assessment and summative assessment. The former is inextricably interwoven with teaching and learning since it 'contributes to learning by providing a feedback' while the latter shows 'the extent of a learner's success in'. The remaining parts of the paper unfold as follows. The following section provides a framework for understanding the problematique of the study. The next three sections provide a brief background on the objectives of the study; a number of posers in the form of research questions; and the significance of the study. Section 6 discusses the methodology used for the study while section 7 sheds light on the results/discussion. Sections 8, 9 and 10 focus on the discussion of findings, recapitulation and recommendations respectively.

2. Statement of the Problem

New thinking and innovation are being incorporated into teaching, learning and assessment methods globally. Africa cannot be left behind. The earlier we woke up to the realization of the fact that there is a paradigm shift the better. The way teaching and assessment are being carried out in modern times differs from those experienced decades ago.

Consequently, 21st century skills have to be inculcated in students through teaching and assessment if they are to be adequately integrated into future jobs and careers and become globally competitive when seeking admission for further studies abroad. Invariably, the study investigated, explicated and explored the extent to which students in Nigeria are prepared for the assessment of 21st Century skills by WAEC.

3. Objectives of the Study

The study was motivated by the quest to:

- (1) determine the extent to which Senior Secondary (SS) students are conversant with the dynamics of 21st century skills.

meeting the assessment criteria used to gauge the intended outcomes' of a syllabus or scheme of work. Unlike formative assessment, summative assessment contributes to the final mark a student scores at the end of term. The two pillars upon which summative assessment rests are reliability and validity.

- (2) investigate the extent to which SS students engage in social networking and online resources as a form of interaction.
- (3) know whether a proper grasp of the essentials of 21st century skills by SS students can serve as contributory factors for the assessment of these skills by WAEC.

4. Research Questions

- (1) To what extent are SS students familiar with the dynamics of 21st Century Skills?
- (2) To what extent are SS students engaged in social networking and online resources as a form of interaction?
- (3) How would the proper grasp of the fundamentals of 21st century skills by SS students serve as causative factor for the assessment of these skills by WAEC?

5. Significance of the Study

- (1) The invaluable significance of this study is reflected in the unique way it helps us to better understand how a proper grasp of the 21st Century skills by students can provide justification for the West African Examinations Council to commence full scale assessment of the skills at WASSCE in the local context of Nigeria as well as on a regional scale.
- (2) Furthermore, this study offers critical potentials for further research, as much as application of informed ideas and lessons from the Nigerian context of the assessment of 21st Century skills to the broader

dimension of global studies on the subject matter.

6. Methodology

The study adopted a blend of data from both primary and secondary sources.

(1) Research Design

The study employed a survey design. This design became crucial for this study due to the fact that the study involved the collection of quantifiable data in connection with variables which are examined to detect pattern of association (Bryman, 2004).

(2) Study Population, Sampling Procedure and Sample Size

The population for this study comprised 258 senior secondary school students from the six geopolitical zones of the Federal Republic of Nigeria. The participants were sampled using purposive/judgemental sampling procedure as follows:

Urban: Schools – Nos. 6; Students – Nos. 88

Semi-urban: Schools – Nos. 9; Students – Nos. 82

Rural: Schools – Nos. 6; Students – Nos. 88

(3) Study Location and Site

The primary data were gathered during field research in Adamawa State (North East), Anambra State (South East), Benue State (North Central), Kano State (North West), Lagos State (South West) and Rivers State (South South). The field sites were secondary schools chosen based on certain stratification criteria viz., urban, semi-urban and rural.

(4) Instrument and Instrumentation

21st Century Skills Questionnaire was developed and used to elicit relevant responses from the respondents. The questionnaire consisted of two parts: Part A and Part B. Part A consisted of six items on the respondent's socio-demographic profile. Part B consisted of 25 items that sought respondent's perception on the various aspects of the 21st Century skills.

All the items had a four-response format, some of which were "Always", "Sometimes", "Never" and "Not sure".

Other responses included "I own already", "Don't own but would like one", "Don't own but I know how it works" and "Don't know about it" while "I have an active account", "Don't have but would like one", "Don't own but I know how it works" and "Don't know about it" were responses to some items. Responses like "Strongly agree", "Agree somewhat", "Disagree" and "Have not used such site to form an opinion" as their responses are also included. The questionnaire was validated through peer critique and review by a senior researcher.

(5) Data Collection

About 300 questionnaires were administered to students at their various school locations by research assistants in the six geo-political zones. In administering the questionnaires, respondents were required to complete and submit the completed questionnaires on the spot. This ensured high return rate.

To further enrich the study and increase its generalization, primary data obtained through the above method were complemented with secondary data gathered through archive and database in determining the historical, social and scientific backdrops for the study. In other words, ideas were drawn from the existing body of literature on relevant themes related to this study.

(6) Analysis of Data

Portions of the data obtained were analyzed descriptively using analytical tools such as the Statistical Package for the Social Sciences (SPSS) and Microsoft Excel.

7. RESULTS AND DISCUSSION

The result of the study is presented and discussed in order of the research questions as follows:

| Gender | Ethnic Group | Social Status | Level of Education |
|-------------------|------------------|-------------------------|--------------------|
| Male 140 (54.3) | Hausa 34 (13.2) | Upper Class 52 (20.2) | SS 1 11 (4.3) |
| Female 117 (45.3) | Igbo 70 (27.1) | Middle Class 178 (69.0) | SS 2 135 (52.3) |
| | Yoruba 71 (27.5) | Lower Class 19 (7.4) | SS 3 112 (43.4) |
| | Others 83 (32.2) | Omission 9 (3.5) | |

Table 1: Socio-demographic Characteristics of Respondents

As shown in Table 1 above, there are two hundred and fifty eight (258) respondents of which there are one hundred and forty (140) male respondents accounting for 54.3% of total respondents, one hundred and seventeen (117) female respondents accounting for 45.3% of the total respondents and one (1) missing system accounting for 0.4% of the total respondents. More so, thirty-four (34) respondents were Hausas accounting for 13.2% of the total respondents, seventy (70) respondents were Igbos accounting for 27.1% of the total respondents, seventy-one (71) respondents were Yorubas accounting for 27.5% of the total respondents and eighty-three (83) respondents are from other ethnic groups accounting for 32.2% of the total respondents. Nor is that all, the parents of fifty-two (52) respondents are of upper class social status accounting for 20.2% of the total respondents, the parents of one hundred and seventy eight (178) respondents are of middle class social status accounting for 69.0% of the total respondents, the parents of nineteen (19) respondents are of lower class social status accounting for 7.4% of the total respondents and nine (9) missing values accounting for 3.5% of the total respondents.

7.1 Research Question One

To what extent are SS students familiar with the dynamics of 21st Century Skills?

Table 2 shows the students’ responses to their level of practice of 6 items on the questionnaire in order to show

the extent to which they are familiar with the dynamics of the 21st century skills. The items were assessed on a 4-point Likert scale ranging from ‘not sure’ to ‘always’.

| | | | | | |
|----|---|-----------|-----------|---------|---------|
| 1. | I create something new and in some way valuable such as an idea, a solution, and literary work and so on (creativity) | 58(22.5) | 176(68.2) | 16(6.2) | 8(3.1) |
| 2. | I think clearly and rationally and understand the logical connections between ideas (critical thinking) | 104(40.3) | 137(53.1) | 6(2.3) | 10(3.9) |
| 3. | I consider a variety of methods and perspectives in solving problem. (problem-solving) | 138(53.5) | 103(39.9) | 7(2.7) | 9(3.5) |
| 4. | I summarize my conclusions and identify the most desirable course of actions as well as other acceptable choices (decision-making) | 146(56.6) | 99(38.4) | 6(2.3) | 7(2.7) |
| 5. | I listen to the opinions of other students with respect (communication) | 178(69.0) | 72(27.9) | 5(1.9) | 3(1.2) |
| 6. | I am able to overcome the conflicts that arise between group members (team work) | 113(43.8) | 123(47.7) | 11(4.3) | 9(3.5) |

Table 2: Level of Familiarity of SS Students with the Dynamics of 21st Century Skills

From Table 2 above, 22.5%, 40.3%, 53.5%, 56.6%, 69.0% and 43.8% of the total respondents are always creative, think critically, solve problems, make decisions, communicate and engage in team work respectively. It can be inferred that a larger percentage of respondents always make use of the 21st century thinking and working skills while fewer respondents have never made use of the 21st century thinking and working skills. In addition, it is obvious from the table that larger percentage of

respondents are involved in communication while only few are creative.

7.2. Research Question Two

To what extent are SS students engaged in social networking and online resources as a form of interaction? Figure 2 shows the students' responses to their level of practice of the ---items on the questionnaire in order to show the extent to which they engage in social networking and online resources as a form of interaction. The items were assessed on a 4-point Likert scale ranging from 'Don't know about it' to 'I have an active account'.

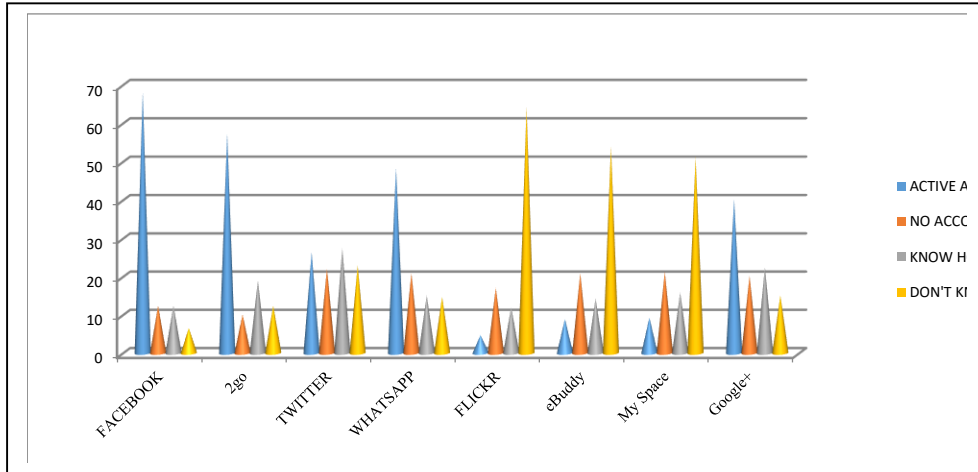


Figure 2: Extent to which SS Students engaged in Social Networking and Online Resources as a Form of Interaction

Figure 2 shows that 68.2%, 57.4%, 26.4%, 48.4%, 4.7%, 8.9%, 9.3% and 40.3% of total respondents have an active account on Facebook, 2go, Twitter, WhatsApp, Flickr, EBuddy, MySpace and Google+ respectively. This indicates that larger percentages of the respondents have an active Facebook account while fewer respondents have a Flickr account. This result also indicates that a larger percentage of the respondents make good use of online resources as a form of interaction.

7.3. Research Question Three

How would the proper grasp of the fundamentals of 21st century skills by SS students serve as causative factor for the assessment of these skills by WAEC?

Table 3 shows the students' responses to their level of practice of 5 items on the questionnaire in order to show how their proper grasp of the basics of 21st century skills can prompt WAEC to start assessing the skills in WASSCE. Again, the items were assessed on a 4-point Likert scale ranging from 'have not formed an opinion' to 'strongly agree'.

| s/n | Items | Strongly agree | Agree somewhat | Disagree | Have not formed and opinion |
|-----|---|----------------|----------------|----------|-----------------------------|
| 1. | Do you think that integration of online social sites into your learning process would enrich your experience - seeing announcements or lesson notes from your school on Facebook, for instance? | 169(65.5) | 52(20.2) | 21(8.1) | 15(5.8) |
| 2. | The way teachers are educating me helps me to see how ideas I learn in class apply to real life | 197(76.4) | 40(15.5) | 10(3.9) | 7(2.7) |
| 3. | Building and maintaining contacts and relationships with other students for problem-solving will make it easier for me to work with them (Networking) | 166(64.3) | 63(24.4) | 15(5.8) | 11(4.3) |
| 4. | The education that I am receiving will make me to be socially responsible | 227(88.0) | 23(8.9) | 6(2.3) | 1(0.4) |
| 5. | Incorporation of Civic Education into the syllabus to address societal problems will be useful in my life and career and make me a better citizen of Nigeria and the world | 196(76.0) | 33(12.8) | 18(7.0) | 6(2.3) |

Table 3: Grasping the Fundamentals of 21st Century Skills as Basis for Large Scale Assessment

From Table 3 above, it can be seen that 65.5% of the total respondents strongly agree that the incorporation of online social site into their education will enrich their learning experience, 8.1% of the total respondents disagree with this assertion while 5.8% of the respondents have not formed an opinion about this. More so, 76.4% of respondents strongly agree that the way teachers educate them help them to see how the ideas they are taught in school apply to real life, 88.0% of total respondents strongly agree that the education they are receiving will make them socially responsible while 76.0% of total respondents strongly agree that the incorporation of Civic Education into the syllabus to address societal problems will be useful in their life and

career and make them better citizens of Nigeria and the world at large.

Similarly, more than average (64.3%) of the total respondents believe in building and maintaining contacts with their colleagues as this would enhance their problem solving skills. However, about 4.3% of the total respondents do not have an opinion about this.

8. Discussion of Findings

In addition to the analysis presented above based on the three research questions, it was also discovered in the course of this study that, generally there is a high level of awareness with respect to 21st century skills regardless of social classes of respondents. And, Most of the participants in the survey believe that integration of

online social sites into their learning process would enrich their learning experience, however, on the downside, it might prove difficult to control the misuse of these social sites by some students.

More so, there is a general perception that 21st century skills are mere adds on; this perception needs to be changed and the skills properly integrated into the teaching, learning and assessment regimes. Intensive professional development with respect to 21st century skills is not yet a commonplace for teachers, school owners and test experts.

9. Conclusion and Recommendations

Data gathered from field work have been efficiently analyzed based on the research questions and the results presented. Obviously, the Nigerian education system, collectively has more experience in measuring academic content than in measuring interpersonal or intrapersonal dispositions. More so, four elements that are crucial to the effective incorporation of the 21st century skills into Nigerian education system were identified viz., applicable educational policies, school curricula, quality of teacher and mode of assessment. Although the 21st century skills have not been properly integrated into the Nigerian education system, it can be seen that most school age students are quite adept with some of these skills. It is therefore suggested that the government of Nigeria, WAEC, school administrators, teachers and other stakeholders in the education sector should key into these innovative ideas by implementing the following recommendations:

- (1) Examination bodies should endeavour to hire Subject Officers/Test Experts that are conversant with the fundamentals of the 21st Century education. More so, WAEC Subject Officers should be conversant with the current trends in their subject areas and learn how the nitty-gritty of the Century skills can be effectively incorporated into the assessment regimes.
- (2) Item Writers/examiners have a great role to play in the development of the final test papers. Such roles include but not limited to the generation of items and moderation of edited items. Consequently, a seminar that will make them to have in-depth knowledge of the substance of the 21st century skills should be organized as a precursor to the Item Writing Workshop.
- (3) The goal of integrating 21st Century skills into WASSCE will not be achieved if we have more recall – ‘knowledge-based’ or comprehension question in our final test papers. More attention should be focused on items at the level of application, analysis and/or evaluation as regards the Bloom’s taxonomy. As such the test specification tables and test blueprints should be adjusted to reflect the current realities.
- (4) Additionally, Scenario Based Test can be experimented as a pilot study before a large-scale assessment is carried out.
- (5) It is imperative to incorporate the proper use of these skills into the school curricula in Nigeria so as to allow Nigerian students get the most benefit from them.
- (6) Today, every student requires 21st century skills to succeed because these skills are very relevant to students’ living in the 21st century for them to achieve greater success in their academic endeavours.
- (7) The 21st century curriculum, instruction, and schemes of work which should focus on providing opportunities for applying 21st century skills across content areas should be developed by relevant authorities.
- (8) More so, the 21st century syllabus that: excites inventive learning methods and integrates the use of supportive know-how and critical/creative thinking skills should be developed by WAEC.

- (9) Teachers, school owners and other stakeholders in the education sector should be professionally trained on the specifics of the 21st century skills.
- (10) Awareness about 21st century skills assessment should be created, orientation and re-orientation of schools and teachers should be embarked upon by WAEC.
- (11) WAEC should strive to benchmark with some other examining bodies in the First World Countries that have been able to successfully integrate these skills into their assessment procedures.
- (12) When assessing the 21st century skills, the assessment should be such that is, borrowing the words of Harleen Singh, 2013, instruments should be designed in such a way that they are 'responsive', 'flexible', 'informative', 'technically sound' and have a 'blend of formative and summative assessment'.
- (13) School learning environments that will support the teaching, learning and assessment of 21st century skill outcomes are very germane and therefore should be created.
- (14) Osun State blazed the trail of digital literacy in Nigeria. The state governor introduced what is known in the local parlance as 'Opon Imo' – Tablet of Knowledge/School Computer Tablet – to students at secondary schools level. Earlier this year, Lagos State also followed suit by introducing a digital literacy project tagged 'Educate Lagos'. It is meant to provide study aids on core subjects from primary to secondary school curriculum and tutorial materials, instructional videos, selected e-books for primary to senior secondary, 3 approved texts, online forum and exam-mate just to mention but a few. These innovative projects are worthy of emulation by other states in Nigeria and countries of the West African sub region.

References:

- Adlon, J.A. (2015). *21st Century Skills Position Paper: Technology and Learning in the Modern Age*.
https://www.academia.edu...in_the_Moder_n_Age. Accessed 20/07/15
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M. and M. Rumble. (2010). "Defining 21st Century Skills" *Assessment and Teaching of 21st Century Skills Draft White Paper*. The University of Melbourne.
- Bryman, A. (2004), *Social Research Methods*, Second Edition. Oxford: Oxford University Press
- Ellis, R. and Collins, N. (2011). "Scenario-based Electronic Learning: A Viable Educational Method" *Journal of Paramedic Practice*, Vol. 4, No. 2, pp. 96-104.
- Erwin, T.D. (1991). *Assessing Student Learning and Development: A Guide to the Principles, Goals, and Methods of Determining College Outcomes*. San Francisco: Jossey-Bass.
- Great Schools Partnership. (2014), *The Glossary of Education Reform: For journalists, parents, and community members Portland*, <http://edglossary.org/21st-century-skills/>.
- Greenstein, L.M. (2012). *Assessing 21st Century Skills: A Guide to Evaluating Mastery and Authentic Learning*. London: Sage Publication.
- Holy Bible, King James Version
- Huba, M. E. and Freed, J. E. (2000). *Learner-Centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning*. Boston: Allyn and Bacon.
- Framework for 21st Century Learning*. Washington, D.C.: Partnership for 21st Century Learning. P21.,
- Rajendran, N. (2015). "21st Century Skills" nikhilaedulife.blogspot.com...html. Accessed 19/07/15

- Singh, H. (2013). *Assessing 21st Century Skills*.
<https://www.slideshare.net/HarleenSing/assessing-21st-century-skills-28570572>
Retrieved 30/06/17
- Terenzini, P.T. (1989). "Assessment with Open Eyes: Pitfalls in Studying Student Outcomes." *Journal of Higher Education*, Vol. 60, No. 6, pp. 644-664.
- The Glossary of Education Reforms
<http://edglossary.org/21st-century-skills/>
Accessed 15/01/15.
- Trilling, B. and Fadel, C. (2009). *21st Century Skills: Learning for Life in our Times*. San Francisco: Jossey-Bass.

Assessment of Soft-Skills in Ugandan Secondary Schools: The Development and Use of Assessment Tools

John Mary Vianney Mitana^{1*}, Anthony Mugagga Muwagga (PhD)², Mauro Giacomazzi³, Omala Saint Kizito (PhD)⁴,
Martin Ariapa⁵, Florence Nsubuga (PhD)⁶

1= PhD Candidate Makerere University and Principal, Luigi Giussani Institute of Higher Education (LGIHE), 2= Associate Professor, Makerere University, 3= Institutional Development Advisor, LGIHE, 4= Senior Examination Officer at UNEB, 5= Monitoring and Evaluation Officer, LGIHE, 6=Senior Lecturer Makerere University.

* Corresponding Author: j.mitana@lgihe.org Tel: +256(0) 772 991945/701504068

Institution: Luigi Giussani Institute of Higher Education, 31th May 2017

Abstract

Literature and empirical evidence suggest that the quality of education cannot be limited to the level of cognitive or academic performance of students alone. Emotional intelligence, also known as soft-skills, is increasingly recognised as critical for students' survival, advancement, and success in academic, professional and personal life. In Uganda, adopting a national assessment framework that measures and assesses soft skills at lower secondary school levels, could positively transform public examinations and learning outcomes. The documentary and empirical evidence in this paper demonstrates how the public examination system could sufficiently assess the soft skills relevant for the current labour market. The paper presents results of a Soft Skills Assessment Tools (SSATs) pre-tests conducted among students and teachers at lower secondary school levels in Uganda. The SSATs pre-tests identified 71 attributes/qualities/items for measuring/assessing the study sample's application of 11 different soft skills. The pre-tests indicate that the SSATs are reliable and valid instruments for testing the levels of soft skills application/proficiency. The outcomes of the SSATs pre-tests bear national policy implications regarding curriculum development, pedagogies, learning outcomes and the integration of soft skills assessment in public examinations generally. Hence, the paper recommends: the revision and pilot implementation of the SSATs nationally; establishment of similar continuous assessment tools for the teachers' career and personal development; and the integration of SSATs in all public examinations.

Key words: Assessment, Soft Skills, Assessment Tools.

1.0. Introduction to the Soft Skills Assessment Tools (SSATs) Pre-Test Study

Luigi Giussani Institute of Higher Education (LGIHE) undertook a broad three-phase study about the integration of soft skills assessment in public examinations at lower secondary school in Uganda. It was implemented as follows: **Phase I** - a Market Survey and needs assessment of the types of soft skills on demand; **Phase II** - Development and pre-testing of the Soft Skills Assessment Tools (SSATs); and **Phase III** - investigating possible mechanism for mainstreaming the SSATs nationally (*study underway*). Phase II of the study, fulfilled the purpose of formulating a testable educational

assessment framework for measuring soft skills aptitude. The SSATs were pre-tested at lower secondary education levels among students and teachers in five districts (i.e. Kampala, Kyenjojo, Pader, Pallisa and Wakiso). Altogether, two pre-tests were conducted (see *Pre-tests I and II* below).

This paper presents the findings from Phase II only; highlighting their implications for the implementation stage. The main objective of the pre-tests was to identify practical approaches of measuring and assessing soft skills in lower secondary schools. The study relied on both quantitative and qualitative methodologies to establish the efficacy of the SSATs. Through psychometric evaluation techniques, the SSATs generated valid and reliable survey and qualitative-type data reflective of, for example, the personality traits, mental abilities/aptitude, opinions or intelligence of the

pre-test sample (i.e. lower secondary school teachers and students).³ Therefore, this paper reports about the validity and reliability of the SSATs as instruments for measuring specific soft skills and recommends their improvement in view of potential integration into the national educational system.

1.1 The Study Context: Thematic Concerns on Assessment of Soft Skills

Studies suggest that Uganda's education system falls short of achieving its intended objectives and aims as per the Government White Paper on Education (1992).⁴ The Ministry of Education and Sports (MoES) Strategic Report 2004-2015 argues that learners are failing to acquire the skills and knowledge relevant for the world of

work or further education. The report stresses that only a minority of students in post-primary education appear to be achieving at expected levels. Hence, many leave school ill-prepared to participate in the labour market as productive citizens or workers (MoES, 2004). Despite the attempts by different institutions such as the Uganda National Examinations Board (UNEB) to assess students' learning outcomes at secondary school levels, the results hardly paint an accurate picture of students' achievements. Relatedly, today's assessment of Uganda's secondary education achievements concentrates more on students' mastery of content and core academic skills, but neglects soft skills. The system mainly emphasises cognitive skills, standardised examination and test scores. Thus, confirming that the soft skills of: emotional intelligence, conflict resolution, communication, resilience, assertiveness, optimism, integrity, self-

³Psychometric evaluation refers to a test designed to reveal an individual's personality, mental ability or opinions etc. It is a branch of psychology concerned with the design, administration, and interpretation of quantitative tests aimed at measuring psychological variables such as intelligence, aptitude, and personality traits etc.

⁴See, the National Assessment of Progress in Education (NAPE) in Uganda report of 2014 and the National Assessment of Educational Progress (NAEP), UNESCO (2015) and Uwezo (2015) reports for perspectives on educational achievements.

awareness and patience demanded by labour market are undeveloped and unappraised (LGIHE, 2016).⁵ Similarly, the soft skills of: motivation, time management, self-regulation, cooperation, conscientiousness, organisation, perseverance, goal setting, help-seeking, self-efficacy, self-regulation, self-control, self-discipline, motivation, convictions, effort, work habits, homework completion, study skills and learning strategies are either underemphasised or completely disregarded (Farrington *et al.*, 2012). This situation created a dichotomy between the general aims of education and what is assessed in schools; or between what is assessed and what is required of students to cope with life beyond secondary school. Despite the growing evidence that soft skills affect academic performance and life after school, policy-makers and educators seem not to

have leveraged that fact (Farrington *et al.*, 2012). Hence, the implications of soft skills on educational practice remain unclear. Nevertheless, the unsupported assumption has been that policy-makers and educators understand the value of soft skills development, have concrete strategies to develop them and reliable tools to assess/measure their effects on students in close reference to content knowledge and academic skills (Farrington *et al.*, 2012).

These concerns provide scope for evaluating what is taught and how it is taught to enhance the acquisition of soft skills as one of the major learning outcomes in Uganda's secondary level education. Although there is an undeniable recognition and demand for a universalised access to education today, it is also widely understood that a nation's ability to educate its youth cannot be measured by access to schooling or enrolment rates alone, but rather by its ability to enable students develop the requisite knowledge, values and skills to function as literate and numerate members of the wider society.

Moreover, the quality of education is no longer solely measured by students' ability to apply cognitive skills as employers and the realities of work require a mastery of soft skills (Awada, 2014). This attests that soft skills are considered "more important by employers than one's cognitive strengths or academic abilities" since they enable people to live a holistic life (Awada, 2014:1). Thus, this paper adopts the idea of holistic assessment of learning outcomes strategy as a means of highlighting that the existing assessment approaches at lower secondary educational levels mostly concentrate on testing academic competences and not soft skills.

2.0 Literature: Conceptualization of Soft Skills Assessment

Holistic quality education cannot be discussed in isolation of holistic quality educational assessment. However, educational assessment particularly high-stakes public/national examinations greatly influence

what takes place in the classroom and how it happens (Altinyelken, 2015; Kellaghan & Greaney, 2004). Available studies and practical experience suggest that the current assessment and examinations fall short of evaluating the quality of education yet education cannot be limited to the knowledge or application of literacy; numeracy and or science only (UNESCO, 2000). Therefore, without the adequate emotional intelligence (soft skills), the Intelligence Quotient (IQ) would not suffice for a person to live and work intelligently. Life thus, requires a number of abilities and skills that this paper describes as the soft skills or what is commonly known as life skills, social skills or emotional intelligence.

Soft skills are a broad set of skills, competencies, behaviours, attitudes, and personal qualities that enable people to effectively navigate life, understand their

⁵In the labour market survey, LGIHE found out that those are some of the soft skills Ugandan employers demand.

environment, work well with others, perform well, and achieve their goals (Lippman *et al.*, 2015). Related studies variously defined the personal attributes that are not usually measured by IQ tests or achievement tests as: soft skills, personality traits, non-cognitive abilities, character skills, life skills and socio-emotional skills. These different descriptions have different connotations despite often being applied interchangeably (Almlund *et al.*, 2011 and Borghans *et al.*, 2008). “Traits” suggests a sense of permanence and also possible heritability. This may also suggest that traits are more natural or inherited than skills. But, the term “skills” suggests that these attributes can be learned. Skills denote a particular ability to do something. Farrington *et al.* (2012) for example, categorises skills into cognitive and non-cognitive, arguing that some such as problem solving and critical thinking require deliberate and high cognition while others such as perseverance require “weak” cognition.

Therefore, the idea of rating skills according to the levels of cognition does not accurately differentiate them since in essence, all skills are cognitive. Farrington *et al.* (2012) asserts that in addition to curriculum knowledge and

academic skills, students must develop sets of behaviour, skills, attitudes, and strategies that are crucial for academic performance, even though these may not be reflected in students’ test scores. Hence, Farrington *et al.* (2012) adopted the term *factors* to go beyond a narrow reference to skills and embrace strategies, attitudes, and behaviours. ‘Factors’ allude to an influence contributing to a result. This potentially and erroneously reduces skills to simply mean a contributory rather than a consequential outcome; as if they cannot be learning outcomes on their own (*pre-test results below*). Education and health related studies commonly refer to soft skills as ‘life skills.’ According to the World Health Organisation (WHO,1994) ‘life skills’ indicate the abilities for adaptive and positive behaviour, that enables individuals to deal effectively with the demands and challenges of everyday life. These are cognitive skills

for analysing and using information, personal skills for developing self-awareness and managing oneself, and inter-personal skills for communicating and interacting effectively with others.

However, the term ‘life skills’ is often applied interchangeably with soft skills or non-cognitive skills (World Bank, 2013). This paper adopts the term ‘soft skills’ and avoids life skills because the former emphasises the malleability of the latter. The understanding of soft skills as the 21st century skills informed the design of the SSATs and the purpose of the pre-tests discussed here. Awareness of Uganda labour market’s soft skills needs and definition provided scope to align the SSATs pre-tests with the soft skills that would be considered as the most important recruitment requirement for entry-level jobs (LGIHE, 2016). Although functional literacy and numeracy are key elements of the measure of the quality of education, they scarcely define quality education as the latter includes all “outcomes of education such as the knowledge, skills,

competencies, attitudes, and values that learners acquire” (Wagner, 2011:23). In terms of soft skills proficiency as a measure of quality education, this paper leans on Wagner’s (2011).

2.0. Methodology

The study built on the preliminary qualitative data collected from a multi-sectoral group of respondents who identified critical current labour market needs-based thematic soft skills areas (LGIHE, 2016). The respondents, proposed the assessment of 11 specific soft skills domains/constructs, namely: problem solving, critical thinking, responsibility, achievement striving, grit, integrity/honesty, assertiveness, teamwork, compassion/empathy, self-control, and self-esteem because they are considered vital for work. The pre-tests relied on quantitative approaches to establish the validity and reliability of the SSATs in measuring the 11 soft skills constructs. Pre-test I identified 79 items associated with the 11 soft skills constructs and tried them on 380

secondary students of Grade 2 on a rating scale of or (a 1-5 Likert score space).⁶ The students responded by means of self-reporting. A sample of 10 students in each of the pre-test schools was rated by two teachers for the soft skills constructs qualities/attributes that the students responded to. However, the psychometric evaluation identified eight items (referring to the qualities or attributes of specific soft skills) as unsatisfactorily explained by the constructs and dropped from the first pre-test tool. Consequently, the low levels of teachers' agreement about their own students and the sharp inconsistencies between students' and teachers' ratings were considered indicative of how little the teachers knew their students. It was then preferable for students to self-report instead of being rated by their teachers. Recommendations from Pre-test I, influenced the design of Pre-test II, which contained 71 items of eleven 11 soft skill constructs. These 11 constructs were tested on 530 students of Grades 3 based on a rating scale of 1-3 or (a 1-3 Likert score space) as well as Literacy and Numeracy tests. The findings are shared below.

3.0. Summary of Findings: Psychometric Soft Skills Assessment Results

The findings presented are drawn from the results of pre-tests I and II. Table 1 below outlines results of the students' self-reporting from pre-test II.

⁶1-5 Likert score space refers/describes the rating scale 1-5 that was used in the first pre-test and later lowered to

1-3 for students while 1-5 was adopted for teacher ratings in pre-test 2.

Part I: Table. 1. Matrix Showing Results of Soft Skills Assessment Tools Pre-Test II⁷

| Tested Soft Skills Constructs | Items =Qualities/attributes/ abilities test scored | Conclusion |
|---|---|---|
| Test administered to students on the scale of 1 to 3 | | |
| 1.Problem-solving | Assessed using these 6 items: 1) First find out exactly what the problem is. 2) Thinking head to prevent problems. 3) Establish cause of problem. 4) Act impulsively. 5) I choose the easiest solution. 6) Take action related to solution. | All 6 items were found suitable, reliable and valid for assessing this skill construct. |
| 2.Critical Thinking | Assessed using these 5 items: 1) Easily recognize when in the wrong or right. 2) Identify best response to a problem. 3) Justify my opinions. 4) Put myself first mostly. 5) Pay attention to my inner feelings. | All 5 items were found suitable, reliable and valid for assessing this skill construct. |
| 3.Responsibility | Assessed using these 6 items: 5) Late reporting to school. 2) Finish all assignments before leaving class. 3) Hand my book for marking whenever required. 4) Inform teacher of my absence in advance. 5) Successfully complete the day's work. 6) Get to class before the first morning lesson. | All 6 items were found suitable, reliable and valid for assessing this skill construct. |
| 4.Achievement Striving | Assessed using these 5 items: 1) Learn things that seem hard. 2) Persist to the very end. 3) Repeat till right. 4) Persist with difficult task till successfully done. 5) I do my best in all my activities. | All 5 items were found suitable, reliable and valid measure of construct. |
| 5.Grit (Consistency and Perseverance) | Assessed using these 8 items: 5) Focus for a short time then lose interest. 2) Often set goal but later choose to follow a different one. 3) Sometimes distracted from previous ideas by new ones. 4) Finish whatever started. 5) Not discouraged by challenges. 6) Hard worker.7) Achieved goals after months of work. 8) Overcame challenges to achieve important goal. | All 8 items were found suitable, reliable and valid measure of construct. But 1-3, required further review. |

⁷**Students rating scale 1-3**, where 1=Never; 2=Sometimes; 3=Always: 1=Not at all true; 2=Somehow true; 3=completely true: 1=Disagree; 2=Partially Agree; 3=Agree. **Teachers rating scale 1-3**, where 1=Strongly Disagree; 2=Disagree; Disagree; 3=neither Neither Agree nor Disagree; 4=Agree; 5=Strongly Agree (this is because their initial tool in pre-test 1 had failed).

| Tested Soft Skills Constructs | Items =Qualities/attributes/abilities test scored | Conclusion |
|---|---|--|
| Test administered to students on the scale of 1 to 3 | | |
| 6.Integrity/Honesty | <p><i>Assessed using these 5 items:</i></p> <p>1) Help friends answer questions during an exam. 2) Ask classmates for help to answer questions during an exam. 3) Ask for permission to use friend’s materials (e.g. books, pens). 4) Sneak out of class when teacher not watching. 5) Tell lies to save self from punishment.</p> | <p>All 5 items were found suitable, reliable and valid measure of construct. Yet, 1, 2, 4 & 5 skewed results towards socially acceptable implications & need reviewing.</p> |
| 7. Assertiveness | <p><i>Assessed using these 10 items:</i></p> <p>1) When asked to do something, insist upon knowing why. 2) Enjoy starting conversations with new friends. 3) Show feelings through facial expressions. 4) Try to maintain eye contact when talking to others. 5) Look at friends’ face to understand their feelings. 6) Express feelings by words and actions together. 7) Take decisions that are fair to others. 8) Do not blame others even when hurt. 9) Try to calm friends down when they argue. 10) Apologize to others when caused them hurt.</p> | <p>All 10 items were found suitable, reliable and valid measures of construct.</p> |
| 8.Cooperation/Team work/Sense of Belonging | <p><i>Assessed using these 8 items:</i></p> <p>1) Work well with peers. 2) Solve problems with peers without becoming rude. 3) Pay attention to how others are feeling. 4) Understand the feelings of peers. 5) Happy to be at this school. 6) Feel lucky to be student of this school. 7) Treated well by people in school. 8) Feel uncomfortable at school.</p> | <p>All 8 items were found suitable, reliable and valid measures of construct. Still, item 8 skewed results towards socially acceptable implications & need reviewing.</p> |
| 9.Compassion/Empathy | <p><i>Assessed using these 4 items:</i></p> <p>1) Listen attentively to others. 2) Unconcerned when others are sad. 3) Comfort people when they are sad. 4) Happy when others are happy.</p> | <p>All 4 items found almost unsuitable, unreliable and invalid measures of construct. All poorly separate respondents. But, item 2 skewed results towards socially acceptable implications & need reviewing.</p> |
| 10.Self-control/Patience | <p><i>Assessed using these 7 items:</i></p> <p>1) Easily get disappointed. 2) Friends are often annoying. 3) Waits in line patiently. 4) Lose attention easily. 5) Refuse things that are bad even if they are fun. 6) Good at resisting temptation. 7) Do things that feel good in the moment but regret later on.</p> | <p>All 7 items found suitable, reliable and valid measures of construct. But, 1, 2, 4 & 7 poorly separate respondents and are skewed results towards socially acceptable implications & need reviewing.</p> |
| 11.Self-esteem | <p><i>Assessed using these 5 items:</i></p> | <p>All 5 items found suitable, reliable and valid measures of</p> |

| Tested Soft Skills Constructs | Items =Qualities/attributes/ abilities test scored | Conclusion |
|--|--|--|
| Test administered to students on the scale of 1 to 3 | | |
| | 1) At times, lower self. 2) Feel that have a number of qualities that most people admire. 3) Can do things just like most of the other people. 4) Feel do not have much to be proud of. 5) Certainly feel useless at times. | construct. But, 1, 4 & 5 poorly separate respondents and are skewed results towards socially acceptable implications & need reviewing. |
| SECTION C: Results of Implications of Variable /Other Factors | | |
| Home and other school factors | <p><i>Assessed using these 11 items:</i></p> <p>5) Did you attend nursery school? 2) Did you repeat any classes? 3) If yes, which classes are those? 4) How many family members are you? 5) Who mainly stays and takes care of you at home? 6) Do you have both parents? 7) What is the level of education of your parents/guardian (<i>Father, Mother, Guardian</i>) 8) What is the main source of income for your family/ home? 9) How often do you participate in the following co-curricular activities at school (<i>Games, Sports, Clubs</i>).10) Do you have a school library? 11) How often do you visit a library to do personal research?</p> | All 11 items found suitable, reliable and valid measures of construct. |

Table.2. Students’ Self-Rating – Pre-Test I

| Indicators | Items/Questions For Students |
|--|--|
| Quality of teacher-pupil relationship | <p><i>Assessed using these 11 items:</i></p> <p>1) My teacher punishes students without even knowing what really happened. 2) I feel safe and comfortable with my teacher. 3) My teacher always treats all students the same. 4) My teacher doesn’t care about what I think. 5) My teacher always keeps his or her promises. 6) My teacher really cares about me.7) Teachers and students treat one another with respect in this school. 8) I feel like I can talk to my teacher about things that are bothering me. 9) Teachers and students in this school don’t seem to like one another.</p> |
| Inclusion of new and innovative teaching methods | <p><i>Assessed using these 5 items:</i></p> <p>1) What I study in this class really makes me think. 2) I am learning a lot in this class. 3) My teacher expects everyone to work hard in this class. 4) My teacher expects me to do my best all the time in this class. 5) My teacher wants us to learn beyond just memorizing.</p> |

Table.3. Teachers Self-Rating Pre-Test I

| Indicators | Items/Questions For Teachers |
|--|---|
| Collaboration among teachers | <p><i>Assessed using these 3 items:</i></p> <ol style="list-style-type: none"> 1) I often feel that I don't have much to offer my fellow teachers in my school. 2) I feel comfortable sharing my challenges with my fellow teachers. 3) My contribution is valued by my fellow teachers. |
| Teacher's job satisfaction and enthusiasm to teach. | <p><i>Assessed using these 12 items:</i></p> <ol style="list-style-type: none"> 1) I am satisfied with my role as a teacher. 2) Each day at my school seems like it will never end. 3) I find real enjoyment in my work. 4) I consider my teaching job to be rather unpleasant. 5) I lead a purposeful and meaningful life. 6) My social relationships are supportive and rewarding. 7) I am engaged and interested in my daily activities. 8) I actively contribute to the happiness and well-being of others. 9) I am competent and capable in the activities that are important to me. 10) I am a good person and live a good life. 11) I am optimistic about the future. 12) People respect me. |
| Role of head teacher in continuing professional development. | <p><i>Assessed using these 4 items:</i></p> <ol style="list-style-type: none"> 1) My head teacher supports the use of alternative and innovative teaching methods. 2) On average teachers in my school prepare their daily lesson plans. 3) My head teacher provides feedback regarding my classroom performance. 4) My head teacher advises me about my teaching methods. |

4.0. Discussion of Findings from Pre-tests I and II

Although, both pre-tests assessed numeracy and literacy skills, only the outcomes of the SSATs have been reported. According to the matrix in Table 1 the SSATs were found to be suitable, valid and reliable instruments for measuring the 11 soft skills constructs. All the soft skills qualities described as test items (Table 1) indicate that each of the skills can be assessed independently based on their own merits and attributes (or psychometric rigor). The SSATs made the student's abilities or application of each skill easily identifiable and assessable. Six out of the 11 SSATs require minor

modifications because some of their test items tended to skew results towards socially acceptable expectations. This implies that the students were merely reporting according to pre-determined socially correct/acceptable attributes to appear to be doing the right thing (Table.1).

5.1 Comments on Teacher vs. Student Self-Rating

Teachers were assessed to determine how well their soft skills abilities would vary from or complement the students'. The results were different for both students and teachers; they highlighted how the student-teacher classroom/school relationships depend on the correct

application of the correct soft skills in the correct contexts (Table 2 and 3). During self-rating, students rated teachers on eight major indicators, although only two indicators are cited in Table.2 above. The items assessed and highlighted how soft skills improve the quality of teacher-students' relationships and the teaching-learning experience. Teachers self-reported on three major indicators namely: collaboration among teachers, teachers' job satisfaction and enthusiasm to teach and role of the head-teacher in continuous professional development (Table.3). All three indicators led to the assessment of teacher-teachers' interactions in a classroom or school environment.

The indicators assessed teachers' job satisfaction, teamwork abilities, and how the school system contributes to their career advancement or personal professional development. The third tool that was used in Pre-test I for teacher-student rating. Two teachers in each participating Pre-test I school in rated the same students over a sample of test items. However, when evaluated against students' self-ratings the teacher ratings of the students of the results specific teacher ratings, the level of teacher rating consistency was found to be low (4.8%). Pre-test I concluded that the results from the exploratory analysis of two teacher-students ratings are indicative of limited teacher-student interactions and reflected the inconsistency of teacher ratings. This led to that specific SSATs (teaching-student rating soft skills tool) being dropped from Pre-test II.

5.0. Conclusion

In conclusion, this study has demonstrates that with the right tools, it is possible to assess soft skills. However, soft skills aptitude is sometimes influenced by both home and school factors. The 11 constructs of the SSATs indicate that learners have the potential to apply those skills variously depending on their social contexts. This largely explains why some items skewed or pre-determined student responses (Table 1). Self-reporting as

seen in both pre-tests allows both students and teachers to self-evaluate, reflect and desire learning and change.

6.1 Recommendation and Policy Implications

An educational assessment framework that depicts students' soft skills aptitude such as the SSATs provides scope for policy review and realignment of curriculum, and pedagogies with soft skills assessment methods, public examinations and student learning outcomes generally. This provides scope for further research and policy consultations to fill existing knowledge gaps regarding the assessment of soft skills as learning outcomes in the secondary education system. Thus, the study recommends:

- i. Refining and adopting the SSATs for pilot implementation nationally;
- ii. Establishment of similar continuous assessment tools for teachers as part of career and personal development; and
- iii. Integration of the SSATs in all public examinations for accurate portrayal of students' learning outcomes.

References

- Almlund, M., A. Duckworth, J. J. Heckman, and T. Kautz (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Wobmann (Eds.), *Handbook of the Economics of Education*, Volume 4. Amsterdam: Elsevier. Forthcoming.
- Altinyelken, H. K (2015) *Education for All 2000-2015. Achievements and Challenges*. Background Paper Prepared for the Education for All Global Monitoring Report 2015
- Borghans, L., A. L. Duckworth, J. J. Heckman, and B. Weel (2008, Fall). The economics and psychology of personality traits. *Journal of Human Resources* 43 (4), 972{1059.

Farrington *et al* (2012). The University of Chicago consortium on Chicago school research

2014 Teaching Adolescents to Become Learners the Role of Non-cognitive Factors in Shaping School Performance: A Critical Literature Review

Heckman, (2008)

Kellaghan, T. & Greaney, V. (2004) *Assessing Students Learning in Africa*. World Bank.

Washington D.C

Lippman, L.H., Ryberg, R., Carney, R. and Moore, K.A.

(2015). *Key “Soft Skills” that Foster Youth Workforce Success: Toward a Consensus Across Fields*. Washington, DC: USAID, FHI 360, Child Trends. Published through the Workforce Connections project managed by FHI 360 and funded by USAID

Ministry of Education and Sports (MoES) Strategic Report 2004-2015

UNESCO (2000) *Education for All: Assessing Learning Achievement. Status and Trends*

2000. Paris, UNESCO

Wagner, D.A (2011). *Smaller, Quicker, Cheaper Improving Learning Assessments for*

Developing Countries. International Institute for Educational Planning. Paris

WHO (1994) *Life Skills Education in Schools*. Introduction and Guidelines to Facilitate the

Development and Implementation of Life Skills Programmes. WHO. Geneva

The explicit assessment of skills for the 21st century - a case study by Cambridge International Examinations

Jane Henry, Paul Bullen-Smith & Ashley small

Introduction

Increasingly it seems that politicians, business leaders and those involved in education believe that education systems need to address the fact that young people in the 21st century require a particular set of skills and competencies to be successful in modern life and that these competencies need to be explicitly included in school curricula, and form part of the national school qualifications, assessed by public examination boards.

These skills have tended to revolve around specific skills such as communication, collaboration, creative thinking, analysis and problem solving, which are often not the explicit focus of traditional academic teaching which has tended to focus on subject knowledge and its application. In this paper we will look at how we at Cambridge International Examinations have designed and launched a range of skills based programmes, Cambridge Global Perspectives® from primary through to advanced secondary, preparing students for this new world.

To do so we have developed several innovative and appropriate ways of assessing these skills explicitly, fairly, reliably and validly.

Theoretical Framework

There is no single definitive list of these skills and competencies, but three frameworks are outlined below, from which it can be seen that there is some consensus about what should be included.

A paper on the assessment of 21st century skills by Emily Lai and Michaela Viering (Lai & Viering, 2012) and Irenka Suto's paper (Suto.I, 2013) integrate some of the research that has been done in this area.

Both papers consider some of the frameworks developed by different organisations, namely, Partnership for 21st Century Learning [P21] (Partnership for 21st Century Skills, 2009); the Assessment and Teaching of 21st Century skills organisation [ATC21S] (Binkley, et al., 2010); and the National Research Council [NRC] on (Measurement in Education, National Research Council, Committee on Assessment of 21st Century skills, 2011). The P21 framework identified the different skills considered important for higher education and the world of work:

- Learning and innovation (to include creativity and innovation, critical thinking and problem solving and communication and collaboration)
- Information, media and technology skills (to include information literacy, media literacy and technology skills and information/communications/technology literacy)
- Life and Career skills (to include flexibility and adaptability, initiative and self-direction, social and cross-cultural skills, productivity and accountability, leadership and responsibility).

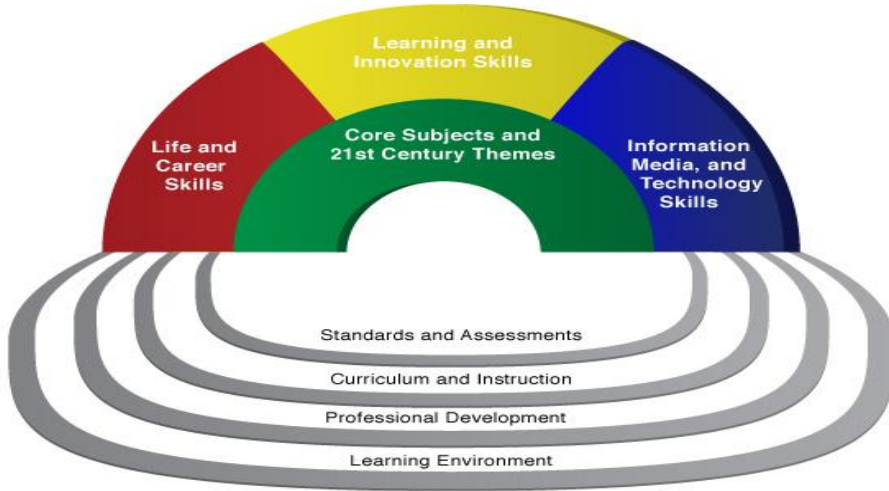


Figure 1: Partnership for 21st Century Learning

The ATC21 framework focuses on:

- Ways of Thinking (to include creativity and innovation; critical thinking, problem solving and decision-making; and metacognition)
- Ways of Working (to include communication and collaboration)
- Tools for Working (to include information literacy and information communication technology (ICT) literacy)
- Living in the World (to include citizenship, life and career skills and personal and social responsibility).

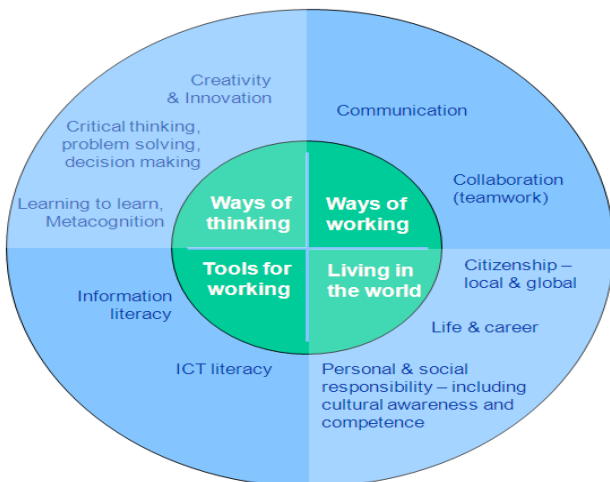


Figure 2: ATC21S

In comparing these frameworks and the research carried out for the European Commission report, it seems they have a number of areas in common in both the domains and in the skills within the domains. A report entitled, “Measuring 21st Century Competencies” for Asia Society Global Cities Education Network (Soland, Hamilton, & Stecher, 2013) also notes that it seems to be a general consensus that these skills, or competencies, can be divided into three broad domains: Cognitive competencies, Interpersonal competencies and Intrapersonal competencies. Within these domains, the skills most frequently highlighted across the different organisations reviewed are: critical thinking and problem solving; ICT literacy, operations and concepts; communication and collaboration.

Aims of our assessment

As an international examination board, Cambridge International Examinations has produced in Cambridge Global Perspectives a suite of interdisciplinary, skills-based syllabuses.

The suite of syllabuses comprises an IGCSE and O Level for ages 14 to16, as well as an International A Level for ages16 to19. We will also be launching shortly a Global Perspectives programme for ages 5 to14, giving a suite of assessed qualifications from primary through to advanced secondary progressing and developing students’ skills.

The syllabuses have a common theme: the development of a range of skills that are transferrable to other subjects, interdisciplinary in their nature, and global in their context. The background to the development of these syllabuses, as stated in the IGCSE syllabus 2018-2020, is that:

“Young people face unprecedented challenges in an interconnected and information-heavy world, not least in how they will gain a sense of their own active place in the world and cope with changes that will impact on their life chances and life choices. Cambridge IGCSE Global Perspectives candidates will have opportunities to acquire and apply a range of skills to support them in these challenges.”

The aims of our IGCSE are to enable learners to:

- become independent and empowered to take their place in an ever-changing, information-heavy, interconnected world
- develop an analytical, evaluative grasp of global issues and their causes, consequences and possible courses of action
- enquire into, and reflect on, issues independently and in collaboration with others from different cultural perspectives
- work independently as well as part of a team, directing much of their own learning with the teacher as an active facilitator
- consider important issues from personal, local and/or national and global perspectives and understand the links between these
- critically assess the information available to them and support judgements with lines of reasoning
- communicate and empathise with the needs and rights of others.

While the A Level syllabus 2017-2019 states that:

“It is widely recognised that we live in an increasingly digitised and inter-connected world. The means by which we access information and the pace with which this takes place are profoundly changing the way we learn, communicate and work. Increasingly, young people are

faced with access to a multiplicity of competing ideas. In such an information-rich society, young people need the skills and dispositions to be able to think critically.”

In addition, the aims of the A Level are to encourage learners to develop by:

- providing opportunities to acquire disciplined and scholarly research skills
- promoting a critical, questioning approach to information using the language of reasoning
- prompting self-reflection and independence of thought
- creating opportunities to understand and engage with key global issues wherever they live and work
- nurturing an awareness and understanding of, and respect for, the diversity of perspectives on global issues
- offering an interdisciplinary approach to global issues
- encouraging development of independent learning skills in preparation for study in higher education and lifelong learning
- promoting an understanding of appropriate research skills
- engaging in the research process on an academic topic of their own choice which reflects their interest
- providing opportunities for the exercise of the higher-order thinking skills of analysis, synthesis and evaluation
- providing opportunities to develop oral presentation and communication skills.

Objectives of our assessment

The range of syllabuses have been developed over the past few years, on each review being refined and developed in the light of research and best practice.

However, their core remains unchanged which is to develop a set of syllabuses that explicitly focusses on skills and how to develop and to assess them. All Cambridge syllabuses aim to develop skills to a greater or lesser extent, but most of them are grounded in a subject-specific context.

In IGCSE History, for example, History content is assessed explicitly while at the same time key skills are as well, as it is

“the skills of analysis and evaluation of sources of evidence, synthesis of information and creative thinking in solving ‘problems’ (causes and consequences of actions) that are required to be evident. In addition, there is an implied need for independent research and communication, and a reflection on learning across all sections of the syllabus.”

In A Level French, the content assessed is French, but it is

“...the skills of reflection on prior learning and the appropriate language to use, plus an analysis and evaluation of both language and an understanding and empathy with another culture that will be assessed. In addition, communication will obviously be at the heart of all four papers and innovative/creative thinking will be necessary on several occasions.”

These are both extracts from recent internal analysis undertaken on the skills implicit in the syllabuses that has allowed us to see more clearly how the skills developed in Global Perspectives, which are common to different subjects, can be so easily transferrable to a student’s other subjects. However, with Global Perspectives it is the skills that are explicit and the subject knowledge gained that is implicit. A teacher in one school summed this up perfectly by saying that *“in other courses I use content to build skills. In Cambridge Global Perspectives, I use skills to deepen a student’s understanding of content.”*

So what are these skills?

All Global Perspectives syllabuses focus on developing and assessing the skills of research, analysis, evaluation,

reflection, communication and collaboration. Most of these are implicit in our other Cambridge syllabuses but in Global Perspectives they have been made explicit and content-free. The explicit focus on reflection and collaboration, and the wider interpretation of the communication skills being assessed, are what make these syllabuses stand out from others. By content-free we do not mean that students do not learn content; rather they focus on the skills, and the content is a by-product, a context. This point will be considered again later in discussing how these skills are taught.

Why these skills?

The view of Cambridge is that these skills are vital to young learners, both for their immediate education and subject transferability, and later on for the world of work and higher education. Furthermore, these skills can be assessed. However, many other skills are also important and are taught, but Cambridge does not believe all of them can be assessed rigorously and consistently. For example, Tony Wagner from Harvard University talks about the Seven Survival Skills (Wagner, 2016), as defined by business leaders in their own words, which include such skills as ‘agility and adaptability’. Others talk about the importance of ‘grit’ and ‘resilience’ that would be exceedingly difficult to assess rigorously. In a 2015 conference speech, Sir Anthony Seldon, now Vice-Chancellor of Buckingham University, UK, argued that ‘*Students certainly need to be skilful at maths, science, languages and humanities. But they also need those skills that computers cannot replicate*’. (Seldon, 2015). Sir Anthony also quotes a recent study from Harvard University, the Carnegie Foundation and Stanford Research Center which found that employers need far more than the skills developed in exams – ‘*they also need what is patronisingly called the ‘soft’ skills ... of creativity, teamwork, empathy, grit, resilience and honesty*’. (National Soft Skills Association, 2015). This report concluded that ‘*85% of job success comes from having well-developed soft and people skills, and only 15% of job success comes from technical skills and knowledge (hard skills)*’ (ibid).

For Cambridge International Examinations as an examining body, the assessment context is paramount. Global Perspectives is not a study skills or general studies programme, but rather a set of syllabuses that need to be assessed as rigorously as any of our other subject-specific syllabuses, whether at IGCSE or at A level. The view of Cambridge is that other skills such as resilience or adaptability should be taught and developed, and elements of these will be developed by students taking our qualifications, but if they cannot realistically be assessed then that is not the focus for our syllabuses. Rather they are skills and traits whose development will be primarily encouraged by, for instance, parents, the school and extra-curricular activities.

From this theoretical and evidenced based background we can now move onto the specifics of how we assess these skills. For all Global Perspectives syllabuses we have developed the same set of assessment objectives from 5-19, as we focus on assessing those skills that we can assess reliably, validly and fairly.

AO1: Research, analysis and evaluation

AO2: Reflection

AO3: Communication and collaboration

Assessment approaches

As may be expected, the skills are assessed through a wide range of assessment methods: the tried and tested written question paper with source material under timed examination conditions; more innovatively, the candidate-chosen essay or individual report; and the Team Project where the *process* of collaboration is assessed as explicitly as the assessment of the *outcome*. This is the approach taken with both our IGCSE and the first year of the A Level

However, with the A Level, an extended 5000 word research report is added as the only assessed component in the students’ second year. A mix of different assessment methods is used in the different levels of the syllabuses, and with different weightings, but common assessment objectives are increasingly being used as syllabuses are reviewed and redeveloped.

Assessment objectives (AO) and their weightings for the latest IGCSE syllabus are as shown in Figure 4 below.

| Component | AO1 | AO2 | AO3 % |
|----------------------------------|-----|-----|-------|
| | % | % | |
| 1. Written Examination | 100 | - | - |
| 2. Individual Report | 67 | 8 | 25 |
| 3. Team Project | 28 | 36 | 36 |
| Weighting of AO in qualification | 65 | 15 | 20 |

Figure 4: Assessment Objectives for IGCSE Global Perspectives

The A Level syllabus has the same assessment objectives and a similar but different % weighting, as shown in Figure 5 below.

| Component | AO1 | AO2 | AO3 % |
|----------------------------------|-----|-----|-------|
| | % | % | |
| 1. Written Examination | 100 | - | - |
| 2. Essay | 57 | 29 | 14 |
| 3. Team Project | 57 | 14 | 29 |
| 4. Cambridge Research Report | 80 | 7 | 13 |
| Weighting of AO in qualification | 75 | 11 | 14 |

Figure 5: A level Assessment Objectives Weightings

The most heavily weighted assessment objective (AO1) covers the key academic thinking and analytical skills as required for an academic course, while the 'softer' skills of reflection and collaboration count for a lower percentage.

This reflects not so much their relative importance but rather the requirements of an academic syllabus which is equivalent to other IGCSE or A Level qualifications, and is recognised by universities and other higher education institutions across the world, together with the constraints of reliable assessment of these areas of reflection and collaboration. For example, with the Team Project in IGCSE a common mark is shared by all members of the team. That is, we have a common mark for all members of the team that reflects the group's collaboration. This can only be assessed by the teacher, and all members of the team are awarded the same mark. It is important that

the team's collaboration is explicitly recognised, but acknowledgement of the limitations of reliability of teacher assessment and 'fairness' of a group mark mean that the shared mark is relatively small.

The balance changes at A Level. As most students are using their A Levels for admission to higher education, all marks for the Team Project are individual marks. With A Level the collaboration is still assessed, but marks are awarded for the individual student's analysis and evaluation of the quality of their team's effectiveness and collaborative endeavour in achieving their outcome, through the requirement to write a reflective paper and present a project.

Again, as would be expected, the assessment objective descriptors change between IGCSE and A Level as the skills and ability of the student develops over time. This reflects the 'spiral of learning' approach to education

espoused by Cambridge and cited in 'Implementing the curriculum with Cambridge: a guide for school leaders': *'a spiral approach to skill development with concepts revisited and engaged with at deeper levels in different contexts, dependent on the learners' developmental stage'*.

Worldwide Assessment

These syllabuses are delivered across the world with thousands of candidate entries from all types of schools whether state funded schools in the USA and the Netherlands, or independent ones in Zimbabwe or Malaysia. In fact Global Perspectives has been our fastest growing subject over the past three years.

In addition, Global Perspectives has as a subject been taken on by several of our Ministry of Education customers - Mauritius, Egypt and Kazakhstan - for teaching in their country's curriculum

Research and Assessment results

A further piece of work, 'Collaboration in the 21st century – implications for assessment' (Child and Shaw, 2016) includes a focus on defining the construct of collaboration and then locating collaboration in Global Perspectives syllabuses. This construct was related to the concept of the collaborative 'state'. The researchers identified six elements of the process of collaboration that maintain equal importance in the establishment and maintenance of the collaborative state:

- *Social interdependence*
- *Conflict resolution*
- *Introduction of new ideas*
- *Sharing of resources*
- *Cooperation/task division*
- *Communication*

The researchers went on to analyse the Global Perspectives syllabuses and found that they engendered all six elements of collaboration. A position statement on collaboration was also produced:

'Cambridge research has shown that collaboration can be defined as comprising six elements: social interdependence, conflict resolution, introduction of new

ideas, sharing resources, cooperation and communication. In order to assess collaboration accurately the whole construct of collaboration is considered, not each individual element. This allows a holistic 'best-fit' approach to assessment which enhances the validity of the assessment for this skill. Collaboration and its assessment are enhanced if the members of the collaborative group have different perspectives which need to be negotiated to find the shared outcome. This requires the task to be constructed to present appropriate challenge.'

This form of research feeds into our reviews of all our Global Perspectives qualifications.

Additionally, in practice one good way of looking at the success of a qualification is looking at its ability to discriminate between candidates, using all parts of the mark scheme, and assessing effectively use and level of skills in the Global perspectives construct.

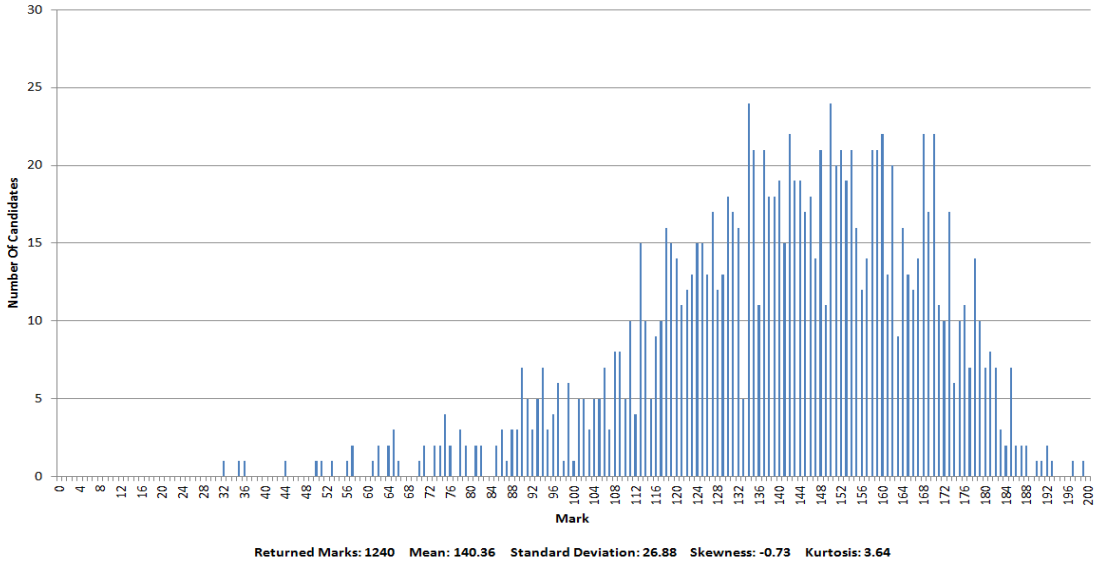


Figure 6: Mark Distribution: (0457) Global Perspectives (IGCSE) June 2016

Discussion

There are two key areas that merit discussion here around the assessment of these skills

Firstly, the possible tension between assessment validity and the assessment of these ‘soft’ skills, and secondly the support for teachers of how best to teach and assess these types of explicitly skills-based programmes.

Assessment has to be valid, reliable and fair. How then should an examination board assess collaboration? In this paper we have shown that through research we have designed a construct for collaboration, so that we know what we mean by collaboration and then designed an assessment, a team project, to assess that construct. As an exam board we have balanced the weighting of marks for individual work with that of the collaborative work. We have erred on the side of validity rather than teacher observation only. Have we got that balance right? Research and teacher evidence convinces us that we have. However, we need to ensure that we monitor this balance closely

Referring to the second point about teacher support, Cambridge International Examinations offer an Online

Learning Area with extensive materials specifically for Global Perspectives. This is freely available to all Cambridge centres delivering or considering delivery of Global Perspectives.

This Online Learning Area provides online structured courses, with guidance for both students and teachers, and multimedia resources, together with teaching and learning activities. Secure personal journals encourage higher-level thinking amongst students as they engage with and reflect on the materials and activities. Spaces with forum discussions allow for collaboration opportunities in a range of ways including between schools, in topic-based groups, and within students’ own class groups. Teachers can also collaborate with other Global Perspectives teachers in different parts of the world.

Students can use ePortfolio tools to gather together and share their research and gain feedback from peers and teachers.

Teachers and students can access the Online Learning Area using any connected device, including tablets, laptops and desktop machines. We strongly suggest that teachers request student accounts, as giving students their own access means that the area can support ‘flipped’ learning opportunities, providing students with access outside the traditional classroom environment. We also provide a full set of face-to-face training as well supporting teachers with our online self-study Coursework Training Programmes and our online tutor led introductions.

Conclusions

What have we learned during these initial few years in the development of these interdisciplinary syllabuses? Firstly, the take-up from centres shows its increasing popularity and evidence from the research is positive. The qualitative feedback from teachers and students is encouraging, as illustrated in the following extracts:

“The Global Perspectives course has honed my writing and research skills. It has also been an enriching and eye-opening experience for me, allowing me to understand current global issues in depth.” (Hong Kong, student)

“It’s really about real life. It’s not about reading a textbook and studying that. It’s about how you see the world and how it’s constantly changing.” (India, student)

“Cambridge Global Perspectives is rigorous, it’s academic, it’s challenging, and because Cambridge is very self-driven, I think that those students are going to be at a distinct advantage when they go to university.” (New Zealand, teacher)

Furthermore, the relationship between Global Perspectives and the more traditional content syllabuses is robust and balanced, as Cambridge understands the need for both skills and knowledge to deliver an effective curriculum. There is no false dichotomy between skills and content; they are two sides of the same coin. All Cambridge syllabuses are focussed on helping schools offer a well-balanced curriculum to deliver the Cambridge Learner attributes (Responsible, Reflective, Innovative, Engaged and Confident). However there are

several areas of challenge to guide future development of the syllabuses. One is the balance between internal assessment, moderated by Cambridge, and external assessment carried out by Cambridge. Teachers in many cases are best placed to assess the skills being developed, particularly that of collaboration, but to ensure robustness and validity this needs to be moderated. How much of the collaboration can be moderated effectively? Teachers need to have sufficiently clear and unambiguous assessment criteria, as do moderators, whose feedback to schools must be clear and consistent. Syllabus support resources must focus as much on the different role of the teacher as a facilitator, and on the importance of developing team work, as on any subject knowledge that students need.

One future possible development is the use of technology through eAssessment of the skills being used, on our Online Learning Area. It is relatively simple to assess the quality of collaboration; what is less clear is the most valid ways of measuring the quality of collaboration. What is the best balance between the AOs that will ensure academic rigour and the development of the ‘softer’ skills such as reflection, and communication that are equally valid? These are the key questions asked whenever one of the Global Perspectives syllabuses is reviewed. At present a 5-14 Primary and lower Secondary Global Perspectives programme is being developed, and although assessment is lower stakes than at IGCSE or A Level, the assessment still has to be robust, valid and fair. As is entirely appropriate when considering a syllabus such as Global Perspectives, there is no one correct approach to assessment of the subject: rather an on-going critical evaluation of the challenges and opportunities is developed over time, and based on the increasing body of research available. As Eric Hoffer, a self-taught moral philosopher noted: *‘In a time of drastic change it is the learners who inherit the future.’*

References

- Assessment and Teaching of 21st-Century Skills (2016)
Available via
<http://education.unimelb.edu.au/arc/projects/completed/2012/atc21sg>
- Child, S., and Shaw, S. (2016). Collaboration in the 21st century: Implications for assessment. *Research Matters: A Cambridge Assessment publication*, 22, 17–22.
- Hoffer E. (2006) *Reflections on the Human Condition*, Titusville, NJ: Hopewell Publications
- Kolb D. A. (1984) *Experiential Learning: experience as the source of learning and development*, Englewood Cliffs: Prentice Hall
- National Soft Skills Association (2015) *The Soft Skills Disconnect*, Available via
<http://www.nationalsoftskills.org/the-soft-skills-disconnect/>
- OECD (2016) Available via
<https://skills.oecd.org/documents/OECDskillsStrategyFINALENG.pdf>
- Partnership for 21st Century Learning (2016) Available via
<http://www.p21.org/our-work/p21-framework>
- Seldon, Sir Anthony (2015) *Lecture to Tatler Schools Live! Conference*, Available via
<http://www.bbc.co.uk/news/education-34417189>
- Suto, I. (2013) 21st Century skills: Ancient, ubiquitous, enigmatic? *Research Matters: A Cambridge Assessment*, 15, 2-8
- The Asia Society (2016a) Available via
www.asiasociety.org/education
- The Asia Society (2016b) *Five reasons why Global Competence matters*, Available via
<http://asiasociety.org/global-competence>
- Wagner, T (2016) *Seven Survival Skills*, Available via
<http://www.tonywagner.com/7-survival-skills>

SUB THEME F: RELEVANCE OF APTITUDE TESTS IN SELECTION AND PLACEMENT

The Perceptions of Users of Aptitude Tests and its Suitability for Recruitment of Graduates in Ghana

Robert G. Dery (Rev.)

+233-243558825, rgdery@waecgh.org

The West African Examinations Council, Accra-Ghana

Abstract

A high number of graduate unemployment has been recorded in Ghana over the past few years. Institutions that put up advertisements for recruitment receive a large number of applications from applicants. Employers are faced with the challenge to select the best applicants from the available large pool of job applicants. Some institutions have therefore adopted aptitude tests developed by the West African Examinations Council (WAEC) Ghana to shortlist applicants for employment. The study sought to find out the perceptions of stakeholders on aptitude tests, its effectiveness and suitability in recruitment decisions. The descriptive survey research design was used for the study. The data for the study was collected through questionnaire using convenience sampling technique, administered to 100 participants from Ministry of Foreign Affairs and Regional Integration, Services Integrity Savings and Loans Company, FBN Bank and some staff of the West African Examinations Council, Accra. Quantitative descriptive analysis was used to analyse the data. The questionnaire was used to gather the perception of institutions that ever used aptitude test for recruitment. The main findings showed that institutions which adopted the aptitude test as an assessment instrument stand a better chance to make sound judgments about the choice of preferred applicants. Based on the findings it is recommended that organizations which seek to engage applicants for employment should adopt aptitude tests as an assessment tool for selection and placement of applicants into their institutions.

Keywords: aptitude tests, selection, placement.

Introduction

In recent years, graduate unemployment, turnover rate from the universities is high. A number of graduates who complete their course seek for jobs from various institutions in the global world. Although the benefits of higher education vary according to the course studied and the institution attended, the demand for university graduates for employment generally exceeds the facilities available in institutions. Employers are therefore faced with the challenge of recruitment of qualified applicants amongst many to take up appointments in public or private institutions.

In Ghana over the past few years, a high number of graduate unemployment has been recorded and this has

led to the formation of an association of unemployed graduates. When an institution advertises a job in a national newspaper, a high number of graduates put in applications for the job. The employer receives a large number of applicants and is faced with the challenge of selecting qualified graduates from the pool for employment into institutions.

There are many leading practices that can be employed to recruit and hire highly qualified graduates. Local and international companies in both the private and public sectors rely on pre-employment tests, such as aptitude and

personality tests, as the most effective method to measure fitness for a position applied for. These tests tell

employers what the graduates need to know, not just what they want to share with the institution. Most of the institutions / companies often use aptitude test scores to predict performance of graduates who seek employment into their outfits.

(Higgins & Sun, 2002) stated that psychological testing was not a recent development. Its origins could be traced back to China over 3000 years ago when the Emperor had his officials assessed, to examine the extent to which they were fit for office. The assessments were based on the assumption that by evaluating the performance of individuals over a relatively short period of time in a particular sphere of knowledge (eg. military affairs), it was possible to predict their future performance in work. Modern aptitude test in various fields predicts a person's performance of activities in the future.

In later research, Wissler (1901) cited that Galton F. focused on a set of psychological dimensions to be measured, standardised measurement procedures and the use of objective measurement techniques, have had a lasting impact on the field of psychological testing.

Dewberry (2011), p.39 argues that "aptitude test measure general cognitive ability" based on some forms of reasoning linked to verbal, numerical, logical, and quantitative, etc. to the extent that these tests measure cognitive ability rather than specific areas of job knowledge and the tests are unlikely to provide a better prediction of job or training performance than a test designed to measure cognitive ability.

Aptitude test is identified as one of the leading practices in the recruitment process of graduates into institutions/companies. It is a tool used to determine and measure an individual's ability acquired through the process of study and the set of skills gained in education. There are several aptitude tests in the market and the one chosen is the differential aptitude test which serves as

series of assessments designed to measure an individual's ability to learn or succeed in a number of different areas performed under strictly timed examination conditions. Some institutions in Ghana have adopted aptitude test as an option for fair and transparent assessment for recruitment of job applicants. The difficulty that arises for institutions is the choice to make from an increasingly large number of highly – qualified applicants who achieve higher grades in their respective fields of study for various jobs skills. Aptitude test might not be the ultimate in the selection of job applicants rather it can enhance the selection process which calls for shortlisted applicants for interview.

Schmidt (2002) reported that "the information used to assess university applicants may not be equally reliable. Although "prior education attainment remains the best single indicator of success at undergraduate level"(p.5).

While aptitude tests are norm-referenced and do provide information for graduate employment, their results are used differently for decision making. According to their scores, employees are placed in various departments. If a high proportion of applicants who score well on a certain test go on to be successful at work and those who score lower are somewhat less likely to be successful at work if employed, we would say that the test has good predictive validity. The main benefit of developing and implementing aptitude test to institutions is that it enables them to select candidates according to their preferences and context requirements. It also assists management to shortlist applicants for interview and subsequently placed in the organization. The designed aptitude test incorporated accuracy, speed and paying attention to details.

Statement of the Problem

Many students from the tertiary institutions complete their field of study every year and find themselves in search of jobs. The challenge of selecting qualified graduates from a pool of unemployed graduates for recruitment has become a herculean task to employers. Despite the high efforts made by governments and institutions to create jobs the problem still persists. The purpose of the paper is to examine the use of aptitude test in making a sound judgement by employers in the recruitment process and also understand every skill of a candidate and decide whether the candidate is fit for the job or not.

Purpose of the Study

The main purpose of the study is to investigate the use of aptitude test as a means for recruitment process in the employment of graduates into institutions or companies and to determine the frequency and feedback from institutions and the extent to which aptitude test is accepted as a tool for recruitment process.

Significance of the Study

In this study the appropriate use of assessment for recruitment of unemployed graduates will segregate the graduates who seek for employment into various institutions. The study is useful in proposing aptitude test as a better option for engagement of graduate for employment.

Research Question

The following research questions guided the study:

1. What is the purpose of using aptitude test for recruitment process?
2. What is the employers rating performance of employed graduates?
3. What is the perception of employers on the use of aptitude test as a means of recruitment?

Scope and Limitation of the Study

The study covered four institutions in Ghana that ever employed graduates from the tertiary institutions by use of the aptitude test conducted by the West African

Examinations Council (WAEC), Accra. Other institutions were not included in this study mainly due to various limitations such as time constraint that the study encountered.

Test Development and Administration

Nature of the test

Aptitude tests are structured systematically ways to evaluate how people perform tasks or react to different situations and gives an accurate insight into a person's potential and the ability to cope with change. They are meant for selection, placement, recruitment, promotion, etc. and is very competitive. It is therefore imperative that all security measures are adopted in the development and administration of the tests (WAEC, 2002).

Test Papers and Stationery

The Records and Aptitude Tests Department (RATD) of the Council prepares the tests such as Quantitative Aptitude, Logical Reasoning, Supervisory Judgement, Verbal Aptitude, English Language, Logical and Analytical Reasoning and also responsible for the conduct of the tests. The test materials are kept in the Strong Room or in a secured place under lock and key until the day of the tests. The tests papers are sent to the examination hall in security bags obtained from the Strong Room. Candidates are properly identified, searched, seated in index number order and are provided with blank objective cards for shading (if multiple-type questions) for their responses and/or answer booklet for writing (if easy-type questions).

Conduct of test

The conduct of the test is carried out in large hall/classroom only and the furniture arranged to reflect examination requirements of WAEC standards in such a manner that the security of the test papers are not at risk.

Rules of the conduct of the test

The relevant rules of the test include;

- ❖ Candidates should be searched and admitted into the hall, in good time to ensure that the test is not unduly delayed;

- ❖ Candidates should be identified with their ID cards before being allowed into the examination;
- ❖ Candidates are to bring into the examination hall their individual pencils, eraser, rulers or mathematical sets since they would not be allowed to share any of these items.
- ❖ Candidates are allowed to sit in index number order;
- ❖ Candidates are not allowed to bring in mobile phones or calculators.

Conduct of Test

The security of the test papers is the utmost concern to the officer who is responsible for the distribution and collection of the test materials from candidates. Whilst the test is in progress, officers are very vigilant throughout the test period to ensure that there is sanity throughout the exercise. When the test is over the officers ensure that no candidate leaves the hall until test papers are first collected from the candidates, who are further checked before dismissal from the examination hall. All the question papers used for the exercise are locked up securely in security bag(s) and returned to the Council. The test responses are processed and results released in order of merit and in good time.

Research Method

Research Design

This research is a cross-sectional survey design. Information was collected from FBN Bank, MFARI, Services Integrity Savings and Loan Company (SISLC) and from some staff of the West African Examinations Council. Poopola et al. (2015) cited Awotunde and Ugodulunwa (2002) that cross-sectional design is concerned with collecting data from members of a population in order to make judgments about conditions that exist, opinions that are held, processes that are going on and effects that are evident regarding phenomenon. The study used quantitative descriptive analysis. The

instrument employed was a questionnaire which was administered, using purposive sampling technique.

Participants

A total of seventy-seven (77) officials in three institutions including some staff of the West African Examinations Council, Accra presented their responses to the questionnaire for the study. The participants represented the institutions that benefited from various services rendered to them.

Materials

A questionnaire was used for the study. It had two sections; A and B. Section A was the demographic data of the respondents while section B of the questionnaire sought the impact of the aptitude test on the operations of the institution. In order to ensure the validity of the questionnaire, the items were constructed after a careful review and revisions of the draft before and after a pilot test on some staff of the West African Examinations Council, Accra. A Professor and lecturer in educational measurement and evaluation from the Department of Educational Foundations, University of Cape Coast reviewed the draft of the questionnaire.

Procedure

The questionnaires were administered by the author within one-month period. The participants in the study were assured of their confidentiality of the information they would provide. In the questionnaire the names of the participants were not required in the study. Participants provided their position (rank) in their places of work. Out of 100 questionnaires administered, 77 returned their responses.

Data Analysis and Interpretation of Results

Data were analysed using SPSS package and reported using descriptive statistics. The statements in the questionnaire were sorted in accordance to a numerical code by defining and labelling each of the variables and assigned numbers to each of the possible responses. The data were analysed thematically by grouping similar statements. The statements were further analysed by coming up with the frequency of each response and the

percentages of each statement’s frequency for ease of interpretation of the data. The analysis is also based on the various positions of the respondents namely

Directors, Relationship Managers, Customer Service Officers, Accountants, HRM, and other individuals who willingly responded to the questionnaire.

Table 1 shows the distribution of the various positions to the questionnaire.

| | | Position | | | |
|-------|---------------------------|-----------|---------|---------------|--------------------|
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Directors | 22 | 28.6 | 28.6 | 28.6 |
| | Relationship Manager | 19 | 24.7 | 24.7 | 53.2 |
| | Customer Service Officers | 11 | 14.3 | 14.3 | 67.5 |
| | Accountants | 11 | 14.3 | 14.3 | 81.8 |
| | HRM | 8 | 10.4 | 10.4 | 92.2 |
| | Others | 6 | 7.8 | 7.8 | 100.0 |
| | Total | 77 | 100.0 | 100.0 | |

Source: SPSS 20.0

Table 1 shows a summary of the respondents on the various aspect of the research activity. A total of 77 respondents to the questionnaire out of 100 that were distributed to the respondents. Out of a total of 77

respondents, “Directors” had the highest number of 22 respondents represent 28.6% and “Others” (persons who do not fall under any of the above categories) represent the lowest number of 6 (i.e. 7.8% of the respondents).

Research Question 1

What is the purpose of using aptitude test for recruitment process?

Table 2: Showing the purpose of the use of aptitude test

| Position | Selection Only | | Placement Only | | Both Selection & Placement | | Effective Instrument | |
|---------------------------|----------------|-----------|----------------|--------------|----------------------------|--------------|----------------------|--------------|
| | Yes | No | Yes | No | Yes | No | Yes | No |
| Directors | 9 (33.3%) | 2 (40.0%) | 2 (20.0%) | 1 (25.0%) | 14 (28.6%) | 1 (100.%) | 13 (27.7%) | 6 (4.29%) |
| Managers | 4 (14.8%) | - | 2 (20.0%) | - | 13 (26.5%) | - | 15 (13.9%) | - |
| Customer Service Officers | 5 (18.5%) | - | 1 (10.0%) | - | 7 (14.3%) | - | 8 (17.0%) | 1 (7.1%) |
| Accountants | 5 (18.5%) | - | 3 (30.0%) | - | 7 (14.3%) | - | 6 (12.8%) | 2 (14.3%) |
| HRM | 3 (11.1%) | - | 2 (20.0%) | - | 3 (6.1%) | - | 4 (8.5%) | - |
| Others | 1 (3.7%) | 3 (60.0%) | - | 3 (75.0%) | 5 (10.2%) | - | 1 (2.1%) | 5 (35.7%) |
| Total | 27 | 5 | 10 | 4 | 49 | 1 | 47 | 14 |

Source: Instrument of the Study

Table 2 shows the number of respondents indicating the purpose of aptitude test as a means for selection and/or placement for recruitment process. A total of 49 responses indicates that aptitude test is used for both selection and placement in the recruitment process, whilst

27 and 10 responses indicates that it is for selection and placement only respectively. On the effectiveness of aptitude test as a tool for recruitment, 47 responses have it as an effective tool for selection and/or placement.

Research Question 2

What are the employers rating performance of employed graduates?

Table 3 shows the rated performance of employed graduates

| Position | Outstanding | % | Very Good | % | Good | % | Satisfactory | % |
|---------------------------|-------------|------|-----------|------|-----------|------|--------------|------|
| Directors | 5 | 71.4 | 8 | 30.8 | 4 | 20.0 | - | - |
| Managers | 1 | 14.3 | 9 | 34.6 | 3 | 15.0 | 1 | 50.0 |
| Customer Service Officers | 1 | 14.3 | 6 | 23.1 | 2 | 10.0 | - | - |
| Accountants | - | - | 3 | 11.5 | 3 | 15.0 | 1 | 50.0 |
| HRM | - | - | - | - | 4 | 20.0 | - | - |
| Others | - | - | - | - | 4 | 20.0 | - | - |
| Total | 7 | | 26 | | 20 | | 2 | |

Source: Instrument of the Study

Table 3 showed the responses of the respondents on the performance of the graduates employed into their institutions. It is interesting to note that 26 responses rated the performance of the employed graduate as very good.

The Managers represent the highest responses with a percentage of 34.6%.

Research Question 3

What are the perceptions of employers on the use of aptitude test as a means of recruitment?

Table 4 shows the perceptions of some employers in institutions

| Position | Ability | % | Good | % | Fair | % | Useful tool | % | Not Useful tool | % | Others | % |
|---------------------------|-----------|------|-----------|------|-----------|------|-------------|------|-----------------|------|-----------|------|
| Directors | 6 | 42.9 | 3 | 25.0 | 7 | 36.8 | 3 | 18.8 | 2 | 33.3 | 1 | 10.0 |
| Managers | 4 | 28.6 | 4 | 33.3 | 1 | 5.3 | 6 | 37.5 | - | - | 4 | 40.0 |
| Customer Service Officers | 1 | 7.1 | 2 | 16.7 | 1 | 5.3 | 5 | 31.3 | 1 | 16.7 | 1 | 10.0 |
| Accountants | 1 | 7.1 | 1 | 8.3 | 6 | 31.6 | - | - | 1 | 16.7 | 2 | 20.0 |
| HRM | 2 | 14.3 | - | | 3 | 15.8 | 2 | 12.5 | - | - | 1 | 10.0 |
| Others | - | - | 2 | 16.7 | 1 | 5.3 | - | - | 2 | 33.3 | 1 | 10.0 |
| Total | 14 | | 12 | | 19 | | 16 | | 6 | | 10 | |

Source: Instrument of the Study

Table 4 shows the number and percentage of respondents on the perceptions about the use of the aptitude test for recruitment. The respondents' perceptions were grouped under five thematic groups; ability, good, fair, useful and not useful. The Table depicts the analysis carried out on the perceptions of the respondents on the use of aptitude test as a process for recruitment. The Directors (18.8%), Managers (37.5%) and Customer Service Officers (31.3%) all show that aptitude test is good and a useful tool in recruitment process. Furthermore, 42.9% for Directors, 28.6% for Managers indicated that those employed have best ability to cope up with the work in the institutions. The test enables employers to identify the ability of an employee based on perfect selection of the right and suitable candidate for the job and place him/her at the appropriate job schedule.

Discussion of Results

These days, any job vacancy is likely to attract a large pool of potential candidates. Pre-screening these applicants can help reduce the number to a more manageable size who will then go forward into a more rigorous screening phase. In general, the results of this study reveal that aptitude test is a major contributor to the recruitment of applicants into institutions or companies. There is a consensus of opinion on the issue from the respondents; directors, managers, customer service officers that aptitude test is an effective instrument which enables employers to make an informed decision on the caliber of persons who should be offered opportunity to serve in the institution.

There are a number of advantages that aptitude tests have in combination with entire job search process. According to (Hunter & Schmidt, 1983), studies show that aptitude tests are extremely successful in predicting potential for success which is in tandem with the responses to research question 2 of item 7 of the questionnaire. These tests are accurate in determining a person's strengths and weaknesses and provide an in-depth look at the candidate that interviews alone may not be able to do and offers

insights to candidate's potential for success. In conjunction with interview, the tests give a very accurate measure of an employee's potential to be successful in the position they pursue. The usefulness of an employee's assessment is a combination of the quality of the test content and the manner in which results are interpreted. Different jobs require employees with different combinations of abilities, skills and personality characters connecting the assessment to the job.

Conclusion

The paper concludes by reiterating the importance of aptitude test in the processes of recruitment of job applicants into institutions. It is observed that aptitude test has the ability to guide employers in the selection and placement of applicants at the right position of the job. Many employers who receive applications from job applicants are faced with challenge of how to select the right persons for the job in terms of the knowledge base of the applicant, so aptitude test will enable a quality decision to be taken to trim the number to a manageable size for the employer to take the necessary steps for employment.

Recommendation

Based on the findings of the study, it is recommended that:

1. Aptitude test should be used by employers as a tool for selecting and placement of job applicants.
2. Employers should adopt aptitude test which will assist them to shortlist applicants from the large pool for interviews.

A further study that should focus on the test batteries and how test content could predict the individual success at the job place since this study focused on the perception of stakeholders on the use of aptitude test as a tool for selection and placement.

Reference

- Dewberry, C. (2011). Aptitude testing and the Legal profession. p.39. Birkbeck, University of London.
- Higgins, L. T., & Sun, C. H., (2002). The development of psychological testing in China. *International Journal of Psychology*, 37(4), 246-254.
- Hunter, J. E., & Schmidt, F. L., (1983). Quantifying the effects of psychological interventions on employee job performance and work-force productivity, Vol 38(4), 473-478 <http://psycnet.apa.org/journals/amp/38/4/473/>
- Popoola K., John I., & Aminat E., (2015). Using Teachers's knowledge of Bloom's Taxonomy of Educational objectives as a Framework for Evaluating the Quality of Classroom Assessment. *Journal of Educational Assessment in Africa*, Vol 10, 2015
- Schmidt, F. L (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, 15(1-2) 75-96.
- WAEC (2002). Test Development Division of Council. Manual for Aptitude Testing Exercise for administration of tests.
- Wisler, C. (1901). The correlation of mental and physical tests. [Monograph Supplement]. *The Psychological Review*, 3(6).
- www.scribd.com. Differential Aptitude Test. Sample report

Options functionality of English Language tests for Candidates' Selection and Placement into Tertiary Institutions in Nigeria: A pilot study

¹Prof. Henry O. Owolabi – henryowolabi2000@yahoo.com +2348033733311

²Prof. Adams O. U. Onuka – adamonuka@yahoo.com +2348033564064

³Dr. Mayowa O. Ogunjimi - ogunjimimayowa@gmail.com +2348033912362

⁴Dr. Yusuf Mustapha – bbmusty@yahoo.com +2348031857059

⁵Dr. Lukman Adaramaja – adaramaja4real@yahoo.com +2348061254474

⁶Dr. Dorcas Daramola – immaculatethabitha@yahoo.com +2347031889036

^{1,3,6}Faculty of Education, University of Ilorin, Ilorin

²Institute of Education, University of Ibadan, Ibadan

⁴Faculty of Education, University of Jos, Jos

⁵Alvan Ikoku College of Education, Owerri

Abstract

Effectiveness of options supplied to an objective test item is reflected by the number of examinees selecting them. West African Examinations Council (WAEC), National Examinations Council (NECO) and National Business and Technical Examinations Board (NABTEB) conduct Senior School Certificate Examinations (SSCE) and its equivalents for the certification of school based and private candidates while the Joint Admissions and Matriculation Board (JAMB) conducts the Universities and Tertiary Institutions' Matriculation Examination (UTME) for selection and placement of candidates into tertiary institutions across Nigeria. Although examination bodies ensure the functionality of items used by them, it has become necessary for researchers in psychometrics to carry out confirmatory tests to determine the extent to which options supplied in certification and placement tests are functioning. This pilot study was therefore designed to determine the functionality of options in the 2015 SSCE English Language multiple choice test items used by two examination bodies in Nigeria – WAEC and NECO. A total of nine hundred and sixty final year senior secondary school students presented for school certificate examinations in 2017 were sampled across the country. Item analysis statistics revealing percentage of examinees selecting each option to multiple choice items were computed. It was found that all options supplied by WAEC to 2015 SSCE English Language were either fairly or highly functional while 4.4% of those supplied by NECO were not functional. WAEC had larger proportion of highly functional options compared with NECO. Presence of nonfunctional options implies that sufficient attention has not been given to careful supply of options during the test development process. A need for further investigation was identified. It was thus recommended that a countrywide study in all the subjects examined at the school certificate level by each of the examination bodies should be conducted. More attention should also be given to issues of quality of options supplied to test items before they are administered to improve public examinations used for selection and placement of candidates into tertiary institutions.

Key Word: Options functionality, test quality, selection and placement of candidates,

Introduction

The use of Multiple Choice Tests (MCTs) has become prominent in the school system. There's no way learners will pass through the school system at any level that they will not write multiple choice tests. MCTs feature prominently in teacher-made tests. They, at the same time, play dominant roles in it. Thus, such instructional decisions as selection, placement, promotion, effectiveness and success taken in the school system are outcomes of MCTs. In a similar vein, examination bodies have found the use of MCTs inevitable and every subject has a dedicated objective test paper comprising only multiple-choice items. Test experts and researchers also heavily depend on the use of MCTs for guiding decision taking, policies, and practices. The increasing reliance of the educational system on MCTs has necessitated further research to ensure they possess the qualities that make their use necessary and for achieving the goals for their use.

MCTs play crucial roles in both internal teacher-made and external high stakes assessments which necessitate the use. Among the many reasons for the use of MCTs may be the fact that they are seen to be objective in terms of development and scoring and thus offer the fairest opportunity to testees and testers to prove their competence and integrity respectively. Osunde (2009) asserted that MCTs are seen as most reliable because of their consistency in scoring and fairness to students. Wide coverage of curriculum contents and objectives of instruction is achievable with the use of MCTs. Brane (2013) affirmed that the MCT has the potential advantages of versatility, reliability and validity. Items on MCTs could also be scored rapidly, providing quick feedback to students and efficient when assessing large number of students over broad subject contents. Lawal (2001) and Kolawole (2007) posited that MCTs are useful in assessing learners' mastery of specific facts, concepts, terms, laws and principles. However, one major drawback in using MCTs is that constructing good multiple-choice items requires plenty of time for writing,

review, and revision. The use of MCTs in certain educational fields is sometimes contested due to some of the negative aspects whether actual or perceived but the format remains popular due to its reliability, utility and cost effectiveness (Steven, 2004).

The structure of an MCT item consists of a problem which is known as the stem, followed by suggested solutions, referred to as options. The options consist of one correct or best alternative called key, which is the answer, and other incorrect or inferior alternatives, known as distractors (Owolabi, 2009; Onuka 2009). Distractors are the multiple-choice response options that are not correct. Their functionality resides in the fact that they are plausible but incorrect. Functionality of options is enhanced because they are developed based upon examinees' common misconceptions, wrong assumptions or miscalculations. Consequently, options functionality is the extent at which distractors of multiple choice item perform the role of drawing the attention of examinees with poor reasoning as expected by test developers. In other words, it is the rate at which the incorrect options serve as distractors by being attractive to examinees who do not have the knowledge or skills for getting the answer. Non-functioning options are therefore those distractors that are selected by a small proportion of examinees. For any options to serve as a good distractor, it should have no connection or direct relationship with the answer, and neither does it serve as a clue to it (Abdulghani, Ahmad, Ponnamperna, Khalil & Aldrees, 2014).

The usefulness of MCTs in achieving objectives of testing is a function of the quality rather than quantity of its options. The effectiveness of options is judged by how plausible or functional the distractors are. A distractor is considered as effective if at least 5% of the testees select it (Rodriquez, 2005; Olatunji, 2007; Tarrant, Ware & Mohammed, 2009; Theodorsson, Shafil, Wardy, Khan, Mahrezi & Shafae, 2010). Distractor analysis is the frequency of selection of each option different from the

key. The higher the frequency of selection of an option different from the key, especially by testees in the lower group in relation to the total number of the examinees in the high and low groups, the higher its degree of functionality as a distractor (Alonge, 2004). A good distractor should be chosen by more examinees in the lower group. If otherwise, that particular distractor should be scrutinized, revised or discarded.

Experts in psychometrics have suggested that three to five options should be written for each MCT item (Grier, 1975; Rodriguez, 2005; Olatunji, 2007). The analysis of options functionality by Tarrant, Ware and Mohammed (2009) revealed that only 13.8% of all items used in their study had up to three functioning distractors while more than 70% had only one or two functioning distractors. The low proportion of items with up to three functioning distractors was not altogether surprising given that all tests were generated by in-house teaching faculty, many of who have minimal training in item writing. While this may happen with teacher-made tests investigated by Tarrant, Ware and Mohammed (2009), there is a greater need to obtain empirical evidence to determine whether options in items on standardized tests function better.

Attempts have been made by psychometrists to determine the most appropriate number of options. Rodriguez, Kettler, and Elliott (2014) carried out a quasi-experimental study to determine the effect of test modifications on distractor functioning. The researchers removed the weakest of three distractors from 39 items in each of reading and mathematics. It was found that the functionality of distractors was neither systematically improved nor weakened by the modifications. It was however found that more than 70% of the distractors had increased discrimination power. MCTs with four options were not found to have more functional options than those with three.

Most examination bodies feature tests with items having four or five options. The West African Examinations

Council (WAEC) supplies four options to their SSCE test items while National Examinations Council (NECO) supplies five options. The study by Haladyna and Downing (1989) revealed that approximately two-third of all four-option items they reviewed across subjects had only one or two functioning distractors and none of the five-option items had four functioning distractors. It is often difficult for test developers to provide three or more equally plausible distractors.

Though it seems attractive to develop MCTs with fewer options, contrariwise, Ramons and Stern (1993) observed that developing tests with 4- or 5- option formats make it very difficult for examinees to guess the answer correctly. Olatunji (2007) found that WAEC's 4-options in Economics had 82% functioning options while NECO's 5-options had 56.7%. Also, Olutola, Owolabi and Daramola (2015) found that WAEC's 4-Options in Biology had 95% functioning options while NECO's 5-options had 93% functionality in the same subject. There is need to study whether the situation is the same with other core subjects, particularly English Language which is not only the medium of instruction but should compulsorily offered and passed for admission into any higher educational institution in Nigeria and most parts of the world.

The purpose of this pilot study therefore is to determine the extent to which options in the SSCE 2015 English Language MCT by WAEC and NECO were functional. The following research questions were therefore addressed:

What are the functionality levels of the options supplied by WAEC in the 2015 SSCE English Language MCT?

What are the functionality levels of the options supplied by NECO in the 2015 SSCE English Language MCT?

What are the levels of functionality of options supplied by WAEC compared with NECO in the 2015 SSCE English Language MCT?

This hypothesis was also tested:

H₀: Functionality of options used in 2015 SSCE English Language is independent of number of options used by examination bodies.

H₁: Functionality of options used in 2015 SSCE English Language is associated with number of options used examination bodies.

Methodology

The survey design was adopted with a deliberate focus on English Language MCTs. The population for the study was final year senior secondary school students writing the SSCE in Nigeria. This group was deliberately chosen because it had gone through preparation for SSCE by the various senior secondary schools. The target population for the pilot study were senior secondary school students in Lagos State and Federal Capital Territory, Abuja. These two locations were considered to be the base of students from the entire country. Lagos had played the role of capital city for Nigeria from the colonial days and for more than three decades after independence and is still the commercial and industrial capital of the country, while Abuja had been the capital city of the Federal Republic of Nigeria for more than three decades. The large conglomeration of citizens from all the States and

regions constituting the Federal Republic of Nigeria makes the two locations suitable representative of the populace. A sample of 16 students drawn from each of Arts, Science and Commercial classes was randomly selected from each sampled senior secondary school. From each location 240 students were selected to take the WAEC version of the 2015 SSCE English Language and since the same number also took the NECO equivalent, a total of 480 final year students took part in the study from each location. Therefore, a total of 960 testees took the tests from Lagos and Abuja. In the study, Forty-eight students were sampled from each of the ten schools representing Lagos and Abuja. The instruments for the study were the WAEC and NECO standardized SSCE English Language MCTs used in 2015 and these were administered on the candidates who had finished writing the 2017 WAEC SSCE during the period they were getting ready for NECO SSCE. Percentage and chi-square were used for data analysis.

Results and Discussion

Research Questions:

- i. What are the functionality levels of the options supplied by WAEC in the 2015 SSCE English Language MCT?

The percentage of testees selecting each of the 320 options was obtained and those attracting less than 5% of the total were regarded as not functional. The options attracting from 5% to 15% of the testees were classified as fairly functional while all those attracting more than 15% of the testees were regarded as highly functional.

Table 1: Functionality of 2015 WAEC SSCE English Language

| Levels | Number of options | Percentage |
|-------------------|-------------------|------------|
| Not Functional | 0 | 0 |
| Fairly Functional | 41 | 12.8 |
| Highly Functional | 279 | 87.2 |
| Total | 320 | 100.0 |

From the results on Table 1, there was no non-functioning option supplied by WAEC in the 2015 SSCE English Language MCT. Only 41(12.8%) of the 320 options functioned fairly well while the remaining 279(87.2%) were highly functional. It implies that the four options supplied for the English Language MCTs to the SSCE 2015 items were good. This supports the suggestion that distractors tend to function better when they are fewer and that it makes it possible for test developers to supply quality options. This here finding partly agrees with those of Olatunji (2007) and Olutola, Owolabi and Daramola (2015) who investigated options functionality of Economics and Biology SSCE conducted by the same examination bodies, WAEC and NECO, used in the present study. They found out that WAEC 4-options in Economics and Biology respectively had 82% and 95% functional option. Though the findings were on different subjects conducted by the same examination bodies, it is an indication that irrespective of the subject, 4-options MCTs tend function well. This corroborates the findings of Olatunji and Owolabi (2009) that performance of

students in 4-options MCTs tends to be better than in those with five options.

Findings by Tarrant, Ware and Mohammed (2009) which revealed that over 70% of the items studied by them had only one or two functioning distractors is however different from findings of the current study. The high proportion of highly functioning options observed in the 2015 SSCE English Language MCTs is indicative of care exercised during the test development process. WAEC has over the years improved the quality of options supplied to items used in examining their candidates and this may be indicative of the knowledge and experience of its examiners, as well as efficiency of the test development and validation procedures.

What are the functionality levels of the options supplied by NECO in the 2015 SSCE English Language MCT?

The same procedure adopted for summarizing data to answer the first research question was also adopted for this research question. The summary is presented on Table 2.

Table 2: Functionality of 2015 NECO SSCE English Language

| Levels | Number of options | Percentage |
|-------------------|-------------------|------------|
| Not Functional | 22 | 4.4 |
| Fairly Functional | 210 | 42.0 |
| Highly Functional | 268 | 53.6 |
| Total | 500 | 100.0 |

The outcome on Table 2 revealed that 22(4.4%) of the options supplied to the 2015 SSCE English Language MCT were not functional, 210(42%) functioned fairly while the remaining 268(53.6%) functioned highly. This finding agrees with that of Olatunji (2007) and Olutola, Owolabi and Daramola (2015) that NECO 5-option items in Economics and Biology had 56.7% and 93% functional options respectively. With 95.6% options functioning out of the 500 supplied by NECO, a high level of efficiency seems to have been attained. The problem with the

remaining 4.4% however is the effect it has on the validity and reliability of the MCT as functionality of options promote these qualities.

What are the levels of functionality of options supplied by WAEC compared with NECO in the 2015 SSCE English Language MCT?

To answer this research question, the frequency and percentage of options supplied to the WAEC SSCE 2015 English Language MCTs were compared with those of NECO as shown on Table 3.

Table 3: Comparison of the functionality of WAEC and NECO option

| Levels | WAEC | | NECO | |
|-------------------|-------------------|------------|-------------------|------------|
| | Number of options | Percentage | Number of options | Percentage |
| Not Functional | 0 | 0 | 22 | 4.4 |
| Fairly Functional | 41 | 12.8 | 210 | 42.0 |
| Highly Functional | 279 | 87.2 | 268 | 53.6 |
| Total | 320 | 100.0 | 500 | 100.0 |

As shown on Table 3, WAEC supplied the larger percentage (87.2%) of highly functional options compared with NECO with 53.6%. In the category of fairly functional options, NECO had 42% of its options while WAEC had 12.8%. NECO had 4.4% of its options not functioning while WAEC had none in this category. Whenever distracters are weak, the question is rendered easy and as many of such as are supplied to an item, the greater the chance of answering the question right by majority of the testees (Owolabi, 2009; Onuka 2009). Observation in this study implies that candidates were exposed to some items which could be easily answered.

A closer look at the differences in the functionality of options, though in different subjects, shows and improvement. Functionality of options used in the 2015 SSCE English Language as observed in this study differ from earlier findings by Olatunji (2007) that WAEC 4-

options in Economics had 82% functional options while NECO 5-options had 56.7% and Olutola, Owolabi and Daramola (2015) showing 95% functioning options in WAEC's 4-options in Biology and 93% in NECO's 5-options. If examination bodies standardize the test development and their item refinement processes across the subjects, then what obtains in English Language may be assumed for the other subjects examined by them. The combination of fairly and highly functioning options in the 2015 WAEC SSCE English Language was 100% while that of NECO was 95.6.

Hypotheses Testing

This hypothesis was tested:

H₀: Functionality of options used in 2015 SSCE English Language is independent of examination bodies.

H₁: Functionality of options used in 2015 SSCE English Language is associated with examination bodies.

Table 3: Test of Significant difference

| Levels | WAEC | NECO | Total |
|-------------------|-------------|-------------|-------|
| Not Functional | 0(8.59) | 22(0.59) | 22 |
| Fairly Functional | 41(12.55) | 210 (64.28) | 251 |
| Highly Functional | 279(186.11) | 268(178.78) | 547 |
| Total | 320 | 500 | 820 |

Cal. X² = 61.56

Calculated Chi Square = 61.56 which is higher than the Table Value = 5.99 at df = 2, the null hypothesis is rejected. The alternative hypothesis that functionality of options used in the 2015 SSCE English Language is associated with examination bodies therefore stands. Apparently, what seems to be source of the difference

observed relate to the fact that the WAEC supplied more highly functioning items than NECO in their 2015 SSCE English Language MCTs. Also, a fundamental difference is that while 4 options were supplied by WAEC, while 5-options were supplied by NECO. In addition, there were 80 items in the WAEC paper while NECO used 100

items. Reasoning along with researchers like Ramons and Stern (1993), supplying 4 or 5 options which make guessing difficult are also more demanding with respect to the time, intellect and skills of the test developer. However, it should be noted that quality of numbers of option rather than quantity could guide against guessing. This is not unconnected with the submission of Olatunji and Owolabi (2009) that the fewer the options the better the item difficulty and discrimination indices. Also, 4-options item has significant gains such as economy in terms of avoidance of time and money wastage both in test construction and administration, better content sampling and maximum item effectiveness.

Conclusion

This study had shown that the quality of options supplied by both WAEC and NECO to SSCE English Language items is generally high. This probe into what goes on behind the scene with respect to test development has pointed in the direction of acceptable validity and reliability of the standardized tests conducted by the sampled examination bodies. None of the options supplied to the WAEC SSCE English Language MCT items was not functional. It was further established that functionality of options is not independent of examination bodies. The examination bodies tend to therefore exhibit peculiar levels of functionality of options used in their English Language MCTs.

Recommendations

The following recommendations were made:
The examination bodies should strengthen the test development processes to maximise the supply and use of functional options to their MCTs in English Language and this should be extended to all other subjects.
The use of fewer and more functional options should be considered especially by NECO in the development of MCTs in English Language.

In order to ensure comparable standards for the examination bodies in Nigeria, there is need to have a

regulatory body to look into the psychometric property focused in this study as well as others not included.

There is need to expand the scope of this research by translating it from a pilot study to a countrywide survey. The countrywide survey should cover all the subjects examined by the bodies with a confirmatory analysis of other relevant psychometric properties of these tests. All the examination bodies in Nigeria should also be covered. These include WAEC, NECO, NABTEB, and JAMB.

This study is an ex post facto analysis. Whereas this has its value, it is in the best interest of the examination bodies, candidates and the entire country to determine these qualities before the items are administered on candidates.

References

- Abdulghani, H. M.; Ahmad, F.; Ponnamperna, G. G.; Khalil, M. S.; & Aldrees, A. (2014). The relationship between non-functioning distractors and item difficulty of multiple choice questions: A descriptive analysis. *Journal of Health Specialties*, 2 (4), 148-151.
- Alonge, M. F. (2004). *Measurement and Evaluation in Education and Psychology*. Ado Ekiti: Adedayo Printing Nigeria Limited.
- Brame, C., (2013) Writing good multiple-choice test questions. Retrieved [May 30, 2017] from <https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/>
- Grier, J. B. (1975). The Number of alternatives for optimum test reliability. *Journal of Education and Measurement*.

- Haladyna, T. M., Downing, S. M. (1989). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2 (1): 37-50
- Kolawole, E. B. (2007). A comparative analysis of psychometric properties of two Nigerian examining bodies for Senior Secondary Schools mathematics. *Research Journal of Applied Sciences*, 2 (8), 913-915.
- Lawal, A. (2001). Evaluation of Students' Learning Outcomes I: Types and Uses of Tests. In I. O. Abimbola (Ed.) *Fundamental Principles and Practice of Instruction*. Ilorin: Belodan (Nig) Enterprises & Tunde Babs Printers.
- Olatunji, D. S. (2007). Effects of Number of Options on the psychometric properties of Senior School Certificate Multiple Choice papers in Economics. *Unpublished M. Ed. Dissertation*, University of Ilorin, Ilorin
- Olatunji, D. S. & Owolabi, H. O. (2009). Difficulty and discrimination of Senior School Certificate Economics objective test items with various option formats. *Ilorin Journal of Education*, 28, 50-64.
- Olutola, A. T., Owolabi, H. O. & Daramola, D. S. (2015). Analysis of functioning and non-functioning decoys in Multiple-choice Biology items. *FUDMA Journal of Science and Educational Research*. 1,
- Onuka, A. (2009). Developing objective tests. Workshop papers on Multiple Choice Test Item writing procedures for academic staff, University of Ilorin, Ilorin, Monday 4th 2009
- Osunde, A. U. (2009). *Essay and Multiple Choice Tests: Bridging the Gap*. Workshop Papers on Multiple Choice Test Item Writing Procedures for Academic Staff, University of Ilorin, Ilorin, on Monday 4th 2009. Pp. 14- 24.
- Owolabi, H. O. (2009). Characteristics of Multiple Choice Items. Workshop papers on Multiple Choice Test Item writing procedures for academic staff, University of Ilorin, Ilorin, Monday 4th 2009
- Ramons, R. L. & Stern, J. (1993). Item behaviour associated with changes in the number of alternatives in multiple choice item. *Journal of Educational Measurement*, 10.
- Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement Issues and Practices*, 24 (2): 3-13.
- Rodriguez, M.C., Kettler, R.J., & Elliott, S.N. (2014). Distractor functioning in modified items for test accessibility. *Sage Open*, 4(4), 1-10.
- Steven, M. D. (2000). Reliability: On the Reproducibility of Assessment Dafa. *Medical Education*. 38 (9), 1006- 1012.
- Tarrant, M., Ware, J. and Mohammed, A. M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A Descriptive Analysis. *BMC Medical Education*. DOI: 10.1186/1472- 6920-9-40

Baseline Survey Report on Aptitude Testing

Dr. Saverio Pido
Ruth Ndagire Ssenyonjo
Betty Nalukenge Habaasa
Dr. Peter Kagaba

Abstract

UNEB is in the process of setting up aptitude test batteries to be used for purposes of recruitment, enrollment, promotion, identification and appraisal. A baseline survey was therefore carried out to seek public opinion about the use of aptitude tests in organizations, identify variables required by various institutions, and inform UNEB about the need to develop aptitude tests. A cross section survey design was adopted and three districts were purposively sampled from each of the four traditional regions of Uganda. The institutions/organizations in these districts were stratified as private, Government and parastatal organizations. Two hundred and seven (207) human resource managers from institutions in the sampled districts participated in the study. Questionnaires were used as tools of data collection. Pearson Chi-Square and Fishers exact statistics were used to establish respondents' characteristics and some specific variables. Findings revealed: that more than 50% of the organizations that participated in the study had ever used aptitude tests; significant variations in willingness to use UNEB developed aptitude tests; mostly government based organizations were willing to use the UNEB developed aptitude tests closely followed by Private ones and their interest did not vary by management level, type of institution and terms of employment. The specific skills to be incorporated in the aptitude tests as mentioned by respondents include: problem solving, critical thinking, logical reasoning and interpersonal relationship skills. It was recommended that the set aptitude tests be valid reliable, and relevant to the job or any purpose for which they are set. It was also recommended that professionals be employed to develop the aptitude tests.

Introduction

UNEB is in the process of setting up a consultancy firm. One of the major functions of the firm is to develop aptitude test batteries that could be used by interested parties for purposes of recruitment/enrollment, promotions, identification of staff training needs and appraisals etc.

According to Graham and Lilly (1984), an aptitude test is a systematic means of testing job candidates' abilities to perform specific tasks and react to a range of different situations. Aptitude tests can be used as a measure of the ability of a person and give a complete picture of the person's mind. The tests mainly consist of multiple choice questions and are held as proper examinations. UNEB has an edge in administering public examinations;

this experience can therefore be extended to aptitude tests.

Bennet, Seashore and Wesman (1996) have specified the major purpose of aptitude tests as:

- Placement purpose: Used to evaluate a specific skill or some particular trait in a person as required by the job.
- Choosing a career: A person will get to know what he/she is good at because the results of the tests will inform on the skills and abilities of a person.
- Academic improvement: If a student takes an aptitude test, it will directly or indirectly test his/her strong and weak subjects.

The scholars further emphasized the need to include aptitude tests in education. As a major assessment body

therefore, it is necessary that UNEB incorporates aptitude tests in its assessments.

A baseline survey was therefore carried out to seek public views and identify specific areas the aptitude tests should capture with a view of empowering UNEB to develop aptitude tests for interested parties.

Objectives

The specific objectives of the survey were to;

- i) obtain opinions from the public about the use of aptitude tests in their organizations.
- ii) identify specific competencies to be elicited by aptitude tests in various institutions.
- iii) inform UNEB about the need of developing aptitude tests.

Methodology

Basing on the four major regions of Uganda (Central, East, West and North), at least three districts (Appendix

1) with large towns expected to have many organizations were purposively selected in each region. Institutions/organizations were stratified as: Private, Government and Parastatal organizations. In each district 10% of the organizations in each category were randomly sampled. A questionnaire was administered to Human resource managers in the sampled organisations.

Frequency tables and graphs were used to summarize the data. Relationships between characteristics of respondents and some specific variables were established using Pearson chi-square and fishers exact p-value where appropriate. The characteristics of the respondents and the findings of the study are presented in Table 1.

Background characteristics of respondents

The background characteristics of respondents that participated in the survey included: gender, job status, type of institution and terms of employment. Table 1 presents a distribution of these characteristics.

Table 1:

Characteristics of respondents

| Characteristics | Frequency(n = 207) | Percentage (%) |
|----------------------------|--------------------|----------------|
| Gender | | |
| Male | 142 | 68.6 |
| Female | 65 | 31.4 |
| Management Level | | |
| Top Manager | 76 | 36.71 |
| Middle Manager | 113 | 54.59 |
| Lower Manager | 16 | 7.73 |
| Other | 2 | 0.97 |
| Type of Institution | | |
| Private | 128 | 61.84 |
| Government | 43 | 20.77 |
| Parastatal | 3 | 1.45 |
| NGO | 29 | 14.01 |
| Other | 4 | 1.93 |
| Terms of employment | | |
| Permanent | 107 | 52.45 |
| Contract | 97 | 47.55 |

Table 1 shows the summarized background information of the respondents. More males (68.6%) than females (31.4%) participated in the study and slightly over half of the participants were employed on permanent terms (52.5%) while those on contract were 47.5%. Majority of respondents (61.8%) were working in private institutions while the rest were mainly employed in government institutions (20.8%) and NGOs (14.0%). With regard to the job status, the highest proportion was middle level managers (54.6%), followed by Top managers (36.7%) while the rest were lower managers (7.7%) and other levels of management (0.97%).

Table 2
Awareness of aptitude tests by respondents

| Characteristics | Frequency | Response (%) | | Statistic Value (Chi-square) | p-value |
|----------------------------|------------|--------------|--------------|---------------------------------|---------|
| | | Yes | No | | |
| Gender | | | | | |
| Male | 142 | 83.10 | 16.90 | 0.63 | 0.424 |
| Female | 65 | 78.46 | 21.54 | | |
| Job Level | | | | | |
| Top Manager | 76 | 81.58 | 18.42 | 0.800 | |
| Middle Manager | 113 | 82.30 | 17.70 | | |
| Lower Manager | 16 | 75.00 | 25.00 | | |
| Other | 2 | 100.00 | 0.00 | | |
| Type of Institution | | | | | |
| Private | 128 | 79.69 | 20.31 | 0.318 | |
| Government | 43 | 90.70 | 9.30 | | |
| Parastatals | 3 | 66.67 | 33.33 | | |
| NGO | 29 | 79.31 | 20.69 | | |
| Other | 4 | 75.00 | 25.00 | | |
| Terms of Employment | | | | | |
| Permanent | 107 | 86.92 | 13.08 | 4.56 | 0.033 |
| Contract | 97 | 75.26 | 24.74 | | |
| Total | 207 | 81.64 | 18.36 | | |

Findings

The findings are reported basing on three aspects: organization opinions on the use of aptitude tests; specific competences required by various institutions to be included in the aptitude tests and necessary information given to UNEB about developing aptitude tests.

Objective one; Organizations’ opinion on the use of Aptitude Testing

Respondents were asked about their awareness of aptitude testing. Table 2 presents responses (Yes/No) made by respondents against their characteristics.

In the results according to Table 2, majority responded positively regarding awareness of aptitude tests (81.6%). Further analysis of awareness of aptitude tests revealed significant variations by respondents' terms of employment ($p < 0.05$). In particular, about twice the proportion of employees on contract (24.7) were

unfamiliar with aptitude tests compared with employees on permanent basis (13.1%).

The participants were further asked whether they would be interested in the use of aptitude tests in their organisations. Their responses are summarized in Table 3.

Table 3:
Views of respondents on the use of aptitude tests in their institution

| Characteristics | Frequency | Response (%) | | P-value |
|----------------------------|------------|--------------|--------------|---------|
| | | Yes | No | |
| Management Level | | | | |
| Top Manager | 76 | 47.37 | 52.63 | 0.159 |
| Middle Manager | 113 | 56.64 | 43.36 | |
| Lower Manager | 16 | 37.50 | 62.50 | |
| Other | 2 | 0.00 | 100.00 | |
| Type of Institution | | | | |
| Private | 128 | 46.88 | 53.13 | 0.484 |
| Government | 43 | 60.47 | 39.53 | |
| Parastatal | 3 | 66.67 | 33.33 | |
| NGO | 29 | 51.72 | 48.28 | |
| Other | 4 | 75.00 | 25.00 | |
| Terms of Employment | | | | |
| Permanent | 107 | 57.01 | 42.99 | 0.096 |
| Contract | 97 | 45.36 | 54.64 | |
| Total | 207 | 51.21 | 48.79 | |

Regarding use of aptitude tests in the institutions or organization, slightly more than a half (51.2%) responded in the affirmative as shown in Table 3. No significant variations in the responses were noted among management level of respondents, type of institution and terms of employment ($p > 0.05$). In other words, interest

in the use of the aptitude tests did not vary by background characteristics of the respondents.

In addition, the participants were requested to provide information on how often they use aptitude tests in their organisation and Table 4 depicts results regarding the frequency of usage of aptitude tests.

Table 4:
Frequency of usage of aptitude tests by type of institution

| Type of Institution | Frequency (n=207) | Response (%) | | |
|---------------------|-------------------|--------------|--------------|--------------|
| | | Very often | Sometimes | Never |
| Private | 128 | 11.72 | 49.22 | 39.06 |
| Government | 43 | 18.60 | 51.16 | 30.23 |
| Parastatal | 3 | 66.67 | 0.00 | 33.33 |
| NGO | 29 | 10.34 | 41.38 | 48.28 |
| Other | 4 | 0.00 | 75.00 | 25.00 |
| Average | | 13.53 | 48.31 | 38.16 |

With regard to the frequency of administering the aptitude test, the results in Table 4 show that the highest proportion (48.3%) did so sometimes, followed by those who never administered the test (38.2%) while the rest administered the test very often (13.5%).

In summary, 69.8% of the government institutions use aptitude tests, followed by parastatals (66.7%), private institutions (60.9%) and finally NGO's (51.7%).

The respondents were then asked about their willingness to hire UNEB to administer aptitude tests in their organizations. The summarized results are presented in Table 5.

Table 5
Willingness to hire UNEB to administer aptitude tests

| Characteristics | Frequency n =207 | Response (%) | | p-value |
|----------------------------|---------------------|--------------|--------------|---------|
| | | Yes | No | |
| Type of Institution | | | | |
| Private | 128 | 59.38 | 40.63 | 0.019 |
| Government | 43 | 79.07 | 20.93 | |
| Parastatal | 3 | 0.00 | 100.00 | |
| NGO | 29 | 68.97 | 31.03 | |
| Other | 4 | 75.00 | 25.00 | |
| Average | | 64.25 | 35.75 | |

In the results according to Table 5, majority responded positively (64.3%) regarding willingness to hire UNEB administer aptitude tests in their institutions. Further analysis by background characteristics revealed significant variations in willingness by type of institution ($p < 0.05$). The highest proportion of those willing to hire UNEB were from government institutions (79.1%) followed by NGOs (69%) and private institutions (59.4%).

Respondents were then asked to identify the purpose for which the aptitude test would serve in their institutions or organizations. The results are shown in Figure 1.

Figure 1: Purpose for which aptitude tests would serve in the institution

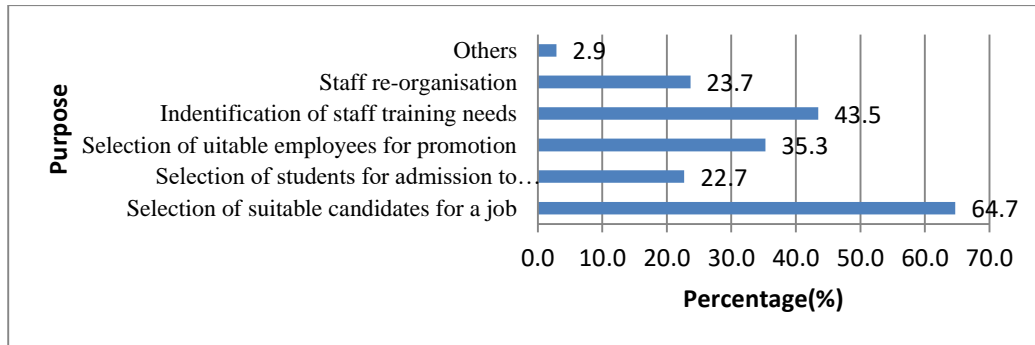


Figure 1 shows that the predominant purposes for aptitude tests include: selection of suitable candidate for a job (64.7%), identification of staff training needs (43.5%) and selection of suitable employees for promotion (35.3%).

Furthermore, assessment regarding the degree of importance of the tests was made. Respondents were asked about their opinion on the degree of importance of aptitude testing in their institutions or organizations. Responses are shown in Table 6.

Table 6

Degree of importance of aptitude testing by respondents' characteristics

| Characteristics | Frequency | Response (%) | | | p-value |
|----------------------------|-----------|----------------|--------------|---------------|---------|
| | | Very Important | Important | Not Important | |
| Management Level | | | | | |
| Top Manager | 76 | 46.05 | 43.42 | 10.53 | 0.916 |
| Middle Manager | 113 | 45.13 | 45.13 | 9.73 | |
| Lower Manager | 16 | 43.75 | 56.25 | 0.00 | |
| Other | 2 | 50.00 | 50.00 | 0.00 | |
| Type of Institution | | | | | |
| Private | 128 | 42.97 | 47.66 | 9.38 | 0.364 |
| Government | 43 | 60.47 | 34.88 | 4.65 | |
| Parastatal | 3 | 33.33 | 66.67 | 0.00 | |
| NGO | 29 | 37.93 | 44.83 | 17.24 | |
| Other | 4 | 25.00 | 75.00 | 0.00 | |
| Average | | 45.41 | 45.41 | 9.18 | |

According to Table 6, the majority responded that the tests were important (90.82%). In particular, 45.4% responded that the tests were very important. A similar proportion responded that the test was important in any institution. However, no significant variations in opinion

were noted with respect to management level and type of institution.

With regard to whether aptitude tests are high quality tools that may be used for selection purposes in their

institutions, majority of respondents' agreed (77.3%) as shown in Table 7.

Table 7

Opinion on aptitude tests

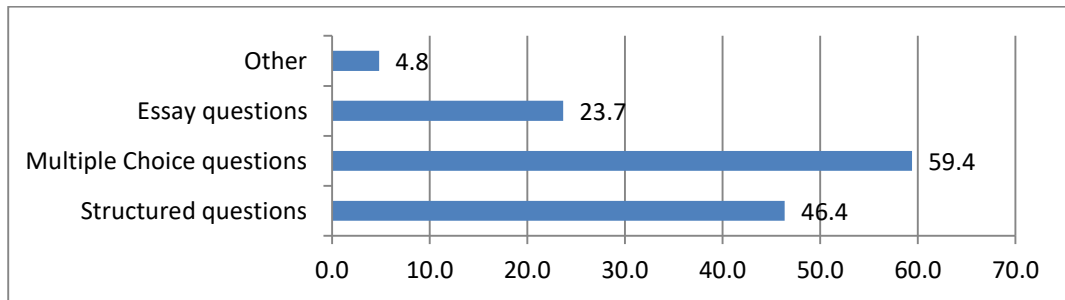
| Aspects on aptitude tests | | Frequency(n = 207) | Percentage (%) |
|--|-----|--------------------|----------------|
| Whether aptitude tests are high quality tools | Yes | 160 | 77.3 |
| | No | 47 | 22.7 |
| Whether you would recommend use of aptitude tests | Yes | 184 | 88.9 |
| | No | 23 | 11.1 |

Similarly, when the participants were asked whether they were willing to recommend the use of the aptitude tests to other institutions as shown in Table 7, majority (88.9%) responded that they would recommend the use of aptitude tests to other institutions.

Format of aptitude tests

Respondents were asked to give their preferred choice of aptitude test format. In their response as shown in Figure 2, the most preferred structure of the tests was multiple choice questions (59.4%) followed by structured questions (46.4%).

Figure 2: Preferred structure of aptitude tests (%)



Method of Administration of aptitude tests

Respondents were further asked about their opinion regarding the administration of the aptitude test. Majority (67.2%) responded that they would like a one-to-one approach of administering the tests; the rest mentioned group approach.

As regards whether or not the respondent would be willing to participate in the process of aptitude test setting, the vast majority (87.4%) said they were willing while only 12.6% were unwilling.

Objective two; Aptitude Test variables required by Institutions

The study further established from the participants the content areas to be covered and specific competences and or skills to be elicited by the aptitude tests.

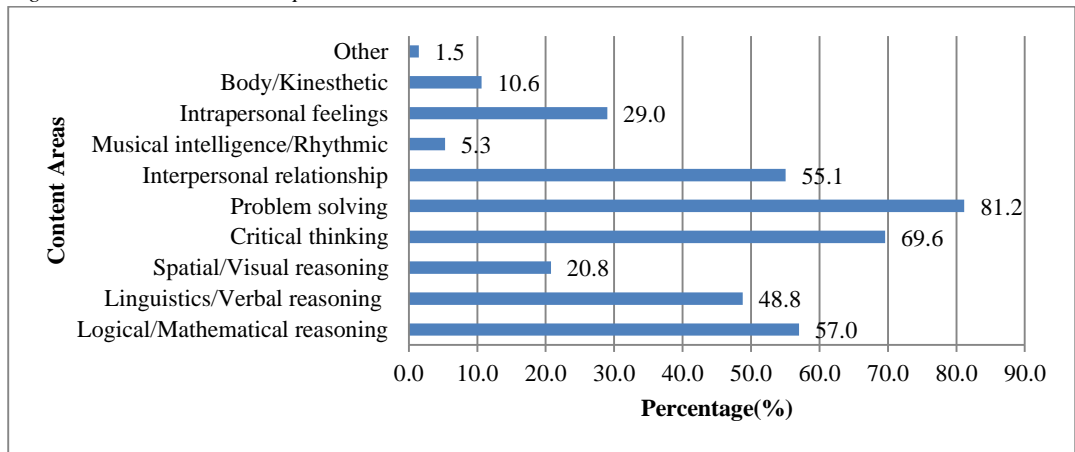
Specific skills in aptitude tests

Respondents were requested to mention the content and skills that they would prefer to be tested in their organizations. Figures 3 and 4 illustrate the content areas and specific skills that respondents would wish the aptitude tests to cover.

Content Areas

Content areas required to be assessed in the aptitude tests as identified by respondents are illustrated in Figure 3.

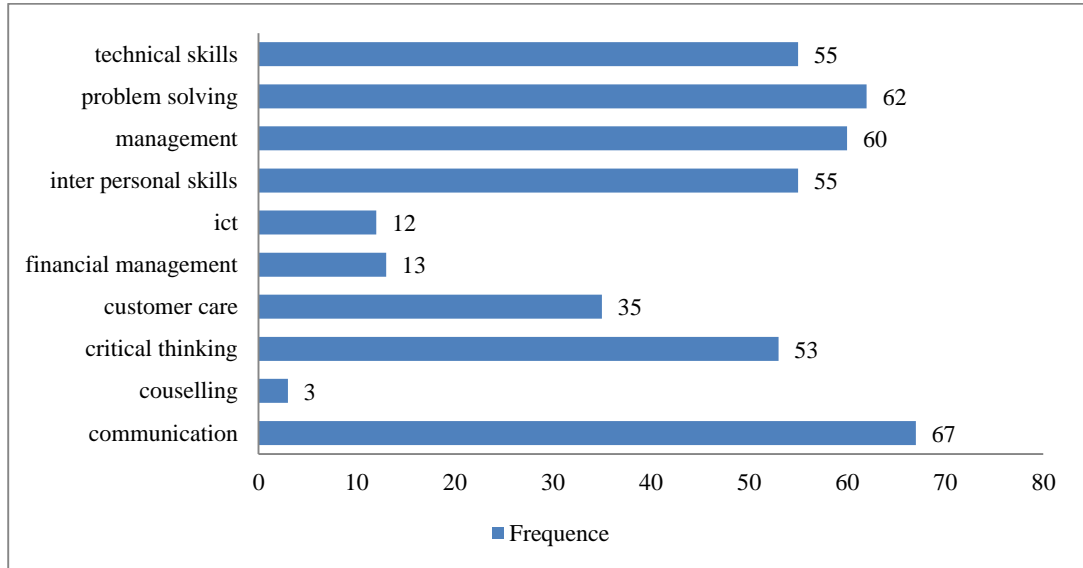
Figure 3: Content areas the aptitude tests would cover



From the results according to Figure 3, four major areas were noted namely problem solving (81.2%), critical thinking (69.6%), logical reasoning (57%) and interpersonal relationship (55.1%) as the main content areas to be covered in aptitude testing.

The competences and or skills required by respondents to be assessed in aptitude tests are illustrated in figure 4.

Figure 4: Responses on competences that aptitude tests should cover



According to Figure 4, the major skills that respondents would wish the aptitude tests to cover include: communication skills (67), problem solving skills (62), management skills (60), technical and interpersonal skills (55) as well as critical thinking skills (53).

Objective three; inform UNEB about the need of developing aptitude tests.

Using the questionnaire, the study sought information from the respondents that would help UNEB in the event that aptitude tests were taken on.

Information given to UNEB

Respondents gave the advantages and disadvantages of using aptitude tests and finally gave advice to UNEB about aptitude tests.

The advantages of using aptitude tests as given by the respondents are presented in Table 8.

Table 8:

Advantages of using aptitude tests

| S/N | Advantages of using aptitude tests | Frequency | % |
|-----|--|-----------|------|
| 1 | Capable applicants easily identified | 60 | 30.0 |
| 2 | Intelligent applicants easily identified | 59 | 28.5 |
| 3 | Easy to administer | 56 | 27.1 |
| 4 | Propel critical thinking | 54 | 26.1 |
| 5 | Avoid bias | 18 | 8.7 |
| 6 | Identifies organizational needs | 12 | 5.8 |

Respondents mentioned the following advantages in using aptitude tests: identifying capable and intelligent applicants (60), easy to administer (56) and encourages critical thinking.

On the other hand, the respondents were also requested to give the disadvantages of using aptitude tests and Table 9 summarizes their responses.

Table 9:

Disadvantages of using aptitude tests

| S/N | Disadvantages of using aptitude tests | Frequency | % |
|-----|---------------------------------------|-----------|------|
| 1 | May not be valid. | 65 | 31.4 |
| 2 | Costly. | 46 | 22.2 |
| 3 | Time consuming. | 44 | 21.3 |
| 4 | May not be reliable. | 33 | 15.9 |
| 5 | Questions may leak. | 30 | 14.5 |
| 6 | Language barrier. | 20 | 9.7 |
| 7 | May be biased. | 7 | 3.4 |
| 8 | Lack of awareness of AT | 5 | 2.4 |

With reference to Table 9, respondents were concerned that the aptitude tests may not be valid (31.4%), be expensive (22.2%), be time consuming (21.3%), not be reliable (15.9%) and could be leaked (14.5%).

However, the respondents gave advice to UNEB on how to overcome the above problems. The suggestions are presented in Table 10.

Table 10:

Advice to UNEB on administering aptitude tests

| S/N | Advice on administering aptitude tests | Frequency | % |
|-----|--|-----------|------|
| 1 | Aptitude tests should be relevant to the job | 50 | 24.2 |
| 2 | Sensitize masses on aptitude tests | 34 | 16.4 |
| 3 | Use professionals | 31 | 15.0 |
| 4 | Charge less compared to the rest | 12 | 5.8 |
| 5 | Develop strong question bank | 8 | 3.9 |
| 6 | Be transparent and honest | 7 | 3.4 |
| 7 | Carry out needs assessment | 6 | 3.0 |
| 8 | Set structured questions | 6 | 3.0 |

The advice of respondents as listed in Table 10 included: aptitude tests being relevant to the job (24.2%); UNEB sensitizing masses on the use of aptitude tests (16.4%) and engaging professionals in the development of aptitude tests (15.0%).

Conclusion

The baseline survey conducted by UNEB involved 207 organizations. Majority of the organisations that participated were private (61.84%), followed by government organizations (20.77%). More than half of the respondents (54.5%) were middle managers and

47.37% were top managers. Majority of respondents (81.6%) were aware of aptitude tests and these were mainly permanent employees compared to those on contract. It was noted that more than half (51.2%) of participants said they were interest in the use aptitude tests in their institutions.

The specific competencies identified from the study to be included in the aptitude tests include: problem solving skills, communication skills, management skills, critical thinking skills, logical/mathematical reasoning competences, interpersonal relationships and technical skills. It was also noted that the most preferred structure of tests were multiple choice tests followed by structured questions.

More than half (64.3%) of the institutions were willing to hire UNEB to administer aptitude tests to them. These were mainly from government institutions followed by non-government organizations. On the other hand parastatal bodies were not willing to have aptitude tests administered to them.

The major purposes that aptitude tests would serve include; selection of suitable candidates for the job, identification of staff training needs and selection of suitable employees for promotion.

Advice given to UNEB in the process of administering aptitude tests involves: aptitude tests being relevant to the job; sensitizing masses on aptitude tests and using professionals in the development of aptitude tests.

Recommendations

Basing on the findings the following recommendations are made:

UNEB should;

- Implement the development and administering of aptitude tests in Uganda because the public especially government organizations have trust in UNEB.
- Sensitize masses about the need to use aptitude tests.
- Develop internal capacity in the development and administration of aptitude testing.

References

- Bennett, G.K., Seashore, H.G.& Wesman, A.G. (1996), *Differential Aptitude Tests. Psychological cooperation, San Antonio, TX. US:*
- Ghiselli, E. (1973). *The validity of aptitude tests in personal selection*
- Graham, J.R. & Lilly R.S. (1984). *Psychological Testice, Prentice Hall.*

SUB THEME G: STANDARD SETTING PROCEDURES IN EDUCATIONAL ASSESSMENT

Standard Setting Procedures: A Challenge for Modern Educational Assessment

David Kani Olema

Busitema University

Abstract

The central role of assessment in education is of greater social concern now than ever before. The stakes are higher in this digital era: the current emphasis on the so called 21st century skills, scarcity and stiff competitiveness of jobs, “paper race,” and the like. This paper intended to provoke further discourse on the following questions: a) What is it that is assessed? b) What level or extent of assessment is true and acceptable? c) How fair are the assessments? In exploring answers to these questions, the paper has had review of various literature and, as much as possible, examples have been drawn from experiences in Uganda. The paper has concluded that, the question of what is assessed is as complex in answering as the answer to the extent of truth and acceptability of assessment. The issue of fairness is more disturbing than is focused on by any assessment system or body. It is therefore concluded that a more encompassing assessment, especially assessment for learning, should be employed rather than relying only on assessment of learning. Further, deliberate reform in the form of assessment and assessment systems are suggested to minimize the most feared “halo effect” of assessment for learning, hence pinning what is assessed, and improving on fairness and acceptability of assessment.

Key Words: *Assessment, assessment standard, Assessment for Learning, Assessment of learning*

Introduction

Without assessment, the process of teaching and learning, hence education is incomplete (William, 2013). Assessment plays a central role in teaching and learning, especially the determination of the learner’s fate in terms of progression in teaching and learning. That role notwithstanding, assessment has the power to “kill” and “create”: the determination of progress or retention, and labelling the learner has everything for their future life in school and after school. The young and old alike are anxious to approach being assessed or undertaking any form of assessment or examination, and to receive the outcome of the exercise. Assessment is fair and unfair at the same time. Unfortunately, it cannot be done away with; “a necessary evil indeed!” (Valencia, 2002)!

The central role assessment plays in education is reflected by it being used as a means of accounting to parents, students, teachers, the government, and indeed a myriad of other stakeholders; that is the role of assessment goes beyond awarding students marks or certificates (Hambleton, 2016; Liljedahl, 2010; Surgenor, 2010). Assessment also places a value tag on the assessed, the assessor, and the system(s) of and in which the assessment takes place. Because the stakes are always very high in any assessment, consequently many questions are often asked, including, but not limited to: a) What is it that is assessed? b) What level or extent of this assessment is acceptable? c) How fair are assessments?

The good and bad impacts of assessment on the individual learner, the institutions that manage the assessment, and humanity calls for a fair and less stressing assessment

(The Gordon Commission, 2012). Such “fair” and “less stressing or less anxiety-evoking” assessments should meet a standard. Thus, the need for a standard assessment (Hambleton, 2016). Yet, the standardization of assessment creates more unfairness and anxiety! Yet again, “fairness” in anything cannot be achieved without standard or some agreed upon rules of the game.

What are standard assessments?

Before talking about standard assessment, it is important to give the meaning of the noun, assessment. The word assessment has its roots in the Latin word *assidere*, literally translated to “sit by” (citation). This literal meaning ideally calls for a person assessing another to *sit by* the one being assessed, and he or she carries out the process. We shall look at this *sitting by* a little later, in relation to the forms of educational assessment commonly carried out. However, assessment has on many occasions been used to mean to (a) evaluate, (b) judge, (c) gauge, (d) rate, (e) estimate, (f) appraise, (g) form an opinion of, (h) check out, (i) form an impression of, (j) make up one’s mind about, (k) get the measure of, (l) determine, (m) value or put value on, (n) cost, and so on (Stevenson, Ed., 2015). Following such an expounded meaning of assessment, a standard educational assessment is as complex as it is convoluted in the meaning of the word. However, in this paper a standard assessment simply refers to one that meets universally acceptable criteria or procedures, and the agreed level of quality or attainment. Though this definition seems simple, it is intricate in its operationalization as it raises questions such as (a) Who sets the assessment standard? (b) How is assessment standard set? Before we provide any solutions to these questions, we need to know why we need standard assessments.

Why standard Assessments?

Aware of the numerous disadvantages of classroom assessment in general and particularly the standard

assessment (Colombia University, 2013), and the different forms of assessment; yet for a common anticipated market, it is important that an agreed level of quality is used. Thus, the need for standard assessment. A range of high stakes decisions are made about learners based on educational assessment outcomes, making assessments the single most important determinant of learners’ future. According to Green (1995), there are four major reasons why standards are needed: (a) certification of minimum competence and other professional certification, (b) prediction of performance at the next level of education (promotion) or employment (aptitude), (c) description of a measurement scale (ranges of performance, etc).

Assessment results may be used to classify learners, according to their performance, into various categories ranging from the positive *pass* to the often negative, *fail*. Such classification places the test takers into several achievement levels (Hambleton, Jaeger, Plake, & Mills, 2016). For example, students who take the Uganda Certificate of Education (UCE) or Ordinary Level (*O’level*), are classified as 1st Grade (has passed exceptionally well), 2nd Grade (has passed well), 3rd Grade (has passed), 4th Grade (has a satisfactory pass), and any grade beyond the fourth (has not passed) to establish and communicate the achievement level of the educational goals at the *O’level*. For post-secondary institutions in Uganda, certificate course examination takers are categorized as having achieved Class I (*Distinction*), Class II (*Credit*), and Class III (*Pass*; NCHE, 2015). In other cases, test takers are categorized by their academic performance in terms of the level of their proficiency, as being *Advanced*, *Proficient*, *Basic*, and *Below Basic*, and indeed, in many other categorizations that exist out there. Such categorizations and or labelling are not only used by educationists for academic purposes, but also by the job market and society for placement. These categorizations, especially the lower end ones, are more often open, unintended doors of

negative psychological consequences for the test takers, and abuse of the results by the wider society.

Therefore, the purpose of having a standard assessment, especially of the large-scale such as the regional or national examinations, should as much as possible be **valid, reliable, and fair** including being accessible to be acceptable (Hambleton, Jaeger, Plake, & Mills, 2016). The process of having standard assessments requires, first agreeing *on the standards*, and *how they are set*. This paper does not focus on the **standard assessment** *per se*, but rather **setting the standards**. The distinction between the two is that, **standard assessment** is an assessment which meets a prior agreed standard or level of quality, while **setting the standard** is the process of arriving at the standard assessment. The latter is the product of the former.

Setting Assessment Standards

It is impossible to set assessment standards without knowing the standards that we need. According to education assessment experts, there are two forms of standards that are considered in educational assessment: the *Content Standard* and the *Performance Standard*. The content standard refers to the curriculum or course content that students are supposed to learn. In setting the content standard, the question at hand is, “what are students supposed to know or in effect what are the learners expected to be able to do?” Whereas, the performance standard, on the other hand, refers to the level of performance expected of the students. In setting the performance standard, the question is, “what level of performance is expected of a student to be graded; 1st Grade, 2nd Grade, (etc.). The two Educational Standards combined help to arrive at the achievement levels that indicate what a student who has passed through, for instance, the O’level in Uganda, should know and be able to do, and level of his or her performance to warrant award of a given class or grade of certificate. Therefore, *Setting Assessment Standards* ideally is arriving at what

the learner at a certain level should know and be able to do.

Agreeing on the standards for an assessment is the most important step in all efforts of delivering a fair, valid, and reliable assessment. According to Bejar (2008), standard setting is “the methodology used to define *levels* of achievement or proficiency and the *cutscores* corresponding to those levels” (p. 1). A cutscore or cutoff mark is the score or mark that sets the pass mark, or grades test takers at one level of achievement or the next, or a mark that sets the line between pass and fail. The result of this mark or cutscore is essentially classification of learners. Classification takes into consideration many factors which are at the core of assessment. These factors include issues of fairness, equity, and validity which constitute the heart of authenticity of assessment outcomes. Yet the issues of fairness and equity may be beyond the assessor’s powers, especially in the case of varying geographical, socio-economic, cultural, historical, physical, psychological, structural, and environmental contexts in which the test takers learn or prepare (Dodge, et al. 2013; Friend & Bursuck, 2006; Winthrop & Ferris, 2010). Despite the emphasis put on the issues of fairness associated with setting performance standards, in the actual demonstrated performance by test takers, wide variations exist among the candidates. So, categorization according to performance requires a common scale, and deciding who has and has not achieved the minimum level of the expected learning outcomes requires setting a standard. Therefore, setting a standard is inevitable. In designing assessment, test-developers seek to establish whether candidates have completed the learning tasks specified in the curriculum and the learning objectives. This achievement of the objectives can only be called forth and determined by the candidates achieving a certain minimum level of competence.

Therefore, standard setting procedures for the levels of competence is paramount. To set the standards, one must

consider broad aspects of educational objectives, policies of the education, and structural and systematic issues that enable or inhibit the achievement of the tasks. Besides, many stakeholders that include (a) Governments, (b) Students, (c) Parents, (d) Job Market and Society (social skills) need the standards. It is therefore the expectation of these entities and more that the educationists also consider, in order to arrive at the assessment standards.

The standard setting methods

Given the many stakeholders expecting certain standards of products from the education systems, setting standards is a complex venture. Consequently, there is no one recommended approach to setting standards. However, there are general frameworks, principles, key concepts and practical considerations of setting standards. These general considerations provide possible guidelines for making informed decisions on the frameworks that best suit the examination bodies. However, in setting assessment standards, the focus is more on setting performance standards rather than setting content standards. In Uganda, the latter is the responsibility of Government through National Curriculum Development Center (NCDC) for primary and secondary education, and National Council for Higher Education for the tertiary institutions. This process of setting content standards is essentially done before setting the performance standard, because the latter is a process of establishing the levels of the earlier for the different purposes. The two are not mutually exclusive! There are two major standard setting methods, which are dependent on the expected assessment outcomes. For instance, when the expected assessment outcomes are relative, then the *norm-referenced* approach is used. Whereas, when absolute or independently derived, then the *criterion referenced* approach for standard setting is used.

Norm referenced approach to setting the standard

The Norm Referenced approach to the expected assessment outcome or interpretation of results to setting standards is where the performance of a candidate is based on the norm (the larger group) that sat the exam. Therefore, in an examination whose standard is set using the norm referenced approach, a relative standard can be set using various statistics, such as at the mean performance of the candidates, percentile ranks, and median points. This makes the norm referenced standards unstable, varying with the ability of the group that sat the examinations, and thus be fixed per examination or year of examination as well, to achieve a normal distribution (i.e., balance the number of “failures” and those who have passed at the different levels). That is, the standard is set after the examination has been done and marks are computed. For Example: When in 2016 the students in a National or Regional examinations generally have high scores in mathematics, the pass mark may be set at 50%, and if candidates of the same education level take the same in 2017 and the general performance in mathematics is poor, the pass mark may be set at 25%.

Criterion referenced approach to setting the standard

On the other hand, the Criterion Referenced approach to the expected assessment outcome or interpretation of results is where the candidates’ results are linked to the content and the competence level under consideration. In other words, the standard is set in relation to “how much” of the expected content has been mastered by the candidate and what level of mastery prior to the examination. For example, according to National Council for Higher Education (NCHE, 2015) in Uganda, for degree and diploma courses, an “A” student should score 80% and above and a “C” student between 60% and 64% in a subject or course (see Tables 1- 3):

Table 1: Prior assigned grades

| Marks | 80 -100 | 75 - 79 | 70 - 74 | 65 - 69 | 60 - 64 | 55 - 59 | 50 - 54 | 0 – 49 |
|--------------|---------|---------|---------|---------|---------|---------|---------|--------|
| Letter Grade | A | B+ | B | C+ | C | D+ | D | F |
| Grade Point | 5 | 4.5 | 4.0 | 3.5 | 3.0 | 2.5 | 2.0 | 0 |

Table 2: Classification of awards for Degrees

| Class | Cumulated Grade Point Average (CGPA) |
|-------------------------------|--------------------------------------|
| First Class | 4.40 – 5.00 |
| Second Class (Upper Division) | 3.60 - 4.39 |
| Second Class (Lower Division) | 2.80 – 3.59 |
| Third Class (Pass) | 2.00 – 2.79 |

Table 3: Classification of awards of Diplomas

| Class | Cumulated Grade Point Average (CGPA) |
|-----------------------|--------------------------------------|
| Class I (Distinction) | 4.40 – 5.00 |
| Class II (Credit) | 2.80 - 4.39 |
| Class III (Pass) | 2.00 – 2.79 |

Other Approaches to setting Standards

There are various other approaches to setting assessment standards including Test-centered Models, Group centered Models, Hofstee Method, and so on.

Test-Centered Models

This approach or model only applies to cases where the scores of the borderline group are clustered. Both educational testing and performance assessments are considered in the model. The judges in this case set standards by reviewing test items against performance of the candidates in these items, basing on “just adequate” level of performance.

Group centered Model

Unlike the test centered model, the judges in the group centred approach sort the examinees into two groups: a) the competent, and b) the non-competent on the basis of the examinees characteristics in relation to the tasks other than test scores. Before their knowledge of the scores, the

judges will consider the tasks that the examinees did and on this basis, determine the two categories. Take an arbitrary example, examinees who attempted Tasks 4, 7, and 9 are competent and those who attempted Tasks 5, 6, 8 are incompetent. Thereafter, the tasks are scored and the scores of the two groups are plotted and the point of intersection is considered the pass mark.

The Hofstee Method

This method is a compromise between the *relative* and *absolute* standards. It is a method that considers the judgement of subject experts, familiar with examination methods, level of candidates and good problem-solvers. This method, in the case of Uganda, would comprise subject experts, seniority in dealing with examinations, principled personalities who are well informed in matters of education. The judges or panelists review the test materials, and they provide the following values (see Table 4):

Table 4: Hofstee’s method test values

| Description of Value | Value |
|---|-----------------------|
| Lowest acceptable percentage of failing examinee | Minimum failing rate |
| Highest acceptable percentage of failing examinee | Maximum failing rate |
| Lowest Score which allows a candidate to pass | Minimum passing point |
| Highest score required for a candidate to pass | Maximum passing point |

Practical country/local experiences in the procedures of setting standards and grading the learners

Various countries have had different reforms in their assessment over the years to the current reforms to produce global citizens with 21st century skills. This can best be expressed in the call: “I am calling on our nation’s Governors and state education chiefs to develop standards and assessments that don’t simply measure whether students can fill in a bubble on a test, but whether they possess 21st-century skills like problem-solving and critical thinking, entrepreneurship and creativity” (President Barack Obama, March 2009, as quoted by Darling-Hammond, et al. 2013, p.1).

The USA for example emphasizes 45 states agreed on what they referred to as the Common Core State Standards (CCSS), which in their effort to meet the current job and life demands focus on a number of learning outcomes. These outcomes incorporate **deeper learning**, which entails students’ ability to analyze, synthesize, compare, connect, critique, hypothesize, prove, and explain their ideas. These are the 21st century skills.

Challenges to setting Educational Standards

There are numerous challenges of Setting Standards for Educational Assessment. These challenges emanate from the current concerns of assessment which include, among others:

1. Being too much or too many assessments. Schools have too many formal assessments and start involving students too early in life. This is a typical concern of assessment in Uganda where even the pre-primary pupils are subjected to formal examinations, the primary and secondary schools

being reduced to studying examinations or how to pass examinations instead of craving to acquire knowledge and skills for life. The primary and secondary schools, out of the 12 weeks of a normal Ugandan educational term, have more than four weeks for examinations. As a result, schools extend pupils’ time of study from the 8:00 a.m. to 3:30 p.m., by starting as early as 6:00 a.m. to 5:00 or 6:00 p.m. without extending the break times (Allen, Elks, Outhred, & Varly, 2016).

2. Being too much cognitively oriented, neglecting the psychomotor and affective domains. Such examinations most likely produce people who are not well balanced in life and who are potentially problematic in society.

3. Being the driver of teaching and learning, that is, assessment of learning instead of teaching and learning driving assessment, that is, assessment for learning. An approach of integrating assessment into the learning environment would improve on competences of the learners.

4. Lacking diversity. Opolot-Okurut (2010), in a study conducted among primary and secondary schools in Uganda, reported that the assessment is heavily reliant on pencil and paper type of assessment.

5. Being top-down instead of bottom-up, or at best bottom-up and horizontal (considering the societal expectations or philosophies.

6. Lacking diversity of facilities and personnel, and hence preparation for assessment at national level is poor.

7. The overwhelming number of students per teacher is certainly a big challenge for quality assessment, and so on.

8. Subjecting students from various backgrounds; especially Socio-economic and other geographical disparities to the same examination. These external socio-

economic and geographical situations rather than the innate abilities end up determining the fate of the learners.

Conclusion

It is therefore concluded that due cognizance should be given to the regional and institutional disparities and diversities in facilities and personnel in schools by adopting a diverse and more encompassing assessment. Such an assessment would cater for the differences that are not of individual, but rather of environmental origin. In other words, assessment for learning and aptitude for future learning should be employed rather than depending more on assessment of learning because of the unfair and diverse teaching and learning experiences. Intensive quality control measures should be established in addition to training assessment specialists for this new approach. Various approaches for setting standards need to be followed to cater for the preferred assessment for leaning. As assessment for learning demands, various approaches to call forth the different skills, diverse approached also need to be appropriately followed in setting standards on case by case basis.

Recommendations

Deliberate reforms in assessment and assessment systems should be explored for inclusive development of the critical human resource needed for socio-economic development of peoples. Robust plans should be made to minimize the most feared halo effect of assessment for learning, hence pinning what is assessed, and improving on fairness and acceptability of assessment.

Reference

Allen, R., Elks, P., Outhred, R., & Varly, P. (2016). Uganda's Assessment System: A road-map for Enhancing Assessment in Education. *Health & Education Advance & Resource Team*, UKaid. Retrieved 20th July, 2016; from: <https://assets.publishing.service.gov.uk/media/585a8c7740f0b60e4a0000dc/Final-report->

Enhancing-Ugandas-Assessment-System-September-2016.pdf.

Colombia University (2013). Pros and Cons of Standardised Testing. *Office of Work/Life School and Child Care Search Service*. Retrieved 25th May, 2017 from: http://worklife.columbia.edu/files_worklife/public/Pros_and_Cons_of_Standardized_Testing_1.pdf

Darling-Hammond, et al. (2013). Criteria for High-Quality Assessment. *Stanford Center for Opportunity Policy in Education, Stanford University*. Retrieved 25th May, 2017 from: https://edpolicy.stanford.edu/sites/default/files/publications/criteria-higher-quality-assessment_2.pdf

Dodge, L. et al. (2013). *Guidelines: Intellectual Disabilities in Educational Settings*. Special Education in North Dakota Department of Public Instruction, Bismarck.

Friend, M., and Bursuck, W. (2006). *Including Students with Special Needs: A Practical Guide for Classroom Teachers, 4/E*. Allyn & Bacon, Boston (USA)

Ferris, E. and Winthrop, R. (2010). Education and Displacement: Assessing Conditions for Refugees and Internally Displaced Persons affected by Conflict. *Paper commissioned for the EFA Global Monitoring Report 2011, The hidden crisis: Armed conflict and education*. efareport@unesco.org.

Green, F. B. (1995). *Setting Performance Standards: Content, Goals, and Individual Differences*. The Second Annual William Angoff Memorial Lecture Series, November 6th 1995. Education Testing Service, Princeton, US.

- Hambleton, K. R., Jaeger, M. R., Plake, S. B., and Mills, C. (2016). *Setting Performance Standards on Complex Educational Assessments. Applied Psychological Measurement, 24* (4): 355-366, doi: 10.1177/0146621031804.
- Liljah, P. (2010). *The four Purposes of Assessment*. Retrieved 25th May 2017 from: <http://www.peterliljedahl.com/wp-content/uploads/4-Purposes-of-Assessment.pdf>
- National Council for Higher Education (NCHE). (2015). Statutory Instruments 2015 No. 21. The Universities and Other Tertiary Institutions (Grading and Classification of Undergraduate Degrees and Diplomas) Regulations, 2015. *Uganda Gazette*, No. 27, volume (CVIII): 77 - 79.
- Opolot-Okurut, C. (2010). Statistics in Uganda Schools: Challenges on Instruction and Assessment. *ICOTS (2010) Invited Paper Refereed*. Retrieved 26th June, 2017 from: https://iase-web.org/documents/papers/icots8/ICOTS8_10B3_OPOLOTOKURUT.pdf
- Stevenson, A. (Ed) (2015). *Oxford Dictionary of English* (3rd ed.). Oxford University Press, ISBN-13:9780199571123, eISBN: 9780191727665
- Surgenor, P. (2010). *Teaching Toolkit: Role of Assessment. UCD Teaching and Learning Resources*. Retrieved 5th July, 2017 from: <https://www.ucd.ie/t4cms/UCDTLT0026.pdf>
- The Gordon Commission (2012). *Assessment, Teaching, and Learning*. The Gordon Commission on the Future of Assessment in Education. 2 (2). April 2012.
- Valencia, W. S. (2002). *Understanding Assessment: Putting together the Puzzle. Current Research in Reading/Language Arts*. Houghton Mifflin Company. All rights reserved. UNI20M302 Litho in U.S.A. G-23043. Retrieved 13th July, 2017 from: <https://www.eduplace.com/state/author/valencia.pdf>
- William, D. (2013). *Assessment: The Bridge between Teaching and Learning*. *Voices from the Middle, 21* (2): 15 – 20.

SUB THEME H: USING TECHNOLOGY TO ENHANCE EFFICIENCY AND EFFECTIVENESS IN EDUCATIONAL ASSESSMENT

Item Calibration for School-based Assessment in Delta State Nigeria

Uruemu Charity **Ogbebor-Kigho**

uruemukigh@gmail.com, ogbeborremu@yahoo.com +2348037027063

Adams. O.U. **Onuka**

ada.otuoze@gmail.com/ adamonuka@yahoo.com +2348033564064

Institute of Education, University of Ibadan, Ibadan Nigeria

and

Henry O. **Owolabi**

Henrvowolabi2000@yahoo.com

Department of Adult and Primary Education, University of Ilorin, Nigeria

Abstract

Item calibration using Item Response Theory (IRT) has been in existence for decades in the developed countries. In Nigeria, development and calibration of achievement tests has been based mainly on the Classical Test Theory (CTT) framework until recently, when public examining bodies in Nigeria began to shift from the use of CTT to IRT in validating test items for assessing students' achievement for the purpose of placement and certification. Calibration of school-based tests in Nigeria secondary schools has not been given enough attention. It has been observed that teachers in Nigeria secondary schools make use of un-validated teacher-made tests for internal examinations. Thus, students' performance in these examinations does not correspond with their performance in their senior secondary certificate examinations conducted by public examination bodies in Nigeria. It is in the light of the foregoing that a school-based Economics test for senior secondary schools in Delta State using the 3PL-IRT model was developed and calibrated. Three research questions were answered in the calibration study. The study adopted a survey research type of ex-post facto design. The population consisted of all senior secondary school students in Delta State Nigeria. Multistage sampling procedure was employed to select participants for the study. The total sample size was 1,352 SS3 students. A validated Economics Achievement test (EAT) was used and the reliability coefficient of the instrument was determined using KR-21 and found to be 0.91. Data were analysed using BILOG-MG3. The findings revealed that the EAT fitted the 3PLM. Out of the 79 calibrated items, 46 items were found to be valid. The EAT items were reliable and valid for assessing students' achievement for internal examinations and also prepare them for external examination conducted by public examination bodies in Nigeria.

Keywords: Item Calibration, Item Response Theory (IRT) and School Based Assessment

Introduction

School Based Assessment (SBA) plays a vital role in enhancing and improving educational standards and national development of a nation. Student grading during SBA needs to be given more attention in order to enhance

student learning outcomes. School-based assessment is an assessment which is rooted in the teaching and learning process. The teacher is involved in the assessment process from the planning stage of the assessment to identifying and developing assessment tasks up to the point of evaluating student learning outcomes. School based assessment is an assessment that is teacher designed, systematic and continuous in nature. SBA is a complex assessment process that encompasses the use of different assessment techniques to collect and provide information for evaluation. To Onuka and Ogbebor (2016), School based assessment involves the teacher and the taught from the beginning to the end in order to both improve the process and also make some judgement on the effectiveness of the teaching and learning process. School Based Assessment involves the continuous assessment of student learning at intervals in the three learning domains - cognitive, affective and psychomotor - using different instruments such as tests, assignments, observations, interviews, questionnaires, and projects. School Based Assessment should be formative, summative, performance-based, and authentic.

SBA is an internal assessment involving school administrators, teachers and students, in an effort investigate the extent to which students have achieved or learnt what is expected of them during the teaching and learning process. To Ogbebor (2017), formative assessment is an aspect of SBA which involves the collection of formal and informal assessment procedures carried out by teachers during the teaching and learning process in order to improve students' learning outcomes. Thus, it is imperative that school based test items used by teachers to measure students' mastery of the course content taught should be reliable and valid.

Researchers (Dosumu 2002; Alele-Williams 2002; Osadebe 2012; Muhammad, Iram & Abdul 2012; Onuka & Ogbebor 2016; Ogbebor 2017) all stressed the need to incorporate the use of valid tests for internal assessment of student's achievement. It is based on this notion that test items should be validated and calibrated for Nigeria secondary schools. Item calibration is an aspect of item response theory (IRT) for validating a test. It is the process of estimating parameters (item difficulty, discrimination, guessing and speediness) of bulky numbers of items. Item calibration provide an model in which different forms of a test have a set of linking items that allow the formation of a common scale on which all item parameter estimates can be expressed. Item parameters varies, depending on the IRT model used, for one parameter model items difficulty is measured, two parameter model adds item discrimination, and a three parameter model in addition has a guessing parameter and so on. According to Ogbebor (2017), Item Response Theory (IRT) is based on latent trait theory that incorporates measurement assumptions about examinee, items, test performance, and how performance can be attributed to knowledge as measured by test items. These models handle item responses that are continuously and dichotomously or polytomously scored. IRT aims at item-level information. Item calibration using IRT helps to enhance the reliability and validity of a test. From the forgoing, SBA is essential in improving student achievement, therefore, the measurement instruments or test used should be valid and of standard in order to help students perform better in public examinations.

Objectives of the Study

In the light of the aforementioned of School Based Assessment test and its validity issues this study aims to

1. Calibration Economics Achievement test items using the three-parameter logistic model of IRT?
2. Determine model fit of the Developed Economics test items

3. Develop a valid school-based test for assessing student learning.

Theoretical Background

The theories underlying this study are Theory of test validity and the item Response Theory (IRT).

Test Validity Theory

The theory of test validity assumes that error scores on two separate tests are uncorrelated. The expected-value concept of true score in the classical test theory model as formulated by Lord and Novick, Guttman, and others, infers scientifically, that true scores and error scores are uncorrelated. This concept does not imply, however, that error scores on two arbitrary tests are uncorrelated, and an additional axiom of “experimental independence” is needed in order to obtain familiar results in the theory of test validity (Zimmerman, 1977). Observed scores on two tests can have a positively correlated value even when the true scores are negatively correlated, and the validity coefficient can exceed the index of reliability. CTT statistics can also be used to determine values of reliability and item co-variance. Hitches inherent in the reliability of a test are circumscribed by the Kuder-Richardson 20 (KR-20) or Kuder-Richardson 21 (KR-21 and Cronbach’s Coefficient Alpha. In test validity theory practical interest on validity coefficient will decrease with increase in test length. These irregularities sometimes occur even when the correlation between error scores is fairly minor, and their magnitude is inversely related to test reliability. The elimination of correlated errors in practice will not enhance a test’s predictive value, but will restore the psychometric properties of the validity that are familiar in the classical theory. Rust and Golombok (2000) refer to such errors as sabotage in assessment.

Item Response Theory

Item Response Theory (IRT), is based on latent trait theory, and integrates measurement assumptions about

the learners to be examined, test items, test scores, and examinees ability level and how test scores generated can be ascribed to what is being taught and learnt as measured by test items (Ogbebor, 2015). IRT is most valid in a latent trait that can be measured on a scale having a midpoint of zero (0) and a unit measurement of one. The range of ability is from negative to positive infinity, with practical considerations usually limiting the range of values from -5 to +5.

In measuring latent traits such as ability, item characteristic curves can be modeled for each individual item, showing the item’s difficulty and discrimination. The use of principles under Item Response Theory can be applied to many different types of models to help increase the reliability and validity of items and tests. Some most common models used for dichotomously scored tests include the Rasch or one-parameter logistic model which was proposed by Rasch in 1960 and Birnbaum’s two and three-parameter logistic models respectively which were proposed by Allen Birnbaum in 1968. Item response theory has three models which are known as three, two, one parameters. The simplest of the models is the one parameter while the most complex is the three parameter model. Item response theory investigates three assumptions which are unidimensionality of a test, local item independence and item characteristics curves. Unidimensionality of a test implies that a test should measure only one trait or knowledge, item local independence also refers to the fact that an examinee’s scores are independent of all other examinees’ scores and finally, item characteristics curve (ICC) is a mathematical monotonically increasing function that describes the relationship between an examinee’s item performance and the trait underlying the item performance (Ogbebor 2017).

Problem Statement

The use of item response theory for item calibrations of school based assessment is not new in developed countries but in Nigeria school based tests for assessing students learning outcomes are not valid. Thus, students' performance in their internal examination does not often appear to correctly predict their performance in their School Certificate Examination that public examining bodies in Nigeria conduct. It is in the light of the foregoing, that this study calibrated an Economics test for assessing student achievement in senior secondary schools in Delta State.

Research Questions

1. What is the Economics Achievement Test (EAT) Information curve?
2. What are the calibrated estimates of the Economics Achievement Test (EAT) items using the three-parameter logistic model of IRT?
3. How best do the EAT items fit into the 3PL model of IRT?
4. How many EAT Items were selected as valid items based on the item selection standard?

Methods and Procedures

Design

The survey research type of ex-post facto design was used in this study. Ex-post facto was appropriate because the variables of interest were not manipulated as they had already occurred.

Population

The population comprises all Senior Secondary School three students' (SSS3) who offered Economics in Delta State during 2014/2015 academic year.

Sampling Techniques and Sample

The Multi-stage sampling design was adopted. Simple random sampling technique was used to select one senatorial district (Delta Central) of the existing three senatorial districts in Delta State. Simple random sampling technique was also used to select four (4) urban and four (4) rural Local Government Council

Areas (LGAs) (i.e. a total of eight (8) LGAs). Thirdly, simple random sampling was used to select four (4) public schools and four (4) private schools a piece from each of the selected Local Government Areas (LGAs). Thus, the total number of schools sampled was sixty four (64) senior secondary schools. An Intact class was selected from each of the sampled schools, thus, one thousand three hundred and fifty two (1,352) students participated in the study.

Instrumentation, Data Collection and Analysis

One instrument was used for this study: the Economics Achievement Test (EAT). A Developed Economics Test was constructed using the following procedure: one hundred and fifty (150) multiple choice items were drafted by the researchers. The items covered three levels of cognitive domain which includes Knowledge, Comprehension and Thinking as suggested by Obemeata (1991). The items covered the Senior Secondary Economics Curriculum for Nigerian students. The 100 items drafted by the researchers were subjected to review by experts in Economics who are also examiners in WAEC. Their recommendations in terms of clarity of words, ambiguity of items and also plausibility of the distracters were strictly adhered to. The information provided on each of the items was used to reconstruct thirteen (13) of the items.

In addition, the test items were trial tested. These items were administered to five hundred (500) SS3 students that were not part of the sample for the study. The data gathered were subjected to item analysis based on the criterion set (*p-value* 0.3 – 0.8, *d-value* 0.3- 0.9) for selection of items. In addition, 21 items were deleted while 79 items formed the EAT for item calibration using Item Response theory 3PL model. The KR-21 reliability method was used to obtain the reliability coefficient of the Developed Economics Test. The obtained value was 0.91. On the basis of this reliability coefficient, the instrument was considered reliable for the study and was employed to collect data. The resultant data were collated and BILOG-MG3 procedure for item calibration.

FINDINGS/RESULTS

Research Question One: What is the Economics Achievement Test (EAT) Information curve?

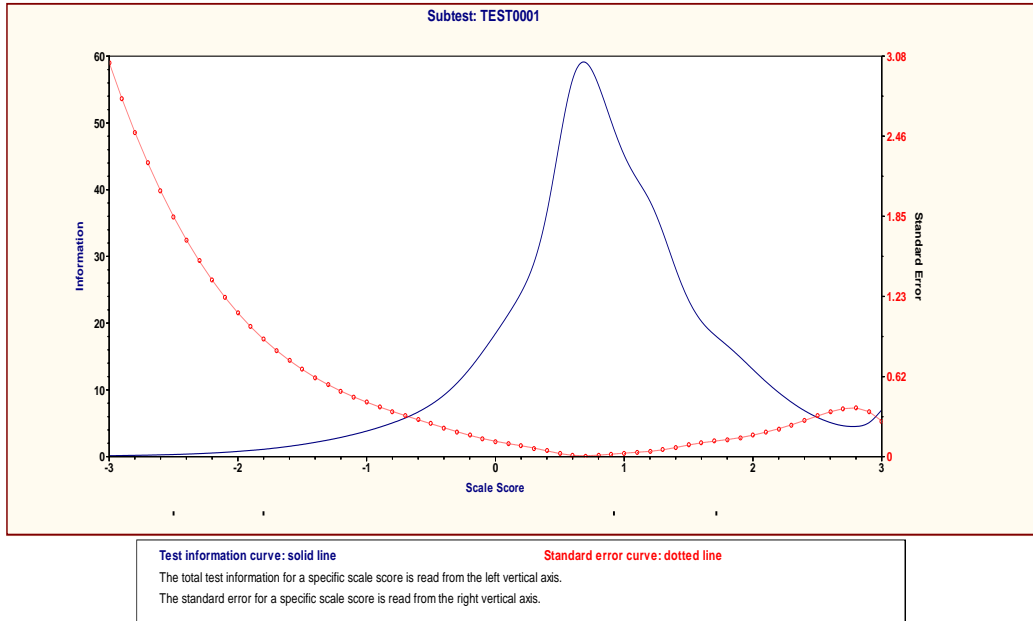


Figure 1: EAT Test Information Curve (TIC).

Figure 1 gave the maximum amount of information provided by the whole 79 EAT items that were calibrated as 58.5% at an ability level of 0.8 (the point at which the curve peaks). From the Test Information Curve (TIC), items with difficulty level between -0.7 to 2.5 (the point at which the dotted lines cross the solid lines and point where the dotted line and the solid line meet on the right hand side of the curve) gave valuable information.

The finding of this study supports the observation of DeAyala (2009) that an item provides its maximum information at its location that the Item Information Function is uni-modal and symmetric about δ , and that all items in an instrument provides the same maximum amount of information of α^2 0.25 at their respective locations.

Research question two: What are the calibrated estimates of the Economics Achievement Test (EAT) items using the three-parameter logistic model of IRT?

Table 1: Showing Item calibrated estimate using 3PL Model

| ITEMS | a | b | c | χ^2 value | Prob. | DF |
|-------|------|-------|------|----------------|-------|-----|
| 1 | 0.83 | 0.02 | 0.50 | 105.6 | 0.000 | 1.0 |
| 2 | 0.39 | 0.71 | 0.27 | 62.2 | 0.000 | 1.0 |
| 3 | 2.74 | 1.21 | 0.33 | 21.2 | 0.012 | 1.0 |
| 4 | 2.52 | -0.00 | 0.30 | 28.2 | 0.002 | 1.0 |
| 5 | 0.56 | 0.49 | 0.27 | 91.6 | 0.000 | 1.0 |
| 6 | 0.46 | 0.77 | 0.32 | 49.5 | 0.000 | 1.0 |
| 7 | 1.45 | 0.18 | 0.43 | 21.8 | 0.006 | 1.0 |
| 8 | 1.65 | 0.66 | 0.24 | 91.9 | 0.000 | 1.0 |
| 9 | 0.55 | -0.16 | 0.14 | 143.4 | 0.000 | 1.0 |
| 10 | 0.52 | 0.19 | 0.17 | 119.5 | 0.000 | 1.0 |
| 11 | 1.04 | 0.16 | 0.44 | 32.8 | 0.000 | 1.0 |
| 12 | 3.39 | 0.80 | 0.29 | 48.6 | 0.000 | 1.0 |
| 13 | 0.75 | 1.59 | 0.44 | 58.7 | 0.000 | 1.0 |
| 14 | 1.52 | 0.38 | 0.29 | 99.5 | 0.000 | 1.0 |
| 15 | 5.05 | 0.51 | 0.42 | 61.6 | 0.000 | 1.0 |
| 16 | 1.54 | 0.93 | 0.24 | 37.5 | 0.000 | 1.0 |
| 17 | 0.84 | 0.67 | 0.16 | 90.3 | 0.000 | 1.0 |
| 18 | 4.37 | 1.19 | 0.33 | 42.5 | 0.000 | 1.0 |
| 19 | 1.43 | 0.08 | 0.50 | 104.1 | 0.000 | 1.0 |
| 20 | 1.12 | 0.45 | 0.32 | 83.1 | 0.000 | 1.0 |
| 21 | 1.46 | 0.23 | 0.23 | 7.1 | 0.524 | 1.0 |
| 22 | 4.97 | 0.63 | 0.27 | 68.8 | 0.000 | 1.0 |
| 23 | 1.02 | -0.76 | 0.10 | 165.0 | 0.000 | 1.0 |
| 24 | 0.52 | 2.37 | 0.37 | 92.9 | 0.000 | 1.0 |
| 25 | 0.68 | 0.32 | 0.17 | 71.5 | 0.000 | 1.0 |
| 26 | 0.81 | 1.81 | 0.15 | 47.5 | 0.000 | 1.0 |
| 27 | 1.64 | 0.26 | 0.36 | 23.1 | 0.003 | 1.0 |
| 28 | 0.33 | 6.10 | 0.21 | 42.2 | 0.000 | 1.0 |
| 29 | 1.02 | 0.44 | 0.42 | 127.6 | 0.000 | 1.0 |
| 30 | 1.03 | 3.14 | 0.20 | 61.7 | 0.000 | 1.0 |
| 31 | 3.27 | 1.74 | 0.32 | 18.6 | 0.029 | 1.0 |
| 32 | 1.31 | 1.90 | 0.27 | 47.3 | 0.000 | 1.0 |
| 33 | 0.49 | 5.40 | 0.14 | 70.2 | 0.000 | 1.0 |
| 34 | 0.39 | 0.12 | 0.31 | 145.9 | 0.000 | 1.0 |
| 35 | 1.02 | -0.03 | 0.50 | 28.5 | 0.000 | 1.0 |
| 36 | 0.64 | 1.83 | 0.21 | 121.1 | 0.000 | 1.0 |
| 37 | 0.34 | 1.88 | 0.19 | 98.2 | 0.000 | 1.0 |
| 38 | 0.37 | 1.59 | 0.40 | 86.2 | 0.000 | 1.0 |

| | | | | | | |
|----|------|-------|------|-------|-------|-----|
| 39 | 0.84 | 0.52 | 0.19 | 9.3 | 0.000 | 1.0 |
| 40 | 0.28 | 4.13 | 0.18 | 191.2 | 0.000 | 1.0 |
| 41 | 1.05 | -0.16 | 0.24 | 111.3 | 0.000 | 1.0 |
| 42 | 0.77 | -0.59 | 0.19 | 137.5 | 0.000 | 1.0 |
| 43 | 0.89 | -0.10 | 0.17 | 75.0 | 0.000 | 1.0 |
| 44 | 0.82 | 0.89 | 0.28 | 49.5 | 0.000 | 1.0 |
| 45 | 1.35 | 0.81 | 0.29 | 49.3 | 0.000 | 1.0 |
| 46 | 0.50 | 2.49 | 0.21 | 93.5 | 0.000 | 1.0 |
| 47 | 0.69 | 0.43 | 0.27 | 110.9 | 0.000 | 1.0 |
| 48 | 2.00 | 1.43 | 0.36 | 56.7 | 0.000 | 1.0 |
| 49 | 1.95 | 0.73 | 0.25 | 30.4 | 0.002 | 1.0 |
| 50 | 0.99 | 0.24 | 0.33 | 31.8 | 0.000 | 1.0 |
| 51 | 2.02 | 0.92 | 0.46 | 119.0 | 0.000 | 1.0 |
| 52 | 0.34 | 1.53 | 0.29 | 175.4 | 0.000 | 1.0 |
| 53 | 2.15 | 0.35 | 0.25 | 22.9 | 0.000 | 1.0 |
| 54 | 1.34 | 1.06 | 0.16 | 40.3 | 0.000 | 1.0 |
| 55 | 0.72 | -0.39 | 0.17 | 84.2 | 0.000 | 1.0 |
| 56 | 0.53 | 0.53 | 0.24 | 109.0 | 0.000 | 1.0 |
| 57 | 1.09 | -0.10 | 0.17 | 75.2 | 0.000 | 1.0 |
| 58 | 1.01 | 0.78 | 0.30 | 34.4 | 0.000 | 1.0 |
| 59 | 2.39 | 0.47 | 0.24 | 60.4 | 0.000 | 1.0 |
| 60 | 0.25 | 5.56 | 0.27 | 148.2 | 0.000 | 1.0 |
| 61 | 1.75 | 0.57 | 0.34 | 39.2 | 0.000 | 1.0 |
| 62 | 2.16 | 0.77 | 0.28 | 66.5 | 0.000 | 1.0 |
| 63 | 0.82 | -0.43 | 0.22 | 80.9 | 0.000 | 1.0 |
| 64 | 3.22 | 0.70 | 0.31 | 45.5 | 0.000 | 1.0 |
| 65 | 0.76 | -0.31 | 0.10 | 113.9 | 0.000 | 1.0 |
| 66 | 2.10 | 0.75 | 0.32 | 66.0 | 0.000 | 1.0 |
| 67 | 1.18 | -0.01 | 0.22 | 54.1 | 0.000 | 1.0 |
| 68 | 1.95 | 0.23 | 0.44 | 18.3 | 0.019 | 1.0 |
| 69 | 2.94 | 0.79 | 0.30 | 34.4 | 0.000 | 1.0 |
| 70 | 0.92 | 1.04 | 0.21 | 56.7 | 0.000 | 1.0 |
| 71 | 0.67 | 0.36 | 0.42 | 55.4 | 0.000 | 1.0 |
| 72 | 0.12 | 7.61 | 0.22 | 193.7 | 0.000 | 1.0 |
| 73 | 0.67 | 0.23 | 0.32 | 57.2 | 0.000 | 1.0 |
| 74 | 2.13 | 0.20 | 0.37 | 37.3 | 0.000 | 1.0 |
| 75 | 2.71 | 0.79 | 0.24 | 55.4 | 0.000 | 1.0 |
| 76 | 2.11 | 0.88 | 0.26 | 64.2 | 0.000 | 1.0 |
| 77 | 0.47 | 1.48 | 0.26 | 209.6 | 0.000 | 1.0 |
| 78 | 2.55 | 2.04 | 0.44 | 67.7 | 0.000 | 1.0 |
| 79 | 0.98 | -0.42 | 0.15 | 46.7 | 0.000 | 1.0 |

χ^2 value = 5682.4
 $P_{(0.000)} < 0.05$

The EAT items were analysed using the IRT 3PL model of BILOG. Table 1 shows the estimated items difficulty (b), discrimination (a) and guessing (c) for the 79 calibrated EAT items. The difficulty index, (b) of the MAT items ranges from -0.76 to 7.61, it implies that the most difficult items have positive values while the easier items have negative values. Thus, it can be inferred that item 72 (difficulty index of 7.61) is the most difficult while item 23 (difficulty index of -0.76) is the least difficult.

The discriminating (a) values of the EAT ranges from 0.12 to 5.05 this connotes that the most discriminating item is item 15 (a= 5.05) while the less discriminating item is 72 (a=0.12). Thus, discrimination index of the EAT are in the positive direction which is an indication that all EAT items are satisfactory in the context of discrimination index. Therefore, it could be inferred that all the items are valid under item discrimination. Furthermore, items guessing values of the EAT ranges from 0.10 to 0.50 this indicates that item with high tendency to guess is item 35 (c=0.50) while the item with less tendency to guess is item 23 (c=0.10).

Research Question Three: How best do the EAT items fit into the 3PL model of IRT? To determine whether the EAT items fit the 3PLM, Chi-square (χ^2) value of BILOG-MG3 was used as shown in Table 1. The Economics Achievement Test (EAT) fitted the 3PL model with χ^2 value = 5682.4 $P_{(0.000)} < 0.05$. The result implies that EAT calibrated fitted the 3PL model of IRT using the χ^2 approach.

Research Question Three: How many EAT Items were selected as valid items based on item selection standard? Three things were considered in selecting valid items these includes

- i. Item difficulty index: According to IRT from scientific software international - SSI (2003), items with difficulty range of -1.0 to +1.0 could be considered valid items as they are neither too difficult nor too easy (Scientific Software International - SSI 2003; Aliyu & Ogbebor 2015).
- ii. Item discrimination: Items discrimination should have or display a positive direction.
- iii. Guessing tendency of items: valid items under guessing index to be those not above 0.35 (Aliyu & Ogbebor 2015; SSI 2003). This indicates that the probability of tendency to guessing should be < 0.35 .

Table 2: Showing Selected Items Using the Criterion

| ITEMS | A | B | C | Remarks |
|-------|------|-------|------|--------------|
| 1 | 0.83 | 0.02 | 0.50 | Selected |
| 2 | 0.39 | 0.71 | 0.27 | Selected |
| 3 | 2.74 | 1.21 | 0.33 | Not Selected |
| 4 | 2.52 | -0.01 | 0.30 | Selected |
| 5 | 0.56 | 0.49 | 0.27 | Selected |
| 6 | 0.46 | 0.77 | 0.32 | Selected |
| 7 | 1.45 | 0.18 | 0.43 | Not Selected |
| 8 | 1.65 | 0.66 | 0.24 | Selected |
| 9 | 0.55 | -0.16 | 0.14 | Selected |
| 10 | 0.52 | 0.19 | 0.17 | Selected |
| 11 | 1.04 | 0.16 | 0.44 | Not Selected |
| 12 | 3.39 | 0.80 | 0.29 | Selected |
| 13 | 0.75 | 1.59 | 0.44 | Not Selected |
| 14 | 1.52 | 0.38 | 0.29 | Selected |
| 15 | 5.05 | 0.51 | 0.42 | Not Selected |
| 16 | 1.54 | 0.93 | 0.24 | Selected |
| 17 | 0.84 | 0.67 | 0.16 | Selected |
| 18 | 4.37 | 1.19 | 0.33 | Not Selected |
| 19 | 1.43 | 0.08 | 0.50 | Not Selected |
| 20 | 1.12 | 0.45 | 0.32 | Selected |
| 21 | 1.46 | 0.23 | 0.23 | Selected |
| 22 | 4.97 | 0.63 | 0.27 | Selected |
| 23 | 1.02 | -0.76 | 0.10 | Selected |
| 24 | 0.52 | 2.37 | 0.37 | Not Selected |
| 25 | 0.68 | 0.32 | 0.17 | Selected |
| 26 | 0.81 | 1.81 | 0.15 | Not Selected |
| 27 | 1.64 | 0.26 | 0.36 | Not Selected |
| 28 | 0.33 | 6.10 | 0.21 | Not Selected |
| 29 | 1.02 | 0.44 | 0.42 | Not Selected |
| 30 | 1.03 | 3.14 | 0.20 | Not Selected |
| 31 | 3.27 | 1.74 | 0.32 | Not Selected |
| 32 | 1.31 | 1.90 | 0.27 | Not Selected |
| 33 | 0.49 | 5.40 | 0.14 | Not Selected |
| 34 | 0.39 | 0.12 | 0.31 | Selected |
| 35 | 1.02 | -0.03 | 0.50 | Not Selected |
| 36 | 0.64 | 1.83 | 0.21 | Not Selected |
| 37 | 0.34 | 1.88 | 0.19 | Not Selected |
| 38 | 0.37 | 1.59 | 0.40 | Not Selected |

| | | | | |
|----|------|-------|------|--------------|
| 39 | 0.84 | 0.52 | 0.19 | Selected |
| 40 | 0.28 | 4.13 | 0.18 | Not Selected |
| 41 | 1.05 | -0.16 | 0.24 | Selected |
| 42 | 0.77 | -0.59 | 0.19 | Selected |
| 43 | 0.89 | -0.10 | 0.17 | Selected |
| 44 | 0.82 | 0.89 | 0.28 | Selected |
| 45 | 1.35 | 0.81 | 0.29 | Selected |
| 46 | 0.50 | 2.49 | 0.21 | Not Selected |
| 47 | 0.69 | 0.43 | 0.27 | Selected |
| 48 | 2.00 | 1.43 | 0.36 | Not Selected |
| 49 | 1.95 | 0.73 | 0.25 | Selected |
| 50 | 0.99 | 0.24 | 0.33 | Selected |
| 51 | 2.02 | 0.92 | 0.46 | Selected |
| 52 | 0.34 | 1.53 | 0.29 | Not Selected |
| 53 | 2.15 | 0.35 | 0.25 | Selected |
| 54 | 1.34 | 1.06 | 0.16 | Not Selected |
| 55 | 0.72 | -0.39 | 0.17 | Selected |
| 56 | 0.53 | 0.53 | 0.24 | Selected |
| 57 | 1.09 | -0.10 | 0.17 | Selected |
| 58 | 1.01 | 0.78 | 0.30 | Selected |
| 59 | 2.39 | 0.47 | 0.24 | Selected |
| 60 | 0.25 | 5.56 | 0.27 | Not Selected |
| 61 | 1.75 | 0.57 | 0.34 | Selected |
| 62 | 2.16 | 0.77 | 0.28 | Selected |
| 63 | 0.82 | -0.43 | 0.22 | Selected |
| 64 | 3.22 | 0.70 | 0.31 | Selected |
| 65 | 0.76 | -0.31 | 0.10 | Selected |
| 66 | 2.10 | 0.75 | 0.32 | Selected |
| 67 | 1.18 | -0.01 | 0.22 | Selected |
| 68 | 1.95 | 0.23 | 0.44 | Not Selected |
| 69 | 2.94 | 0.79 | 0.30 | Selected |
| 70 | 0.92 | 1.04 | 0.21 | Not Selected |
| 71 | 0.67 | 0.36 | 0.42 | Not Selected |
| 72 | 0.12 | 7.61 | 0.22 | Not Selected |
| 73 | 0.67 | 0.23 | 0.32 | Selected |
| 74 | 2.13 | 0.20 | 0.37 | Not Selected |
| 75 | 2.71 | 0.79 | 0.24 | Selected |
| 76 | 2.11 | 0.88 | 0.26 | Selected |
| 77 | 0.47 | 1.48 | 0.26 | Not Selected |
| 78 | 2.55 | 2.04 | 0.44 | Not Selected |
| 79 | 0.98 | -0.42 | 0.15 | Selected |

As shown in Table 2, 33 items (3, 7, 11, 13, 15, 18, 19, 24, 26, 27, 28, 29, 30, 31, 32, 33, 35, 36, 37, 38, 40, 46, 48, 52, 54, 60, 68, 70, 71, 72, 74, 77, & 78) did not satisfy the criterion for selection, while forty 46 items formed the final Economics Achievement Test.

Discussion

The finding of this study shows that the Test information curve gave maximum amount of information on the 79 EAT items that were calibrated as 58.5% at an ability level of 0.8 (the point at which the curve peaks). From the Test Information Curve (TIC), The finding of this study supports the observation of DeAyala (2009) that an item provides its maximum information at its location that the Item Information Function is uni-modal and symmetric about δ , and that all items in an instrument provides the same maximum amount of information of α^2 0.25 at their respective locations.

The study found that most difficult items have positive values while the easier items have negative values, discrimination index of the EAT are in the positive direction which is an indication that all EAT items are satisfactory in the context of discrimination index. While tendency to guessing was measured using <0.35 to select valid items. Furthermore, item selected after calibration was 46 while 33 items were not selected as valid. This finding is in tandem with Ogbebor (2017) that item difficulty ranges from -0.7 to 2.5 of the developed Mock Economics Test using TIC were selected as valid. This study supports the result of Gruijter and Kamp (2002) which states that 34 items that fitted the chosen model were selected as most appropriate items in measuring students' ability in Mathematics. In same disposition, Adedoyin (2010) used χ^2 test with probability greater than the alpha level of 0.05 significant levels to select items that fitted models. This study is in line with the report of Ojerinde, *et al*, (2012) which state that selection of items was based on criterion set for item difficulty index ranges from -1.0 to +1.0, discrimination index from -3.0 to +3.0. The result is in consonance with

the study of Aliyu and Ogbebor (2015) that items selected were based on the range from -0.96 to 0.97 for item difficulty, item discrimination ranges from 0.05 to 2.23 while probability of guessing ranged from 0.00 to 0.32 for mathematics achievement test.

In same vein, the result of this study supports the result of Harris (1989) who revealed that item selection ranges from -3.0 to +3.0 for discrimination, while for difficulty index items selection ranges from -1.0 to +1.0. Additionally, this study corroborates the report of the Scientific Software International (SSI 2003) that tendency towards guessing should not exceed 0.35

Conclusions, Implications and Recommendations

Conclusion

This study calibrated Economics Achievement Test in Delta State, Nigeria using BILOG-MG for Item Response Theory 3PL model. IRT models focused on both item and person statistics such as the item parameters, test information curve and test information functions. These features depicted IRT as a better option in giving acceptable information regarding the behaviour of an item as well as the examinees. Based on these facts, this IRT model was used to calibrate EAT. The study also investigated model fit of the Economics Achievement test. The EAT should be used to assess Senior Secondary School III students' achievement in Economics when they have covered what is expected of them using the curriculum as a benchmark. This would enhance the quality of assessment and create bias free and fair tests. Hence, it is needful to have an existing valid Economics test for secondary schools where school administrators and teachers can gain access to valid test items.

Implications

The findings of this study have implications for school management, teachers and learners. The use Item Response Theory is very essential in item development and calibration. However, the Item Response Theory is more effective in calibration of test items as it enhances

selecting items that best measure students' ability and equally gives adequate information concerning the behaviour of an item as well as the examinees.

Recommendations

Based on the findings of this study, the following recommendations were made:

- School-based examinations in secondary schools should make use of items that are validated and calibrated with IRT for effective measurement of learners' ability.
- Valid School based test should be used for internal examinations in Delta State.
- This EAT test should be used by school teachers and other school personnel to prepare Senior Secondary School three (SSS 111) students for certificate examinations.
- Teacher-made tests should be discouraged in our secondary schools.

References

- Adedoyin, O. O. (2010). Using IRT approach to detect gender biased items in public examinations. *Educational Research and Reviews Academic Journals*, 5(7), 385-399.
- Alele-Williams, G. (2002). Measurement and evaluation in Mathematics: The way forward. *Journal of Basic Science*, 1(1), 1-7.
- Aliyu R. T. & Ogbebor U. C. (2015). Extraction of psychometric properties of mathematics achievement test using the 3PL model of IRT. *Journal of educational research and development (AJERD)*, 2(1), 35-51.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- DeAyala, R. J. (2009). *Theory and practice item response theory*. New York: Guilford
- Dosumu, C. T. (2002). *Issues in teacher-made tests*. Ibadan: Olatunji and Sons Publishers.
- Harris, D. (1989). Comparison of 1,2,3 parameter IRT models. *America College Testing Program, NCME Instructional Model*, Iowa City.
- Gruijter, D, N. & Kemp, J. T. C. (2002). *Statistical test theory for education and psychology*. Boston: Kluwer.
<http://www.performancetest.org/uploads/resources/.pdf>.
- IRT from SSI: (2003). BILOG-MG, MULTILOG, PARSCALE, TESTFACT (Ed.) Math; Ida du Toit SSI. Scientific Software International. Author.
- Muhammed, J. Q., Iram, G. G. & Abdul, H. (2012). Rasch calibration of General Science Test at Grade-VIII in Pakistan. *Journal of Education and Vocational Research*, 3(3), 98-106.
- Obemeata, J.O. (1991). Effective teaching of economics in senior secondary. *West African Journal of Education*. XXII (1), 9-13
- Ogbebor, U. C. (2015). *Mock Economics test construction using classical test and item response theories in Delta State Nigeria*. (Unpublished Pre field doctoral proposal, International Centre for Educational Evaluation, Institute of Education, University of Ibadan, Nigeria).
- Ogbebor, U. C. (2017). *Construction and validation of mock Economics test using classical test and item response theories for senior secondary school students in Delta State, Nigeria*. (Unpublished Ph.D. Thesis, University of Ibadan, Nigeria).
- Ojerinde D., Popoola. K., Ojo, F. & Onyeneho, P. (2012). *Introduction to item response theory*. Ibadan: Goshen Print Media Ltd.
- Onuka A. O. U. & Ogbebor, U. C. (2016). Teachers' test construction skills: A tool for improved learning achievement in South-South, Nigeria. Paper

presented at the 34TH AEAA Conference Held in Victoria Falls Zimbabwe, July, 2016.

Osadebe, P. I (2012). *Procedures for construction, validation and standardization of test*. A Seminar Paper Presented at Delta State University Abraka.

Rust, J. & Golombok, S. (2000). The GRISS: A psychometric scale and profile of sexual dysfunction. In C. M. Davis, W. L. Yarber, R. Bauserman, G. Schreer & S. L. Davis (Eds.) *Handbook of sexually related measures* (pp. 192-194). California: Sage.

Parameterization of Test Using Item Response Theory (IRT): A Case Study of English Language Essay Test of the National Business and Technical Examinations Board (NABTEB)

By

Prof. (Mrs.) Ifeoma Mercy Isiugo-Abanihe:

E-mail: Ifyabanihe@gmail.com

Tel: +2347031375555,

Ibidapo Olabode Ilesanmi:

E-mail: bodeibidapo@yahoo.com

Tel: + 2348062945129

and

Irene Ovekairi Iluobe (Mrs.):

E-mail: ireneovekairi@gmail.com

Tel: + 2348033593364

**National Business and Technical Examinations Board (NABTEB), Benin City,
Nigeria**

Abstract

This study determined the item parameters of the English Language test of the 2014 May/June examination of the National Business and Technical Examinations Board (NABTEB) using the Item Response Theory (IRT). The descriptive survey research design was adopted. The population of the study consists of all the students who registered for the 2015 NABTEB May/June Examinations in the South-West and South-South Geopolitical Zones in Nigeria. The sample comprised of six hundred (600) students, made up of three hundred and sixty-three (363) males and two hundred and thirty-seven (237) females, randomly selected from six technical and vocational colleges. Four research questions guided the study. The researchers adopted the NABTEB 2014 May/June NBC/NTC English Language Essay test and marking guide for the study. This instrument was presumed to be valid and reliable because the Board had ascertained that. The responses of the students were scored polytomously by English Language experts, and a person by item matrix was prepared. The data was analyzed using eirt Item Response Theory Excel Assistance Version 3.1 software. From the analysis, the item parameters (difficulty and discrimination) were determined. The result revealed that 64% of the items were within acceptable difficulty level while 36% had unacceptable difficulty level. All the items, representing 100%, discriminated positively. Also, 93% of the items fitted the Graded Response Model (GRM) while there was 7% misfit. Based on the findings, it was recommended, among others, that items that had unacceptable difficulty level should be reframed. It was concluded that failures in the English Language (NABTEB) 2014 May/ June Examinations cannot be attributed to the difficulty and discrimination indices of the items.

Key words: Parameterization, English Language Test, Difficulty, Discrimination, IRT.

Introduction

Examinations are often used by educators as a means of determining students' level of achievement of a specific curriculum and instruction. From the performance of students in examinations, decisions are made. Decisions in education are made based on information gathered from examinations or non-test procedures. The concept of standards of a test is highly important; hence, it has been observed that standards in examinations should be central to the concern of all examination bodies and their stakeholders (Motshabi, Kesamang and Gabalebatse, 2011). Examinations are most often used for promotion, placement or award of certificates. They are used to assess the effects of teaching and learning, and for selection purposes. For examinations to serve the purposes they are meant for, efforts must be made to develop reliable and valid items that will yield accurate measurement of candidates' performance. Thus, examination bodies in Nigeria, such as the West African Examinations Council (WAEC), National Examinations Council (NECO) and the National Business and Technical Examinations Board (NABTEB) adopt various processes to ensure validity and reliability of their examination items.

Over the years, the nation's public examination system has been bedeviled by series of challenges. Adeyegbe

(2002) found that there was a decline in students' performance in Senior Secondary Certificate (SSC) examinations. Onipede (2003) also reported that students performed below expectation in SSC examinations in many subject areas especially in English Language and Mathematics. Students' academic performance in core subjects, such as English Language, in external examinations has been on the decline. English language is a core subject offered in post basic level (secondary level) in Nigeria. It is one of the general subjects in which hundreds of candidates are tested by NABTEB during its May /June and November /December examination series. A credit pass is required in English Language with five other subjects before a candidate is certified by NABTEB and it is also a prerequisite requirement for admission into tertiary institutions. Due to the importance of English Language as a compulsory subject in the NABTEB examinations, this research set out to investigate the validity of its items as a means of verifying the quality of the examination.

The table 1 below shows the total number of students who obtained five credits and above including Mathematics and English Language from 2010 to 2014 in May/June National Business and Technical Examinations Board (NABTEB) Certificate Examinations.

Table 1:

Trend Analysis of Students' who had five credits and above including Mathematics and English Language in NABTEB May/June NBC/NTC Examinations from 2010 – 2014

| Year | 2010 | 2011 | 2012 | 2013 | 2014 |
|---------------------|-------------|-------------|-------------|-------------|-------------|
| Total sat | 61,775 | 109,416 | 147,100 | 103,753 | 79,495 |
| Credits Pass | 20,554 | 35,959 | 48,155 | 28,230 | 29,129 |
| Percentage | (32.31%) | (32.86%) | (35.27%) | (27.2%) | (36.64) |

SOURCE: NABTEB (2015) Statistical Digest

Other researchers (Onukaogu, 1994; Obemeata, 1995; Adepoju, 2002; Labo-Popoola, 2002; and Fakeye, 2012) have identified different factors that could lead to students' failure in examinations. These studies argue that students' academic problems arise from personal inadequacies such as low ability, negative self concept, anxiety, maladjustment, environmental influences such as poor classroom conditions, curricular inadequacies, peer groups and the lack of

home support. They failed, however, to look at the parameters of test items. It has been found that item characteristics can affect the performance of the students (Hamzeh, 2004).

Despite the importance of English Language in the public examinations, students' performance has remained poor. There are multiplicity of factors that could be responsible for this. Item-related factors, such as the quality of item parameters used for the examination could be responsible for the poor performance. Test quality is determined by its psychometric properties which include item difficulty (b), discrimination (a) and guessing (c). These parameters are employed during measurement as the testees' responses to the items are essential for analyzing the psychometric properties. The properties contained in the items determine to a large extent the trait in every examinee's capability and the item characteristics. This approach is conceptualized in Item Response Theory (IRT).

The Item Response Theory (IRT) has parameters which are determined for both items and examinees. Within the IRT context, item difficulty parameter refers to an examinee having the cognitive resistance of the item or task. This is the amount of trait, under measurement, that is just necessary to overcome the task or item. Item discrimination parameter, on the other hand, determines how, in actual fact, items can distinguish between highly capable and incapable students. Item discrimination parameter takes full advantage of supporting and guiding

the process of selecting items with high discrimination power while guessing parameter depicts the likelihood that examinees with low talent performs a difficult task above his capability. From the foregoing, the psychometric parameters are very important, considering the significant role of English Language in students' academic progression.

Statement of the Problem

Most past and present research have been grounded on dichotomous items (correct and incorrect) neglecting the polytomous items where there are multiple responses; and, partial credit is awarded for each response step. It is against this backdrop that this research finds its relevance. This study was motivated by a desire to assess the response pattern of examinees in NABTEB examinations. It was also undertaken to determine the item parameters of polytomously scored NABTEB Essay English Language test using IRT framework. A crucial question here is whether the lack of objectivity in the polytomously scored items is the cause of poor performance of candidates in the NABTEB English Language examinations.

Specific Research Questions

To guide this study, the following research questions were formulated:

1. What is the item difficulty parameter of NABTEB Essay English Language NBC/NTC 2014 Examinations?
2. What is the item discrimination parameter of NABTEB Essay English Language NBC/NTC 2014 Examinations?
3. What is the test function of the NABTEB Essay English Language NBC/NTC 2014 Examinations?
4. What is the item fit statistics of the NABTEB Essay English Language NBC/NTC 2014 Examinations?

Theoretical framework

This study was anchored on Item Response Theory (IRT). The pioneering work of IRT as a theory occurred during the 1950s and 1960s. Frederic M. Lord, the Educational Testing Service Psychometrician, George Rasch a Danish Mathematician and Paul Lazardfeld, an Austrian Sociologist, were the three pioneers of IRT who pursued parallel research independently. Others, who furthered the progress of IRT, include Benjamin Drake Wright and David Andrich. IRT is based on the latent trait theory, and in the cognitive domain of educational objectives, incorporates measurement assumptions about an examinee, item and test performance and how the performance relates to knowledge as measured by individual items on a test. IRT focuses on the item as opposed to the test-level focus of the Classical Test Theory (CTT). The IRT theory models the relationship between the responses of each examinee of a given ability of each item in the test, (Lord, cited in Amasingha, 2015).

Classical Test Theory (CTT)

The Classical Test Theory (CTT) is one of the earliest test theories and may be regarded as roughly synonymous with true score theory. It assumes that each person has a true score that would be obtained if there were no errors in measurement. Over an infinite number of independent administration of the test, a person's true score is defined as the expected number of correct scores. Discussing the concept of CTT, Ojerinde, Popoola, Ojo and Onyeneho (2012), pointed out that it is a theory about test scores which introduces the following three concepts:

Test score, X (often called the observed score)

True score, T and

Error score, E

CTT is a relatively simple theory for testing and has been widely used for constructing and evaluating tests, particularly before the birth of IRT (Duong, 2004).

CTT is basically concerned with the relation among X, T and E in any given population, as explained by the following linear model:

$$X = T + E \quad \dots\dots\dots \text{(Equation 1)}$$

Thus, Observed score = True score + Error score

Item Response Theory (IRT)

IRT is a family of mathematical models that describe the interaction between examinees and tasks (test items). In the development of high stakes tests, IRT methods is presently regarded as superior to classical test theory (CTT), therefore it is the preferred theory. According to Joshua (2014), IRT is seen as the most significant and popular development in psychometrics to overcome the shortcomings of CTT, and maximize objectivity in measurement. He further reiterates that IRT paints a picture that there is an encounter between an individual (testee) and an item during a testing situation. It is equally stressed that IRT attempts to model the relationship between unobserved variable, conceptualized as examinee's ability, and the probability of the examinee correctly responding to any particular test item. It is based on the above premise, that the researchers plan to embark on improving examinees' performance in polytomous test items through the application of the Item Response Theory (IRT).

IRT attempts to model the relationship between an unobservable variable referred to as the examinee correctly responding to any particular test item (Lord, 1986). It is a proven assessment approach that can help assessors to identify good or bad items, create better tests and give better estimates of examinee' abilities. The main idea of item response theory is that, it is a mathematical function describing the probability of specified responses to an item, given some level of quantitative attributes of the respondent. This is explained by Item Characteristic Curve (ICC) which scales items and people onto a common metric, helps in standard setting serves as foundation of equating and makes meaning in terms of student ability. ICC is illustrated by a line in a Cartesian system called Ogive which is defined by a logistic function shown below:

$$P_{ij}(1) | \theta, b = \frac{\text{Exp}(\theta_j - b_i)}{1 + \text{exp}(\theta_j - b_i)}$$

Where

b is the item parameter, and

Θ is the person parameter

The equation represents the probability of responding correctly to item i given the ability of person j while figure 1 below represents ICC which shows the behavior of a good item.

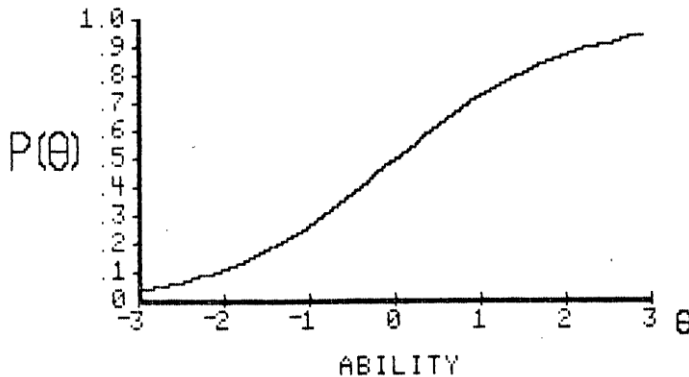


Figure 1: Item Characteristic Curve (ICC) (Source: Baker, 2001)

The item characteristic curve is the basic building block of Item Response Theory; all the other constructs of the theory depend upon this curve (Baker, 2001). The vertical axis represents the probability (.0 to 1.0) of responding correctly to the item while the horizontal axis represents the latent trait/Ability (-3 to 3) of the respondents. IRT is based on the idea that the probability of a correct response to an item is a function of the person and item parameter.

The assumption is that whatever the ability, it can be measured on a scale having a midpoint of zero, a unit of measurement of one, and a range from negative infinity to positive infinity. Since there is a unit of measurement and arbitrary zero point, with such a scale referred to as

existing at an interval level of measurement, the underlying idea is that if one could physically ascertain the ability of a person, the measurement scale would be used to tell how much ability a given person has, and the ability of several persons could also be compared. In IRT, estimates of respondents' traits (θ) are based not only on the responses they provide, but also the characteristics (that is, parameters) of the items they are administered such as their difficulty – reflected by category boundary parameters (b) – their ability to differentiate among respondents – reflected by slope parameters (a) – and their susceptibility to guessing –

reflected by lower asymptote parameters (c), (Mislevy, 2011).

IRT has models for both dichotomously scored items (e.g., correct and incorrect), and polytomously scored questions (e.g., 5 category Likert-type scale). IRT item parameters are set to relate responses to the underlying trait (Embretson and Reise, 2000), thus, IRT can easily model the mixed item formats included in many surveys. Joshua (2014) states that IRT and the presentations of it in models predispose IRT to many applications in practical testing situations.

Assumptions of IRT:

Three common assumptions of item response theory are: Uni-dimensionality, Local Independence and Monotonicity or Item Response Functions (IRF) or Item Characteristic Curve (ICC)

Uni-dimensionality:

IRT assumes that a single latent ability is sufficient to explain or account for examinees performance; this is referred to as uni-dimensionality. Reckase cited in Udom (2004) suggests that uni-dimensionality can be

investigated through the eigenvalues in a factor analysis. A test is said to be uni-dimensional if when plotting the eigenvalues (from the largest to the smallest) of the inter-item correlation matrix there is one dominant first factor.

Local Independence

The assumption of local independence states that the probability of an examinee answering a test correctly is not affected for better or for worse by his or her performance on any other item in the test. Reyes (2010) states that local independence means that once the appropriate number of latent traits is specified for a model, at a given value of the latent trait, item responses should be uncorrelated. The appropriateness of local independence can be tested by factor analysis technique and tetra-choric correlation between items with exactly the same ability. Violation of this assumption may result

in parameter estimates that are different from what they would be if the data were locally independent; thus selecting items for scale construction based on these estimates may lead to erroneous decisions (Chen & Thissen, cited in Reeve, 2000).

Monotonicity:

Logistic IRT models assume a monotonically increasing function (as trait level increases, so does the probability of endorsing an item). It is best displayed on a graph as a curve shaped like an 'S' between the latent trait level on the

X-axis and the probability of a more extreme response on the Y-axis. This curve, called an item characteristic curve (ICC), is assumed to graphically depict the true relationship between the trait and the response to the item. An ICC represents the probability of a correct answer to an item expressed as a function of ability. For data to be amenable to the IRT analysis there should be a fit of the model to the data set, in other words the data available must be, such that, allows the items to be modelled with an ICC that is derived from an item response function. This is most easily examined by plotting the corresponding ICC for all items in the test (Lord, cited in Wiberg, 2004).

Models for Scoring Polytomous Items

Three widely-used IRT models for polytomous scoring are available. These are: the Graded response model, the Partial credit model, and the Generalized partial credit models. Others are the Rating Scale Model and the Norminal Model. The graded response model is of interest in this study.

Graded Response Model (Samejima's model)

Samejima (1969) proposed a model for graded or ordered responses for questions with three or more response categories. A response may be graded on a range of scores, as an example, from poor (0) to excellent (9). For survey measurement, a subject may choose one option out of a number of graded options, such as a five-point Likert-type scale: "strongly-disagree", "disagree", "neutral", "agree", and "strongly agree" (Mellenbergh, cited in

Embretson & Reise, 2000). The model assumes that the response categories are ordered along a continuum. It estimates a set of boundary location parameters for each item. It assumes that discrimination within each item is equal but there is discrimination among items (Guyer and Thompson, 2011). The boundary response function BRF is given as

$$P_{ig}^*(\theta_j) = \frac{\exp\{A_i(\theta_j - b_{ig})\}}{1 + \exp\{A_i(\theta_j - b_{ig})\}}$$

Where A_i is the product of the item discrimination (a_i) with D , b_{ig} is the boundary parameter for boundary g . $P_{ig}^*(\theta_j) = 1$ and $P_{m+1}^*(\theta_j) = 0$

Where $g = m - 1$ and m is the number of response categories.

Then the category response function is defined as $P_{ig}(\theta_j) = P_{ig}^*(\theta_j) - P_{ig} + 1$

Parameter Estimation Methods

The parameters in a test are unknown, therefore under IRT, tasks are performed to estimate the actual values of these parameters. The obtained item parameter estimates provides information as to the technical properties of the

items. Different approaches and techniques are applied in IRT to estimate parameters. The ability parameter can be estimated jointly with the item parameters or estimated with known item parameters, which have been previously estimated. If the ability parameter is not estimated jointly with the item parameters, the item parameters are first estimated from the item responses with the influence of the ability parameter taken away; the ability parameter is either eliminated through conditioning or integrated out through marginalization. Techniques used in parameter estimation include the maximum likelihood procedure (Baker, 1992); logistic regression (Reynolds, Perkins & Brutton, 1994); minimum chi-quadrant (Zwinderman & van der Wollenberg, 1990), and Bayesian modal estimation procedure (Mislevy, 1986; Baker, 1992). The

Marginal Maximum Likelihood procedure was used in this study.

Methodology

Research Design

This study adopted descriptive survey research design. A survey research design is a study in which the characteristics or attributes of an individual, objects or things are described in order to portray accurately the characteristics or interests of a given population. The researchers were able to collect data and described the characteristics features and facts about the population under study. Therefore the survey design was considered appropriate for the study.

Population of the Study

The population of the study consists of one thousand (1000) students in Technical and Vocational Colleges who registered for the 2015 May/June NBC/NTC Examinations in South-West and South-South geopolitical zones in Nigeria.

Sample and Sampling Techniques

Convenient sampling non-probability technique was used to select South-West and South-South Geopolitical Zone because the researchers strongly required

cooperation from schools authorities and students attention. Multi – stage sampling was employed: Stage 1: Schools were clustered into private and public; Stage 2: Random sampling was done to select six (6) schools, three(3) private and three (3) public schools. Each of the selected schools registered more than a hundred students. Stage 3: In each school one hundred (100) students were randomly selected. The total number of participants in this study was six hundred (600) students made up of three hundred sixty-three (363) males and two hundred and thirty-seven (237) females. This is because more males registered than females in Technical and Vocational colleges.

Research Instrument

The instrument used for this study was the NBC/NTC Essay English Language question paper for 2014 May/June Examinations conducted by National Business and Technical Examinations Board (NABTEB). The instrument consisted of two sections: A and B. In section A, students were to answer one (1) question out of four (4) questions while in section B; they were to answer all three (3) questions. The students were to answer four (4) questions in all. The response to each item for all the students in the schools selected was used.

Validity and reliability of the Instrument

The National Business and Technical Examinations Board's May/June NBC/NTC 2014 examinations essay test in English Language was presumed to be valid. This was based on the assumption that NABTEB being an examination board had validated their instruments. Therefore the instrument was presumed to be standardized, valid and reliable.

Collection Procedure

National Business and Technical Examinations Board's Essay English Language question paper for 2014 May/June NBC/NTC Examinations was adopted on the approval of the board. With the permission of the school principals of the various schools sampled, the researcher administered the test with the help of research assistants

to the students and the answer scripts were retrieved on the spot. The students were given two weeks of notice that they will sit for mock examination. The answer scripts were collated by the researchers. The responses from the students was marked and scored by English Language experts using NABTEB marking scheme guide. This is because the instrument has been validated by the examination board. The maximum score was one hundred (100) and the minimum score is zero (0).

Data Analysis

Factor analysis was used to determine the assumption of uni-dimensionality of the data as one of the conditions

to be met when applying IRT Framework. The examinees responses were analyzed using IRT statistical software: eirt Item Response Theory Assistant for Excel Version 3.1 for the test items calibration to determine item parameters. The output included Graded Model Item Parameter Estimates; Likelihood-based values of Goodness of Fit Statistics and Item Information Function Value.

Results

The uni-dimensionality of the instrument that was used for data collection in this study is presented below:

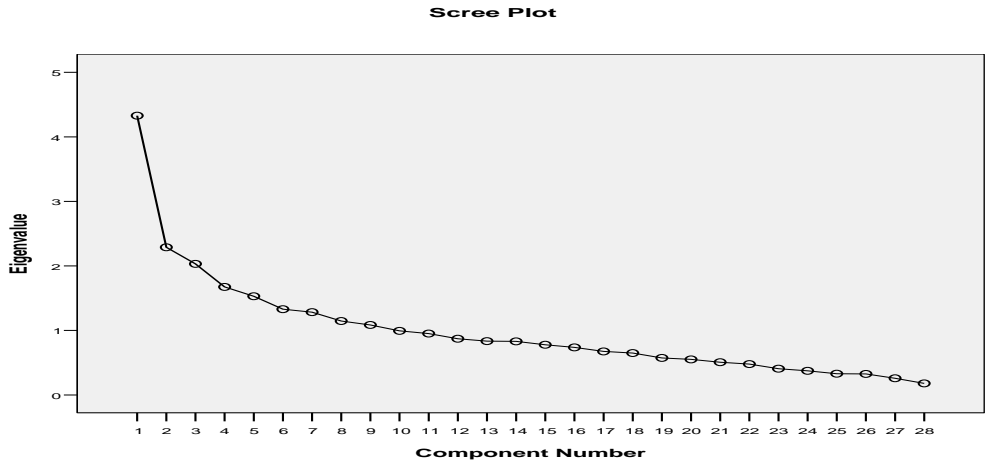


Figure 2: Eigenvalue plot of the components factor

From Figure 2, a clear knee point emerged after the first factor; this is a sign of uni-dimensionality. Also, it could be deduced that the first eigenvalue is substantially greater than the next, that is, there is a dominant factor, the factor structure is deemed to have sufficiently satisfied the assumption of uni-dimensionality hence the instrument/scale satisfied one of the most essential assumption necessary for IRT analysis. This is also in consonance with the submission of Udom (2004); Wiberg, (2004) who affirmed that a dominant factor is a confirmation of uni-dimensionality.

Research Question 1: What is the item difficulty parameter of NABTEB Essay English Language NBC/NTC 2014 Examinations?

This research question was answered using b value of -3 to +3 to select items that fell within this difficulty range from the result obtained from the difficulty estimates obtained from eirt calibration parameter estimates. This is presented in table 1 below:

Table 1:

The Number of Items that Fall Within Acceptable and Unacceptable Difficulty Range Using IRT Framework

| Status of items | Number of Items | Items | % |
|--|-----------------|---|-----|
| Within acceptable difficulty range | 18 | 1a, 5fii, 5fiii, 5fiv,5fv, 6ai, 6aai, 6aiii, 6aiv, 6av, 6b, 6c, 6d, 6e,6fi, 6fii, 7a & 7b | 64% |
| Outside acceptable difficulty range | 10 | 1b, 1c, 1d, 5a, 5b, 5c, 5d, 5ei, 5eii, & 5fi. | 36% |

From table 1, it could be deduced that 18 items representing 64% fall within the acceptable difficulty range while 10 items representing 36% fall outside the acceptable difficulty range.

This research question was answered using positive “a” value to select items that fell within this discrimination range of 0 to +1 from the result obtained from the discrimination estimates obtained from ie iert calibration parameter estimates. This is presented in table 2 below:

Research Question 2: What is the item discrimination parameter of NABTEB Essay English Language NBC/NTC 2014 Examination?

Table 2:

The Number of Items that Fall Within Acceptable and Unacceptable Discrimination Range Using IRT Framework.

| Status of Items | Number of Items | Items | % |
|--------------------------------|-----------------|---|------|
| Positive discrimination | 28 | 1a, 1b, 1c,1d,5a, 5b, 5c, 5d, 5ei, 5eii, 5fi, 5fii, 5fiii, 5fiv,5fv, 6ai, 6aai, 6aiii, 6aiv, 6av, 6b, 6c, 6d, 6e,6fi, 6fii, 7a & 7b | 100% |
| Negative discrimination | Nil | | 0% |

Table 2 shows that all (28) items representing 100% discriminated positively

Research Question 3: What is the test function of the NABTEB Essay English Language NBC/NTC 2014 Examinations?

The test function of the NABTEB Essay English Language NBC/NTC 2014 Examinations is **7.8 at ability of 0.127 θ on ability scale. This is represented in figure 3 below:**

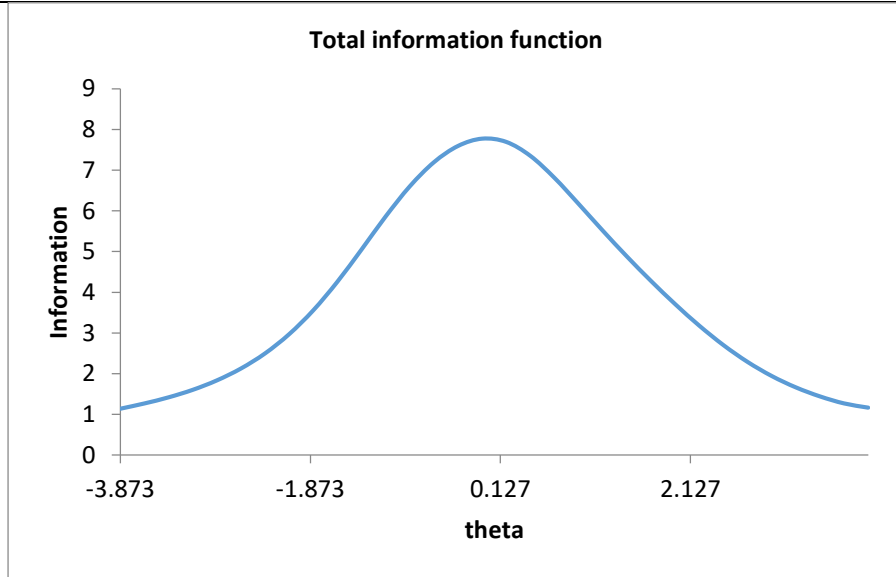


Figure 3 shows the test information function of the NABTEB 2014 Examinations which is the sum of the item information function in any given examination. The maximum information of 7.8 is reached at ability of 0.127 θ on ability scale. In which case, it is indicative of the fact that ability can be estimated with greater precision near this θ (theta) of 0.127. The amount of information decreases steadily as ability tends to the extremes.

Research Question 4: What are the items fit statistics of the NABTEB Essay English Language NBC/NTC 2014 Examinations?

The Goodness of fit statistics obtained using the eirt calibrated parameter estimates showed the items that fitted the graded response model and the items that misfitted the graded response model. This is presented in table 3 below:

Table 3:
The NABTEB Essay English Language NBC/NTC 2014 Items that Statistically Fit or Misfitted the Graded Response Model

| Status of Items | Number of Items | Items | % |
|-----------------|-----------------|---|-----|
| Fit | 26 | 1a, 1b, 1c,1d,5a, 5b, 5c, 5d, 5ei, 5eii, 5fi, 5fii, 5fiii, 5fiv,5fv, 6ai, 6aiii, 6aiv, 6av, 6c, 6d, 6e,6fi, 6fii, 7a & 7b | 93% |
| Misfit | 2 | 6aii & 6b | 7% |

Table 3 shows the result of the chi-square goodness of fit analysis for the NABTEB certificate examinations. It can be deduced that 2 items representing 7% misfitted the graded response model. While 26 items representing 93% fitted the graded response model.

Findings

Ten of the items (36%) fall outside the acceptable difficulty value range. These are items 1b, 1c, 1d, 5a, 5b, 5c, 5d, 5ei, 5eii, & 5fi. While 18 items (64%) are within the acceptable range.

- * All the items (100%) have positive discrimination.
- * The maximum information of 7.8 is reached at ability of 0.127 θ on ability scale which shows that ability can be estimated with greater precision near this θ (theta) of 0.127.
- * 2 items representing 7% do misfit the Graded Response Model while the remaining 93% fit the model.

Discussion of result

The item parameter (difficulty indices) of the NABTEB Essay English Language 2014 examinations. The difficulty value (b parameter) ranges from 0.030 to 21.228, this indicates that the item 5fiii with b value 0.030 is the minimum difficult item while item 1c with b value 21.228 is the maximum difficult item. Although the b value theoretically ranges from $-\infty$ to $+\infty$, typically, the b value from -3 to +3 is used in line with Partchev (2004). The response to research question 1 showed that ten (10) of the items were outside the difficulty range. Partchev (2004) noted that it is traditional to identify location parameter (difficulty) and ability on the same axis and the normal range here is -3 to +3 in that respect, these items (1b, 1c, 1d, 5a, 5b, 5c, 5d, 5ei, 5eii, & 5fi) were found to be outside the recommended level. It is evident that 1c was found to be very difficult for the ability level of the candidates.

The response to research question 2 showed that all the items have positive discrimination representing 100%. Since all the items have positive discrimination, it could be said that the items discriminated quite fairly and positively. This is in consonance with Partchev (2004) who asserted that discrimination indices should be positive except for pathological case where some items have negative discrimination, and further stressed that such items should not be included in a test.

From the test information function of NABTEB 2014 Essay English Language NBC/NTC examinations, it revealed that the test is most informative at 0.127. The test has as much information as 7.8 at that level. It still yields substantially good information at the extreme ends of the scale at -0.4 and -0.0. It means that the test can be used to properly estimate the ability of candidates at that level.

The item fit statistics of NABTEB Essay English Language NBC/NTC 2014 examinations. The test of fit showed that 2 of the items representing 7% were found to misfit the Graded Response Model, while the remaining 26 items (93%) fitted the model.

Conclusion

From the findings it could be deduced that the items were not too difficult for the candidates, though item 1c was found to be very difficult for the ability level of the candidates. It therefore implies that, failure in English Language in this NABTEB 2014 NBC/NTC examinations cannot be accounted for on the ground of the difficulty of the items. Also, failure is not attributable to any inappropriateness of the discrimination parameter. Further, though 2 items: 6aii & 6b misfitted the Graded Response Model, this test was not subjected to IRT Model before use. Due to the benefits inherent in IRT, it calls for NABTEB and other examining bodies to engage the services of Psychometricians and adopt IRT in testing the appropriateness of their items before they are administered.

Recommendations

1. The 10 items that have unacceptable difficulty level need to be carefully reframed.
2. NABTEB and other examining bodies should embrace IRT in item generation, assessment of candidates and analysis of results.
3. Examining bodies should engage the services of measurement experts who are proficient in IRT since IRT is more informative than CTT.
4. There is the need for examining bodies to attend national workshops regularly to keep them abreast of advantages of IRT over CTT and to key into its use. Since it is the globally preferred method of test construction and analysis of results

References

- Adepoju, T.L. (2002). Locational factors ascorrelates of private cost and academic performance of secondary school students in Oyo State, Nigeria. Unpublished Ph.D. Dissertation, University of Ibadan.
- Adeyeye, S.O. (2002). How students' examiners perform at WAEC examinations. *Education and Manpower Vanguard* December, 22
- Amasingha, K. J. (2015). Item response theory: The way forward to objectivity in educational measurement in the school system. *Nigerian journal of educational research and evaluation*. 14, (2)(17).
- Baker, F. B. (1992). *Item response theory: Parameter estimation techniques*. New York, NY: Marcel Dekker.
- Baker, F. B. (2001). *Item response theory: Parameter estimation techniques*. New York, NY: Marcel Dekker.
- Duong, M. (2004). Introduction to item response theory and its applications. *Research Development Paper: CEP 900: Pro-seminar in Learning, Technology, and Culture (Fall 2004)*. London and Philadelphia: Kogan Page.
- Embretson, S.E., and Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Fakeye, D. O. (2012). Teacher's Qualification and Subject Mastery as Predictors of Achievement in English Language. *Global Journal of Human Social Science* 12 (3): From <http://creativecommons.org.com>.
- Hamzeh, D. (2004). The Relationship Between Item Parameters and Item Fit. *Journal of Educational Measurement*. 41 (3) pp 261-270
- Joshua, M. T. (2014). Applications Of Item Response Theory (IRT) In Testing & Associated Challenges (In Africa). Paper presented at IAEA Scholarship Training on Educational and Evaluation. Abuja, Nigeria
- Labo-Popoola S.O (2002). Teacher and School Variables as determinants of students Achievement in Comprehension and Summary Writing aspects of English Language. Anunpublished Ph. D Dissertation in Institute of Education. University of Ibadan, Ibadan.
- Lord, F. M. (1986). *Application of item response theory to practical testing problem*. Mahwah, NJ: Lawrence Erlbaum Associate, Inc.
- Mislevy, R. J. (2011). Evidence –centered designed for simulation-based assessment. (CRESST Report 800). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standard and Student Testing (CRESST).
- Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika*,51, 177-195.
- Motshabi, K., Monamodi, K., and Mompoloki, G. (2011). Best Practices in Educational Assessment For Regional, 29th Conference of the Association for Educational Assessment in Africa (AEAA) Hosted by Kenya National Examinations Council from 1st – 5th August 2011. Laico Regency Hotel, Nairobi.

- National Business and Technical Examinations Board (2015). Statistical Digest
- Obemeata J. O. (1995). Education: An Unprofitable Industry in Nigeria. A postgraduate School Interdisciplinary Research Discourse in University of Ibadan, Ibadan.
- Ojerinde, D. , Ojo, F. R. and Popoola, O.O. (2012). *Introduction to item response theory, parameter models, estimation & application* Goshen printer media Limited, Lagos, Nigeria.
- Onipepe, H. (2003). National Development Hinges on Quality Education. The Comet, Thursday, 2nd January, pp: 21.
- Onukaogu, C. E. (1994). Teacher effectiveness as a factor in the use of English programme In Nigerian University. Ph.D. Dissertation submitted to University of Ibadan.
- Partchev, I. (2004). A visual guide to item response theory. Jena: Fredrick Schiller- University.
- Reyes, J.M. (2010). Assessing invariance of factor structures and Polytomous item response model parameter estimates (Doctoral Dissertation):Texas A & M University
- Reynolds, T., Perkins, K., and Bruten, S. (1994). A comparative item analysis study of a language testing instrument. *Applied Psychological Measurement*, 11, 1-13.
- Samejima, F. (1969). A general model for free- response data. *Psychometrika Monograph Supplement*, 18: 220 - 227
- Udom, S. U. (2004). Item response dimensionality of West African Examination Council(WAEC) objective mathematics test and students performance. Unpublished M.Ed. Thesis. University of Calabar, Calabar, Nigeria.
- Wiberg, M. (2004). Classical test theory vs. item response theory: An evaluation of the theory test in the Swedish driving license test. Sweden: Umea' University.
- Zwinderman, A. H., and van den Wollenberg, A. L. (1990). Robustness of marginal maximum likelihood estimation in the Rasch model. *Applied Psychological Measurement*, 14, 73- 81.

Using an Online Query Tracking System to Track Examination Queries in Zambia

By

Banji Milumbe Shakubanza

bms Shakubanza@exams-council.org.zm

and

Mice Malembo- mhatembo@exams-council.org.zm

Abstract

Tracking systems have been used in various fields, especially in supply chain management to improve efficiency. This concept of tracking items can be applied by examining bodies that deal with different clientele in examination related services such as replacing lost documents, verification and translation of results among others. The primary purpose of this paper is to show how a tracking system would improve efficiency and customer satisfaction in delivering different services related to examinations. A system for tracking queries at Examinations Council of Zambia (ECZ) was developed and piloted internally. A simulation model based on data from the manual system which accounts for different types of services requested for by the clients was created to show how the concept of an online tracking system can be used by an examining body. Different services were logged into the system and their status was tracked. The results indicated that the query tracking simulation model can be used as a tool to help improve efficient and effective service delivery to clients. The system would also resolve the challenges of overcrowding, long distance travels by clients, reduction on use of paper, and loss of documents.

Keywords: *Tracking, Efficiency, online systems, Zambia, examination queries, examinations, tracking queries*

Introduction

Information technology is often used to support the response or to provide the response itself (Turban, 2004). Using information technology has become a new way of conducting business in the 21st Century. Information Technology aims at increasing productivity, reaching new customers and sharing knowledge across institutions for competitive advantage.

The query tracking system will change how business is done at ECZ from the traditional way of receiving queries manually to using online systems. The query tracking system has borrowed the processes used in the supply chain management to manage various queries or services offered by ECZ to clients. The ECZ has endeavoured to be at a front line in the use of technology by looking ahead

at opportunities to improve its operations and provision of services to its clients. The internet has provided a cheap and accessible platform upon which ECZ is building systems to ease or improve its services to the public.

As an examining body, the Examinations Council of Zambia deals with various queries related to results way after the results have been released. The clientele includes examination centres, higher institutions of learning, current and prospective candidates, parents and organisations or employers. To access services, clients have to physically travel to ECZ to submit manual records which in some cases are misplaced or take too long to be resolved as there is no way of identifying which queries have not been attended to. A recent 2017 first and second quarter internal audit report indicated that the amount of money paid by clients for various services did not match

with the the actual services by ECZ. This means that more clients paid to ECZ for various services which they did not receive on time. This is difficult to state which queries were attended to and those that were not as there is no proper system of monitoring received queries. In addition, there is too much congestion of people who are trying to access the services and some have to travel long distances in order to access services thereby spending a lot of money over and above the fees for the required service.

It is for this reason that the Query Tracking System (QTS) was developed to reduce congestion, track the status of

each query online, and be able to check how long each query could have taken before being resolved. Having such a system would ensure accountability and quick response to the clients.

Theoretical framework (input Process output)

The Input Process Output (IPO) Model (Subiyakto, Ahlan, 2014) was used in piloting the tracking system. There are critical inputs that are required to attain accurate information which results in a well-managed examinations query resolution process. These include online systems (QTS), adequate ICT equipment and well trained staff.



Figure 1: Theoretical Framework – Input, Process, Output Model

In the QTS, the various types of services / inputs such as replacement of lost / damaged documents like certificates or statement of results, name amendments, verification of results, translation or equating of results, unprocessed results, remarking of scripts, certificate replacement, and many others comprise the inputs to the system.

The data inputs that are described above are converted into meaningful form which is called processing. The ECZ main computers store these data items and process them to provide customers with solutions to their queries. This would include correcting names of candidates, replacing the actual certificate which would require retrieving the old results and printing a new certificate are.

The output in the QTS comprise amended names, amended results slips from name correction, resolution of unprocessed results or hardcopy of the certificate, translated results or equated results etc. This information is given to the user either as printed copies or displayed on the computer screen when a customer queries or enters the tracking ID.

Review of Supply Chain management

Some literature was reviewed on how the supply chain management works. This was used to help understand the processes involved and then come up with a similar concept for the examinations queries tracking system.

Turban, 2004 describes that a supply chain can be broken into three major parts:-

(i) The upstream supply chain – which involves activities of a manufacturing company with its first tier.

This can be likened to ECZ clients with queries, they are the originators of these queries which need to be resolved by ECZ.

(ii) The internal supply chain – In-house processes used in transforming the inputs received from suppliers into the organisation's outputs.

The internal processes undertaken by ECZ to resolve queries received from the clients.

(iii) The downstream supply chain – are activities involved in delivering the products to the final customers. In the ECZ scenario, this involves giving feedback to the clients once the query is resolved.

Laudon & Laudon 2010 describes a parcel tracking system that is used by United Parcel Services (UPS) to track parcels in transit up to the time it is received. Various inputs were described, the processing and the final outputs of the system.

Methods / Procedures

Research Design

The design of the study involved building a web based query tracking system while the piloting approach was used to test the developed system. The pilot was conducted internally at ECZ with staff responsible for receiving, processing and giving out the final answer or solution of the query to the clients. A total of **ten (10)** members of staff who were involved in the pilot were purposively sampled since they were directly responsible for handling and processing the various queries received from the clients.

Application design – query tracking system

The Query Tracking System (QTS) was developed using XHTML, JavaScript, CSS, PHP, Apache Webserver and MySQL. The online query tracking system will be used by clients who require various services from ECZ as well

as the ECZ staff who are responsible for resolving these queries. Both the ECZ staff and the clients will be able to track queries submitted online and follow up on those that take longer than necessary to be resolved.

A software application was developed by ECZ called the Query Tracking System (QTS), a web-based query tracking system that allows clients to log in the system, submit their query and attach accompanying documents and then check the status of their query through the ECZ Portal.

When a Client submits a query using the QTS, the query is sent to the ECZ webserver and information about the query and client is sent to the ECZ staff alerting them of a new query submitted. A receipt notification is sent to the client confirming that the query has been received at ECZ and a tracking ID assigned. The client is advised of the tracking ID which they will use to check the status of the query in the comfort of their homes.

Piloting the Query Tracking System

A web based Query Tracking System was developed in order to improve service delivery to clients who require post examination services at ECZ and was piloted internally at ECZ. The pilot period was for two (2) months to ensure that even the ECZ staff that were going to use the system were trained and very well acquainted in using the QTS.

Sample of the pilot

The internal piloting of the system targeted ECZ staff that were attending to clients and involved in the resolution of the various queries that are submitted. Of the staff involved in the pilot, some were working from the ECZ service centre while the others were at the ECZ main office.

Pilot Results

The following are the findings, lesson learnt and also recommendations on the feasibility of the implementation of the QTS.

Functionality

In terms of functionality, it was ascertained that the staff who participated in the pilot were able to use the system

successfully and also proposed changes such as inclusion of a facility to attach documents and bank deposit slips. This proposal were accepted and included in the QTS during the piloting period.

Access to the Query Tracking System

All the 10 staff that were involved in the pilot were able to access the Query Tracking System, submit queries online, upload supporting documents, like bank deposit slips and also test the reports.

Tools used for the Pilot

During the pilot, the ECZ internet was used with the staff’s personal computers. The manual records were

used for piloting and were used as input to the QTS. The same staff were used as clients so that they could be able to receive notifications as the queries were being processed. It can therefore be concluded that the online tacking system is indeed a viable solution which will greatly improve the efficiency in the tracking of queries online.

The Query Tracking System

A representation of the QTS is presented in the screen shots below to show how the various queries are received and handled in the system. The first part of the system is to launch a query by selecting the type of service a client requires as shown in Figure 2.

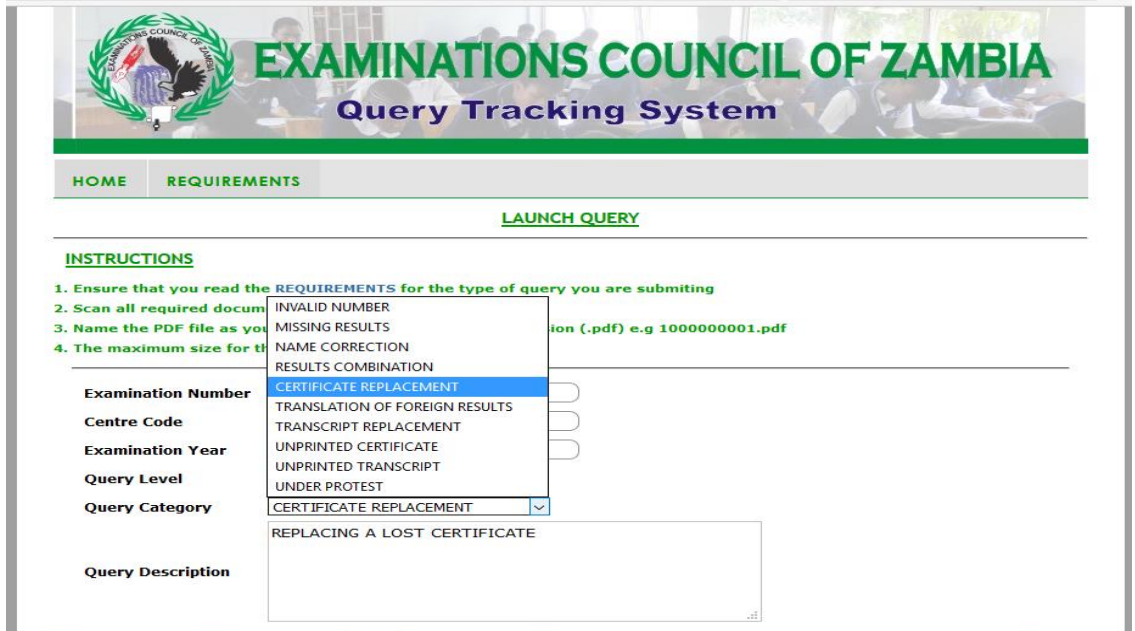


Figure 2: Lauching a Query

Notification after successful launch of a query

The client receives a notification and a tracking ID after submitting the query successfully. Figure 3 shows the message that the client receives together with the assigned tracking ID of their query.

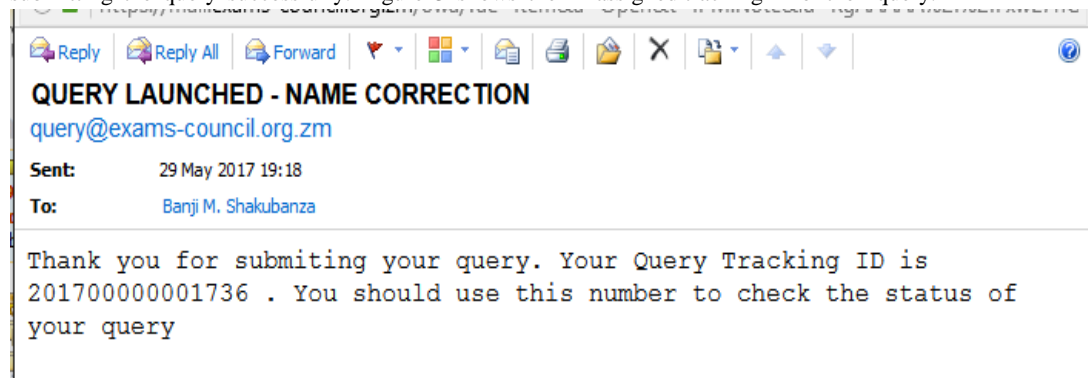


Figure 3: Notification of query launched

Notifications during the process

The client receives email notifications during the resolution of the query. As the resolution of the query is being updated, the client is notified to check on the

updated status of the query until it is resolved and ready for collection of the document if need be. Figure 4 is a snapshot of the message that is sent to the client.

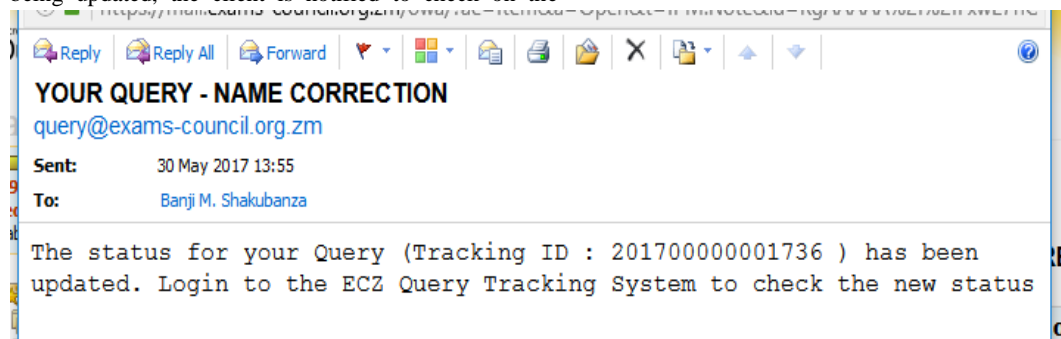


Figure 4: Notification on updated query

Discussion

The query tracking system was piloted internally by the ECZ staff and the results of the pilot revealed that the QTS creates value for ECZ clients as well as ECZ itself in a number of ways;

Efficiency and Effectiveness

From the results, the participants indicated that the QTS facilitates better communication and collaboration between ECZ and Clients thereby improving efficiency and effectiveness in the provision of services to the public. In addition, the QTS promotes efficiency by reducing costs and speeding up the cycle time.

Participants indicated that it is beneficial to the clients as it offers flexibility and they would not need to travel to ECZ offices to submit their queries. The client would not spend transport money unnecessarily except where there is need to collect the final product in person. Furthermore, the QTS lessens work, saves on time and cuts down on physical movement of people and documents as well as reducing transport related costs by the client.

Reducing Congestion

The participants indicated that using the QTS would reduce congestion at the ECZ offices as clients would only visit the office to collect the final product if need arises. The long queues that are experienced during peak periods would be eliminated.

Improved Record keeping

The pilot results showed that there would be good record keeping as all the queries are electronically kept and makes it easy to make a followup on unresolved queries. There would be no loss of documents as they are all electronically kept and the problems of manual systems that are prone to errors would be eliminated.

Ease of Use

The participants found the application to be very user friendly, easy to use, and appreciated the system very much. All the participants easily learnts how to use the QTS and found it very convenient to use. The major benefit that was envisaged was that the mistakes that come with manual systems will come to an end.

Reporting

Participants observed that reporting would be easy as the QTS provides several report by the click of the button which are not possible with the manual system.

Conclusion

From the pilot study that was conducted, it was discovered that in the digital economy, organisations and examining bodies in particular will transform themselves

from traditional modes of operation to e-business depending on how well they will adapt their structure and processes to take advantage of emerging technologies and what architecture and infrastructure they will use. Some of the motivations that contribute to the development of many web-based systems nowadays are because of its efficiency in handling rapid access of documents and its ability in supporting multiusers simultaneously, thus saving a lot of time and hassle free. Othman, M., Ismail

S. N. & Raus, M. I. (2009). Based on the results obtained from the pilot exercise conducted, it can be concluded that the QTS worked as expected and since there are already other online systems like the online candidate registration that have been implemented by ECZ, then the same internet facilities being used currently are good enough for the implementation of the QTS Application. The QTS is, therefore, expected to be used by clients from any part of the country and any part of the world. Clients living outside the country will be able to access the services with ease.

References

- Henry, L.C (2001). *Information Technology for Management*. New Delhi: Tata McGraw-Hill Publishing Company Limited.
- Laudon, K.P. & Laudon, J.P. (2010). *Management Information Systems: Managing the digital firm (13th Ed.)*. Upper Saddle River, New Jersey, 07458: Pearson Prentice Hall.
- Laudon, K.P. & Laudon, J.P. (2007). *Essentials of Business Information Systems (7th Ed.)*. Upper Saddle River, New Jersey, 07458: Pearson Prentice Hall.
- Othman, M., Ismail S. N. & Raus, M. I. (2009, October), The development of the web-based Attendance Register System (ARS) for higher academic institution: From feasibility study to the design

phase, *International Journal of Computer Science and Network Security*, 9(10), 203-208.

- Subiyakto, A., & Ahlan, A. R. (2014, July). Implementation of Input-Process-Output Model for *Measuring Information System Project Success*. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, 12(7), 5603-5612.
- Turban, E., Mclean, E. & Wetherbe, J. (2004). *Information Technology for Management: Transforming Organisations in the Digital Economy* (4th ed. Wiley Student Edition). New Delhi, India: John Wiley & Sons.