The 2014 International Association for Educational Assessment Conference (IAEA 2014) in Singapore, May, 2014.

Theme: **Assessment Innovations for 21st Century.**

Subtheme: **Standards Setting and Large Scale Assessment.**

Title: **Development and use of multiple choice items as instrument for Science performance measure for large class sizes.**

Dr Ursula N. Akanwa[1]
sisngoakanwa@yahoo.com;
Department of Science Education;
College of Agricultural & Science Education
Michael Okpara University of Agriculture; Umudike
P. M. B. 7267; Umuahia, Abia State; Nigeria

Dr Charity A. Okonkwo[2]
caeokonkwo@yahoo.co.uk
School of Education;
National Open University of Nigeria
14/16 Ahmadu Bello Way, Victoria Island
P.M.B. 80067; Victoria Island, Lagos, Nigeria

Abstract

"Education for All" and the attainment of the Millennium Development Goals have placed pressure on carrying capacity of Nigeria's educational system, resulting in significant increase in the number of children in her educational system. This led to very large class sizes and the inherent challenge of ensuring that the learning goals and quality of education are sustained. Provision of fair assessment with timely feedback for students is a difficult task with science laboratory classes, hence, the use of subjective assessment. This study sought to develop and use multiple choice items as new instrument for science performance measure for large classes with a view of automating the process if the results are worthwhile. Biology (a natural science) and Physics (an applied science) were used for this study. First year students of Michael Okpara University of Agriculture, Umudike offering 100 level courses in Biology and Physics Laboratory courses of the University's College of Natural and Applied Science are the population while the sample consists of the 100 level students of College of Agricultural and Science Education offering the required laboratory courses. The practical work for the identified courses will be simultaneously assessed by the College subjective mode of assessment while the researchers will also assess the same courses using the developed multiple choice items targeting goals of the subjective assessment measure. The content validity of the test items were determined with test blue prints while the reliability of 0.81 was established using Pearson product moment correlation coefficient. SPSS 21 was used to carry out the data analysis. Pearson product moment correlation was used to determine the coefficient of correlation between the laboratory subjective examination scores and the multiple choice scores and paired t-test was used to test mean differences. Results showed a significant coefficient of correlation and a significant different between the means as revealed by the t-test result. Since there was a significant correlation coefficient the process involved will be proffered for adoption and implementation to the university and other institutions challenged by the assessment of practical work of large classes.]

Keywords: fair assessment, science laboratory classes, large class sizes, multiple choice items, Biology practical, physics practical, new performance measure, automation.

Introduction

"Education for All" and the attainment of the Millennium Development Goals have resulted in rapid expansion in Nigeria's higher education. Students' numbers have grown considerable in many courses and subjects especially at the undergraduate levels. Large classes pose significant teaching challenges not least in the assessment of student learning especially for laboratory classes because there has not been commensurate expansion of facilities especially human resources. Hence, the Nigerian science teacher is faced with overwhelming task of handling large classes especially in laboratory practical activities. The student cohort is becoming ever more diverse and more active, and student class sizes are increasing. Classes of more than two hundred students are becoming more common. The class size presents real challenges to the design, management and fairness of assessment practices.

Globally, the student cohort is becoming ever more diverse and more active, and student class sizes are increasing. For some faculties, classes of 1000 – 2000 are becoming more common; for others a class of 80 would seem to be a large class; doubling or tripling class numbers can present real challenges in the design, management and fair assessment practices especially when student – centered learning is the University expectation (Strong, 2013). This global trend is the case with Nigerian university system especially for science laboratory classes at 100 and 200 levels where students do general science courses. These large classes present assessment challenges for teachers and for learners. Some of the challenges already identified in literature (Strong, 2013:1) are also applicable to Nigerian educational system. These have issues for both staff and students.

Issues for staff

- Engaging students and encouraging deep learning
- Difficulty giving high quality, individual feedback
- Ensuring consistency in marking
- Heavy workload in managing assessment
- Assessing graduate attributes
- Avoiding plagiarism

Issues for students

- Feeling included in the discipline, relevance of tasks to program
- Inadequate opportunities for feedback, and timing of feedback
- Unclear about assessment methods and marking criteria/standards
- Inadequate support for completing assessment tasks
- Lack of diversity in assessment tasks
- Managing assessment workload

Strong (2013) observed that some of the issues arose from the fact that teachers tend to use traditional assessment methods – such as examinations and lengthy written assessments in large classes. The National Policy on Education (FME, 2004) defines class size as the population of a given class in terms of number of students and recommends an average class size of forty learners to a teacher whereas, class size in this study surpassed two hundred students in a laboratory class activity. Thus, in response to the pressure and challenges of assessing large groups of students, academics are responding through:

- Greater attention to the communication of clear assessment criteria to students;
- The development and use of marking guides to be used by teaching and assessment teams;

- The increase of various forms of exemplars to guide student effort – as well as guide marking and grading – including the modeling of disciplines – based thinking, writing and performance; and
- The continuous refinement and dissemination of assessment policy and practice in relation to large student groups.
  (Australian University Teaching Committee, 2002)

Recent rises in students' numbers have led the authors to have a re-think on the mode of assessment for laboratory practical works and the observation made by Strong (2013) that some of the issues arose from the fact that teachers tend to use traditional assessment methods – such as examinations and lengthy written assessments in large classes. Marking large number of students' laboratory work exercises subjectively is time – consuming, labour intensive and prone to errors of consistency (Newstead and Dennis, 1994). Thus the use of multiple choice items will allow for more efficient examination of students because the questions can be marked rapidly using computers, optical mark reader or staff members with no special knowledge in the area being assessed. It also eliminates the need for double marking, thereby saving time after the examination process. Besides, it enables the lecturer to test a wide range of topics in a single assessment.

Nowadays, objective testing is widespread. It is used for widespread, large scale national objective tests. For instance, it is used as the American College Testing examination (ACT), the Scholastic Aptitude Tests (SAT), the Graduate Record Examinations (GRE). While the Law Society Admissions Test (LSAT), are used as performance indicators for all students seeking admission to undergraduate and postgraduate courses (Good Practice Guide in Question and Test Design. These tests are increasingly delivered via computer networks (http://www.ets.org). In Nigeria, the Joint Admission and Matriculation Board (JAMB) also uses objective test as performance indicators for admission of students into Nigerian higher institutions of learning. Many universities now operate widespread objective testing in a range of departments. The National Open University of Nigeria (NOUN) uses objective testing for formative assessment of learners at all levels of undergraduate programmes and is also used for summative assessment at 100 and 200 levels of the undergraduate programmes because of the large class size. In the United Kingdom, use of objective testing for formative and summative examinations is increasingly popular. The Open University of United Kingdom (OUUK) use large scale testing across a range of disciplines. UK Medical schools, including University of Birmingham, University of Bristol and University of Dundee also use objective testing. These institutions employ objective testing most of which involves computers or optical mark readers.

The authors of this paper are interested in the use of multiple choice items for assessing laboratory practical exercises in Nigerian Universities at 100 levels because of the large class sizes involved so as to cue – in to the advantages of using this test format over the subjective format hitherto used. These advantages are:

- Significant reduction in marking time;
- Wider coverage of topic content;
- Easy of analysis of individual questions;
- Prompt provision of feedback to students and teachers; and
- Pre-testing of items in order to evaluate their effectiveness and level of difficulty.

The authors sort to develop and use multiple choice items as new instrument for science performance measure for large classes with a view to automating the process if the results are worthwhile. The success of this study will lead to possible deviation from the norm – use of traditional approach of examination with lengthy subjective assessment items; and the benefits therein. Moreover, multiple choice items have often been used successfully to assess learning in large classes, and is sometimes the most efficient way to conduct assessment (Strong, 2013). With careful design, multiple choice items could be used to assess deep learning; diagnose students' learning difficulties and give valid and reliable results (Okonkwo 2008; 2005). Use of technology for automating the process of assessment on the other hand would help manage the assessment process (Okonkwo 2010a, 2010b; 2011 and Okonkwo and Ikpe, 2008). The National Open University of Nigeria (NOUN) uses technology in the assessment of students learning outcomes in the University. The Institution has computerized formative assessment for undergraduates, postgraduate diplomas and for masters' programmes. Technology is also successfully used in her summative assessment of learners at the 100 and 200 levels of undergraduate programmes successfully. It should be noted that item writing for multiple choice items demands training of the items writers to ensure quality and good spread of the items across all difficulty levels of the domains of Blooms (1956) taxonomy of educational objectives.

Concept of Assessment

Assessment according to Brookhart (1999;9) means to gather and interpret information about students' achievement, and achievement means the level of attainment of learning goals set for a course. His notion that good assessment yields good information about the result of the instruction; and it is itself a necessary component of instruction is supported by the authors. Thus, accordingly, assessments are an instructor's way of gathering information about what students have learned, which could be used to make important decisions about students and their learning materials. Hence, it is important that student's assessment of science laboratory practical activities give dependable information – meaningful and accurate. That is, the results of assessment should be valid and reliable in order to serve as indicators of the particular learning goals for the courses being assessed for the results/scores to accurately represent the levels of achievement attained by the learners. Therefore, assessment of students' laboratory work in higher education is important because important consequences follow from it. This calls for the use of valid and reliable instrument for assessment of the learning outcomes.

It is necessary to ensure that whatever assessment used is really congruent with whatever objectives for teaching are. It is not desirable to teach a whole bit of a course and the not have them examined as is always done with the use of subjective assessment such as essay test items. Because, even though students are expected to be intrinsically interested in the learning content but they can easily slide out of the content that they perceived will not be examined as a result of human nature. Hence, course development and assessment should be synonymous with each other. It should not be a process in which academics set out to design courses and then the assessment is the add-on, where suddenly marks have to be provided as evidence of teaching of the courses. The Cross Sectoral Assessment Working Party (2011:7) opined that "Assessment of and for students' learning is the process of gathering and analyzing information as evidence about what students know, can do and understand. It is part of the ongoing cycle that includes planning, documenting and evaluating students' learning". Thus, assessment should be regarded as a key professional skill for teachers; part of effective planning for teaching and learning; and recognized as central to classroom practice.

On the other hand, performance based assessment is the process of using students' activities rather than tests or surveys to assess skills and knowledge. It is supposed to be used when activities can be linked directly to the course; in academic programmes that develop complex integrated skills whose focus is on the creation of products or performances as in science oriented programmes. Therefore, they are not ideal in the large class sizes currently challenging higher institutions due to limitations of time and/or scope of courage; and as such other assessment techniques such as tests can better serve the needs of the assessment. The authors nevertheless advocate that performance assessment of the students should continue to be taken care of by building on their daily laboratory activities as formative assessment measures. This is necessary to enable the faculty to determine students' skills and abilities and for students to learn more about how to improve their own skills; and to link their teaching to desired learning outcomes. While the multiple choice items should be used for summative assessment

Teachers should realize that assessment is the driving force behind student learning (Beevers, Cherry, Foster & McGuire, 1991; Brown, Bull, & Pendlebury, 1997). It is a common practice for students not to read additional materials made available to them to practice. They most often do not read such supplementary materials because such materials are rarely assessed due to marking load an indication that students are becoming more strategic in their study habits and they are reluctant to undertake work which does not count towards their final grade. However, computer aided assessment (CAA) offers the opportunity to test students more regularly without increasing the marking load on staff. Computer aided assessment lends itself to: use of multiple choice items to help students diagnose their strength and weaknesses in their courses; the result provides feedback to teachers and students as they work through the learning materials and practical activities; a sequence of tests taken throughout a course of study with all the relevant feedback leads to course improvement, and assessment is always a learning opportunity for students. Hence, a good assessment system must ensure that students use this opportunity for effective learning.

Assessing with Multiple Choice Items

Test must be carefully constructed in order to avoid the contextualization of knowledge (Paxton, 1998) and it is wise to use objective testing as only one of a variety of assessment methods within a module. However, in terms of growing students' numbers and decreasing resources, objective testing can offer a viable addition to the range of assessment types available to a teacher. Multiple choice items are forms of Objective Test Items. This form of test items requires the user to choose a response to an item whose correct response is pre-determined. Multiple choice items are well suited for CAA that involves automated marking. The electronic marking of the responses is completely non-subjective because no judgment has to be made on the correctness or otherwise of the answer at the time of marking. It should be noted that the items are as objective as the test designer makes it. Multiple choice items are strongly associated with assessing lower order cognition such as recall of discrete facts. Because of this, their use in higher education has been questioned (Velan, James, McNeil & Kimar (2008). Nevertheless, multiple choice items can be designed to assess higher order cognition (such as synthesis, creative thinking and problem solving). Some educationist (McBeath, 1992; Okonkwo, 2005) have surmised that all six levels can be tested using multiple test items. These researchers have developed valid and reliable objective test items at different levels of Bloom's taxonomy. But the questions must be drafted with considerable skill if such items are to be valid and reliable. This needs training and takes time. When multiple choice items are to be used for assessment purposes, either formatively and/or summatively, for easy of marking them and/or for solid educational reasons, they are to be integrated effectively into the assessment design as we have done in this study. Multiple choice items can aid teaching and learning in the following ways already identified (UNSW Australia, 2013):

- Providing students with rapid feedback on their learning;
- Being continually available without increasing the marking load (if delivered online, with automated feedback);
- Lending themselves to design using quiz software tool either within or independently of Learning Management Systems such as Blackboard or Moodle. With such software, multiple choice items presentation can be automated and it facilitates test administration, scoring and feedback provision.
- Allowing objective scoring. There can only be one right answer to a well-designed item, so marker bias is eliminated.
- Allowing scoring by anyone, even automatically, thereby increasing efficiency, particularly in teaching large group of learners.
- Being immune to students' diverse capabilities as writers.
- Containing recyclable items. Across the discipline, test writers can progressively develop and accumulate items in pools or banks for re-use in different combinations and settings.

Nevertheless, multiple test items have challenges. Among the challenges identified by UNSW Austratia (2013) are that:

- They are time consuming to develop and require skill and expertise to design well.
- They are generally acknowledged to be poor at testing higher order cognition such as synthesis, creative thinking and problem solving.
- They can be answered correctly by guessing. If poorly designed, they can provide clues to encourage guessing.

Overview of Science Practical Work

Practical work is the teaching of practical skills (Cilliers, Basson, Kirchner and Rutherford, 2000). Researchers like Bradley & Meaker (1998) have posed serious questions about the cost effectiveness and purpose of practical work. While, Motlhabane (2013) have observed that there has been global efforts to reduce cost involved in the manufacturing and distribution of science equipments in schools. The initiative has been welcomed by many teachers according to Motlhabane because the usual observed practice has been:

- Lack of equipment previously in some schools;
- Lack of proper use of equipment even when available in schools;
- Inadequate time for laboratory activities;
- Lack of trained personnel to handle the equipment;
- Inadequate number of equipment compared to class sizes

One of or combination of these usual practice has resulted in demonstration of science practical to group of learners by teachers of science. Therefore, learners observe what the teacher is doing and if the group is too big, they will not be able to see what is happening because of the size of the equipments. In order words, learners in most cases will not be able to manipulate equipments and participate actively in the actual doing parts. But, by implication, science practical work should generally incorporate three components namely:

1. Learners must “hear” what is happening.
2. They must “do” the actual science practical.
3. They must “see” what is happening.

These components are necessary to make practical activities complete. However, in practice, due to large class sizes, lack of proper training of teachers to handle the equipments and/or even proper use of equipment have led to teachers theorizing practical activities. Hence, it is better to use well designed and articulated interactive computer programmes and simulations to teach science practical. The teaching will be simultaneously examined by the use of multiple choice items. Then the teacher will be able to guide the learners from the feedback obtained from the test results. This process will enhance active participation of the learners and is capable of taking care of the lapses inherent in the current practice.

In this study, science laboratory classes mean hands – on as well as minds – on practical work activities such as laboratory experiments. With hands-on activities learners do experiments on their own, while the teachers act as facilitators. But, in distance education, learners can engage in distance practical work. In which case, learners can observe practical work via video or television without hands-on participation (Motlhabane, 2013). In the case of demonstrations, learners can observe a practical being demonstrated by the teacher in the classroom. Learners can also participate by making, doing, measuring, observing, asking and answering questions (Bradley and Maake, 1998:4). Thus, in that case, well designed multiple choice items such as ours, coupled with video demonstration of experiments and/or simulations can adequately serve the purpose of practical activity and assessment for large class sizes. Since, in general, practical work can include all types of investigations or experimentations by learners on their own or in groups, as well as demonstrations by teachers (Van der Linde et al, 1994:49). Furthermore, a review of research in science teaching (Shulman and Tamir’s, 1973) identified the rationales generally advanced for the use of the laboratory in science teaching as:

- The subject matter of science is highly complex and abstract;
- Students need to participate in enquiry to appreciate the spirit and methods of science; and
- Practical work is intrinsically interesting to students.

Hence, the advocates of the practical work were to train students in the way of practicing scientists so that students could become good scientists in the future (Abimbola, 1994). Abimbola (1994:5) opined that “continuity to accord a central role to laboratory work in science teaching does not seem reasonable and feasible anymore in the developing countries”. He suggested recording on video tape well-planned demonstration experiments that can be showed to students at appropriate times. The practice would save teachers and administrators some money without further expenses on many consumable items; and modern day students are likely to enjoy watching a video recording that carry out laboratory work. He noted that most traditional laboratory activities are gradually being banned in developed countries either because of their health hazards or special interest groups.

A typical example is animal dissection that used to be the core of biological experiments being gradually phased out because of the influence of animal right activists (Abimbola, 1994), despite the spirited defense of the animal use in the guidelines issued by the National Association of Biology Teachers (1980). Alternatively, computer simulations of dissection experiments interactive computer activities could be used in place of real life laboratory experiments; while computer programmes for problem-solving exercises could be used in the major sciences. Already, the West African Examination Council are using alternative to practical for Senior School Certificate Examination for private candidates for sciences in Nigeria. Therefore, the authors of this paper are proffering is by extension, practicable considering the fact that the students are to be already exposed to practical activities either in real time or by alternative and modern approaches – simulations and interactive computer programmes. Also, according to Abimbola (1996),

> The objective stated for any unit of instruction has been found to be generally in the ratio of the cognitive domain, 50%; the psychomotor domain, 25% and the affective domain 25%. If this proportion is so, and if students are able to master the cognitive component of their lessons, it should have transfer value on the affective and psychomotor domains.

Thus, Abimbola and Danmole (1995) have recommended the use of content analysis method by concept maps to help students to understand the conceptual knowledge in science. Since, according to Bloom (1956), most of the subject matter content of most disciplines is informational. Whereas, we are proffering the use of multiple choice items as new instrument for science performance measure for large classes with a view of automating the process. This assessment paradigm will complement simulation and use of computer interactive programmes in science classrooms for large classes as worthwhile approach for science teaching and learning. This approach if well planned and executed, will inform good practice for standard setting and large scale assessment.

Research Questions

1. Is there a correlation between students performance in the Physics multiple choice test items and their performance in the subjective test items?
2. Is there a correlation between students performance in the Biology multiple choice test items and their performance in the subjective test items?

Hypotheses

1. There is no significant relationship between the means of multiple choice scores and the subjective scores from examinations in Biology.
2. A significant relationship does not exist between the means of multiple choice scores and the subjective scores from examinations in Physics.

Methodology

The design of the study is a correlation design. The population of the study consists of all first year students of Michael Okpara University of Agriculture, Umudike offering 100 level courses in Biology and Physics Laboratory courses of the University's College of Natural and Applied Science while the sample consists of the 100 level students of College of Agricultural and Science Education offering the required laboratory courses. The sample consists of all first year students in the college of Agricultural and Science Education of Michael Okpara University, Umudike who offered, PHY 117 (Practical Physics) and BIO 111 (Introduction to Biology) the practical aspect. The total number that offered Physics was 83 and Biology was 134. This is because integrated science/Education and Biology/Education students did not offer Practical Physics (PHY 117). The instruments for data collection were the subjective practical examinations conducted by the College of Natural Science and Applied Science for biology (BIO 111) and Physics (PHY 117) . A fifty items multiple choice questions were developed for Biology and Physics by the researchers using the same course outline used to teach and examine the said students. The instruments were content validated using test blue prints and test-retest on some other students from Colleges not part of the study was used to establish the reliability of the test items. Pearson product moment correlation coefficient was used to determine reliability index. It was found to be 0.81. Data were analyzed with SPSS 21. The coefficient of correlation and the mean differences were determined using Pearson product moment correlation coefficient and paired t-test.

Results:

Table 1: Correlation Coefficients of Biology

		A	B
Exam	Pearson Correlation	1	.531**
	Sig. (2-tailed)		.000
	N	134	134
MCQ	Pearson Correlation	.531**	1
	Sig. (2-tailed)	.000	
	N	134	134

**. Correlation is significant at the 0.01 level (2-tailed).

Table 1 shows 0.531 Correlation Coefficient which is significant at 0.01 level of significance for Biology.

Table 2: Correlation Coefficient of Physics

		exam	Aptitude
Exam	Pearson Correlation	1	.538**
	Sig. (2-tailed)		.000
	N	83	83
MCQ	Pearson Correlation	.538**	1
	Sig. (2-tailed)	.000	
	N	83	83

**. Correlation is significant at the 0.01 level (2-tailed).

Table 2 shows a Correlation Coefficient of 0.538 at 0.01 level of significance for Physics.

Table 3: Paired Samples Test of Subjective Examination and MCQ of Physics Scores

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference: Lower	Upper			
Pair 1	exam - MCQ	1.41566	16.12335	1.76977	-2.10497	4.93629	.800	82	.426

Table 3 shows a non significant t of 0.800. This means that hypothesis 1 is tenable.

Table 4: Paired Samples Test of Subjective Examination and MCQ of Biology Scores

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	exam - MCQ	4.84328	11.58678	1.00095	2.86345	6.82311	4.839	133	.000

Table 4 shows a significant difference exists between the two tests scores. This means that the hypothesis 2 is rejected.

Discussion

Results showed a significant relationship between the subjective examination and multiple choice question scores for both Physics and Biology. This actually answered the research questions in affirmative. Since a significant relationship exists between the subjective examination scores and that of the multiple choice question scores, the challenges encountered by teachers and learners in the assessment of large sized Science practical classes as identified by Strong (2013.1) can be eliminated if adopted. The use of multiple choice questions in assessment of large size classes has already been in by JAMB, NOUN, OUUK to mention but a few. Automating or use of computers in the assessment will even make it easy as the marking can be done faster and even if it is marked or graded manually, it does not need any expertise as anyone can use the key to grade them. This also ensures fair assessment and prompt release of feedback to students.

Adoption of MCQ for Science practical work addresses the issues raised by Bradley et al (1998) and Motlhabane (2013). However a significant difference exists in the result of paired t-test for Biology while it was not so in Physics. It could be attributed to view of Shulman and Tamirs (1973) who believe that practical work plays a role in conceptualizing Science. However Abimbola (1994) believes that a well planned demonstration can be shown to students and this will reduce expenditure and assessment could be done using MCQ. The researchers believe that MCQ can be used to assess large size classes at the 100 and 200 levels when almost all the Science students offer similar courses at these levels.

Conclusion

It is worth noting that the significant correlation means that MCQ could be adopted and automated because of the benefits of early release of results.It also eliminates subjective which is possible in subjective practical examinations where laboratory equipment are not enough.

Recommendations

The following recommendations are made:

1. Multiple Choice Questions could be used to assess large-sized practical Science classes especially at 100 and 200 levels.
2. It is possible automate the assessment process since it is easier to manage.
3. The study be repeated by other researchers for other science subjects.

Reference

Abimola, I.O. (1994). A Critical Appraisal of the Role of Laboratory Practical Work in Science Teaching in Nigeria. Journal of Curriculum and Instruction, 4 (1 & 2) pp 59-65.

Abimbola, I.O. & Danmole, B.T. (1995). Origin and Structure of Science Knowledge: Implications for Concept Difficulty in Science. Ilorin Journal of Education, 15, 47-59.

Abimbola, I.O. (1996). Advances in the development and validation of instruments for assessing students, Science Knowledge. Paper presented at the National Conference on Educational Assessment held at Lagos Airport Hotel, Ikeja, Lagos, 9-13 September 1996.

Australia University Teaching Committee (AUTC) (2002), Assessing large classes. Assessing Learning. Australia Universities. Ideas, strategies and resources for quality in student assessment. www.cshe-munimell.edu.au/assessing learning.

Beevers, C, Cherry B. Foster, M & McGuire, G.(1991). Software Tools for Computer Aided Learning in Mathematics" Ashgate Publishing Company.

Bloom, B.S. (Ed). (1956) Taxonomy of educational objectives-Handbook I:Cognitive, domain, New York: David Mckay.

Bradley, J. & Meaker, P. (1998). Research report on practical work in colleges of Education, Pretoria: National Research Foundation.

Brookhart, S.(1999). *The art and science of classroom assessment: The missing part of pedagogy.* Washington, D.C: ERIC clearing house on Higher Education. ED432938. Retrieved 7/03/2014 from http:// Chiron.voldasta.edu/whuitt/files/artsciassess.html.

Brown, G., with Bull, J. and Pendlebury, M. (1997). "*Assessing Student Learning in Higher Education" London: Routledge.*

Federal Ministry of Education (2004). National Policy on education, Federal Government Press. Lagos, Nigeria.

Good Practice Guide in Question and Test Designe: Notes constructed as part of the PASS-IT Project for Preparing Assessments in Scotland using IT (http://www.pass-it.org.uk/).http://www.calm.hw.ac.uk/GeneralAuthoring/031112-goodpracticeguide-hw.pdf

Mothlhabame, A. (2013). The Voice of the Voiceless: Reflections on Science Practical work in Rural Disadvantaged Schools. Mediterranean Journal of Social Sciences. Vol 4 No 14 MCSER publishing, Rome-Italy.

National Association of Biology Teachers, (1980), NABT guidelines for the use of live animal at the Pre-university level. The American Biology Teacher, 42(7), 426-427.

Newstead, S. and Dennis, I. (1994). "Examiners examined: The Reliability of Exam Marking in Psychology ", *The Psychologist*, Vol 7, pp 216-219.

McBeath,R. (1992) " Instructing and Evaluating Higher Education: A Guidebook for Planning Learning Outcomes" ETP New Jersey .

Okonkwo, C.A. (2008). An appraisal of students performance in physics using the Diagnostic physics test (DPT) in Nigeria Secondary Schools. Journal of Educational Research and Policies, 3(1), 31-36.

Okonkwo, C.A. (2005). Development and Validation of Diagnostic Physics Test (DPT) for Senior Secondary School students. The Educational Psychologist. A Journal of the Nigerian Council of Educational psychologists, 1 (2), 264-277.

Okonkwo, A.C. (2011). Adapting on Demand Examination System in National Open University of Nigeria. Turkish Online Journal of Distance Education (JOJDE), 12 (4), article 10. Accessed 17/2/2012.http://tojde;anadolu;edu;tr/tojde 45/articles/articles/article 10.htm.

Okonkwo, C.A. (2010a)., Rethinking and Structuring an Assessment system via effective deployment of

Technology international journal of education and development using information and communication technology (IJEDICT). 6 (2), 69-83.
Okonwo, C.A. (2010b). Using e-Assessment to enhance the operational efficiencies of the National Open University of Nigeria (NOUN). Journal of Educational Assessment in Africa, 5, 117-138.
Okonkwo, C. A. (2008). An appraisal of students performance in physics using the Diagnostic Physics Test (DPT) in Nigeria Secondary Schools. Journal of Educational Research and Policies, 3 (1), 31 – 36.
Okonkwo, C. A. (2005). Development and validation of Diagnostic Physics Test (DPT) for senior secondary school students. The Educational Psychologist. A Journal of the Nigerian Council of Educational Psychologists, 1 (2), 264 -277.
Okonkwo, C.A. & Ikpe, A. (2008). Re-engineering Examination at the National Open University of Nigeria. The on Demand Examination initiative. A paper presented at the International Association for Educational Assessment (IAEA) 34th Annual Conference Cambridge, 7-12 September,2008.Online publication Accessed (30/03/2009). http://www.iaea 2008. Cambridgeassessment.org.uk/ca/digitalassets 164845-okonkwo- CA.pdf
Cilliers, J. A, Basson, I., Kirschener, P.A. & Rutherford, M. (2000). Current views of the purpose of introductory physics laboratory in South African Journal of Higher Education, 14 (1): 20-30.
Paxton, M. (1998). " A Linguistic Perspective on multiple choice questioning" *The Higher Education Research and Development Society of Australia newsletter*, April 1998 20.1
Shulman,L.S.& Tamir, P. (1973). Research on teaching in the natural sciences. In R.M.V. Travers (Ed), Second handbook of research on teaching. Chicago: Rand McNally.
Strong, P. (2013). Teaching and assessing large classes. Teaching @UNSW/Assessment Toolkit. Assessing large Classes. Document version. Date 08/04/2013. http://teaching.unsw.edu.au/assessing-large-classes.
The Cross Sectoral Assessment Working Party (2011,June) Teachers' Guide to Assessment. Association of Independent Schools of the ACT incorporated. ACT Government, Education and Training. Archdiocese of Canberra and Goulburn Catholic Education Office.
UNSW Australia (2013). Teaching @UNSW/Assessment Toolkit. Assessing by Multiple Choice Questions. Document Version Date 20/03/2013. http://teaching.unsw.edu.au/assessing-multiple-choice-questions.
Vander Linde, H.J., Vander Wal., R.W.E., & Wilkinsan, A. (1994). Practical work in science teaching in developing communities. South African Journal of Education 14 (1); 48-52.
Velan, G. M.; James, P.; McNeil, & Kimar, R. K. (2008). *Integrated online formative assessment in the biomedical students: benefits for learning.* BMC Medical Education 8 (52).