Assessment of items effectiveness of senior secondary school mock chemistry examination conducted by Akwa Ibom State

[2]
[I] Dr. Osokoya, Modupe M, Uko, Idorenyin A
Institute of Education University of Ibadan, Ibadan

Abstract

Item analysis is a vital process in test development, employed to determine the psychometric properties of psychological instruments, to ensure the validity and reliability of their measures. This study focused on assessment of items effectiveness of2014 and 2015 Mock Chemistry Examinations, conducted by Akwa Ibom State. A descriptive study was carried on 700 students randomly selected from 18 secondary schools in Akwa Ibom State. Results of item analysis indicated that the two tests
(2014 and 2015) had moderate reliability coefficient of (). 49 and (). 75 respectively. The 2015 test had more difficult items than 2014. The discrimination level of items of2014 test had 26 good items and 24poor items, while that 0/2015 had 25 good items and 25 poor items. The point biserial correlation of items for the 2014 test showed 54% of items having low point biserial, 34% showed acceptable point biserial while 12% had negative point biserial. For the 2015 test, 50% showed very low point biserial, 42% showed acceptable point biserial while 8% of the items had negative point biserial. The good items are recommended for use while the bad ones are restructured for subsequent mock examinations to prepare the students for public examination.

Keywords: Items effectiveness, difficulty index, discrimination index, point biserial, Chemistry

Introduction

In the world today, there is need to acquire appropriate scientific skills to help an individual to cope with the challenges presented by the needs of modern workplace in our public and private sectors. One of the key science subjects at the senior secondary school level in Nigeria is Chemistry. It is a versatile subject, with its content being relevant to a vast majority of individuals across various fields. It is a prerequisite for offering most sciences-oriented courses in the tertiary institutions. The learning of Chemistry entails acquiring the basic knowledge of the universe, aimed at equipping man with skills needed to profitably mould and restructure his environment (Emmanuel, 2013).

Chemistry education means the teaching learning process of chemistry which involves curriculum/ syllabi, teaching —learning and examination.

Within the past decades the quality of Chemistry teaching has been under increased scrutiny in the identification of factors that relate to Chemistry success or failure of students. Korau (2006) reported factors affecting students' performance to comprise student factor, teacher factor, societal factor, textbooks and home related variables among others. These factors may produce useful clues for promoting Chemistry learning and achievement if properly harnessed. Despite these efforts, evidence from literature has shown that student's performance in Chemistry is still low at both internal and external examination. Saage (2009) identified poor primary school background in chemistry, lack of interest on the part of the students, lack of incentive for the teachers, incompetent teachers in primary schools, inability to work hard and psychological fear of the subject can be factors that are responsible for poor students' performance in chemistry. However, Nbine (2012) recommended the need for efficient and effective teachers who are professionally and academically qualified to promote Chemistry learning in schools. Nevertheless, this low performance of students in Chemistry can also be attributed to the nature of items teachers use in assessing students in Chemistry. As such, advancement on this stance demands additional investigation into the nature of Chemistry test items through quality attention to test construction guidelines. According to Odinko (2014) the primary purpose of evaluating students' learning outcomes in a given educational programme is to provide information for decision making about the programme (summative). It could also be seen as a method of acquiring and processing evidence needed to improve the learning and teaching activities (formative). Mock examination is a good example of summative examination used for placement.

Test such as mock examination is an important process in evaluating students' achievement and learning. The Mock examination which is normally conducted in Akwa Ibom State is thus a form of assessment to prepare students in SS2 for the Senior School Certificate Examination (SSCE). It is also a form of assessment to measure how much of the syllabus had been covered and assimilated by the students before going into the final examination class (SS3) for external examinations (WAEC, NECO and NABTEB). It usually consists of 50 multiple choice items and five essay questions. The multiple-choice objective test is what will be used in the cause of this research, to measure how good the items are in terms of psychometric properties in assessing students' achievement in the senior secondary two Chemistry examinations in Akwa Ibom State. This is because student's achievement in test or examinations can be used to determine the performance of an educational system and practice.

The Ministry of Education Akwa Ibom State Mock Examination is usually the last test written by the Senior secondary two students of the state before their Senior Secondary Certificate Examination (SSCE) conducted by

public examining bodies such as West African Examination Council (WAEC) National Examination Council (NECO), National Board for Technical Examination (NABTEB).

Discrimination index, and difficulty index; which is the ratio of the whole cluster that correctly choose the item are item statistics that can aid assessment of item effectiveness (Michigan State University report, 2014). According to their report, a great index shows a simple item, while difficult item is signified by a lesser difficulty index. Though some data analysts prefer to use difficulty index which is the ratio of how difficult or easy an item is for a group of testees. This index may be obtained by adding scores of examinees in the lower collection and those of the upper collection and dividing by the sum of examinees in the upper and lower collection (Adeleke 2010). For classroom achievement tests, most item writers desire items with difficulty indices with lower boundary of 0.30 or 0.40 and upper boundary of 0.60. This view is supported by Zurawski (2009) and Gul-ArNavi Khan (2015).

According to Suruji and Rana (2014) pinpointing defective test and signifying the aspect that the student's strength is or not is the two impacts of item analysis. They added that successes of separate test item can also be achieved through item analysis with respect to its difficulty and discrimination indices i.e. to differentiate between success and failure in a test. In analyzing for effectiveness, items with lesser achievement may be recollected and reviewed if they tally to a particular instructional objective in the course. Items statistically achievable but are not correlated to specific instructional objectives should be revised before being used.

Difficulty index explains how easy or difficult the question was for a specific group of students. The greater the difficulty index the simpler the question will be for the students; the lower the difficulty index the more challenging the questions will be. Discrimination index of a multiple-choice item is the chances to differentiate between the bright testees and poor performing testees (Alonge, 2003). Oyejide (1991) defines discrimination index as the strength of an item to differentiate the higher achievers from the lower achievers, Discrimination index of a test ranges from O to +1. The more closed the value is to l, the superior the item (Oyedeji, 1991; Kelly, 1989). For item difficulty the range of 0.4S p 0.6 is accepted and for discrimination indices q 20.3 is accepted (Thorndike, 1973). With these properties one is sure to have good items which also correlate with the rest of the items which is to be selected for final administration.

Point biserial correlations are essential component in item analysis of a set of test questions. They observe the extent to which scores on one item correlate with scores on the overall items in a scale. It also assesses item redundancy; that is the degree to which all item pools are measuring the same content (Cohen & Swerdlik, 2010). Point biserial correlation (rpbis) also helps in determining effective items. It shows how an item correlates with the total score. It is a relationship between a dichotomous variable and a continuous variable (ranging from zero to maximum number of the items in the test). Point biserial values ranges from -l to +1 (Varma 2006). Brown (2001) states that a greater positive point biserial correlation shows that testees with great scores are also getting the items right as expected and students with low scores are also getting the items wrong. A low pbis show that testees who got the item correct tend to perform not too well on the overall test and student who got the item incorrect tends to perform well on the overall test which indicates an anomaly. Problematic items will always show low pbis.

Effectiveness of test items is important because result obtained from an effective test item gives the true ability level of the testee. It is therefore pertinent that results obtained from Akwa Ibom State Senior Secondary Il Chemistry Examination should be effective in reporting the true ability levels of students, hence the

need to assess item effectiveness of the Akwa Ibom State Senior Secondary Il Chemistry Examination. Psychologists, educators, guidance counsellors, etc., use test results for a variety of purposes. Fahmi and Naser (2013) in their study on analyses of the use of a single best answer format in an undergraduate medical examination reported that 23percent out of the 100 items had an undesirable point biserial. Thus, it is imperative to identify ineffective items to ensure that the overall results are meaningful. On identification, it is also necessary to expunge bad items from the items pool because it lowers the trustworthiness of the test and also confuse the test takers during the test administration process (Varma, 2012).

Statement of problem

In Akwa Ibom State, the Ministry of Education usually conducts qualifying examination for students in SS2 to assess their readiness to write the final examinations for certification which is piloted by two known public examining agencies in Nigeria (West African Examinations Council (WAEC) and National Examination Council (NECO)). Chemistry is a core science subjects which is required by students at a minimum of PASS level of their secondary education to meet the eligibility criteria for admission to study specific courses in some of the higher institutions in Nigeria. Though mock examination is used as a benchmark to qualify learners to write the final examinations conducted by WAEC and NECO. However, little or no literature exists to provide evidence that the items used in mock examinations in Akwa Ibom State are analyzed in order to determine their psychometric properties before the appropriate items are selected. It is therefore difficult to testify to the effectiveness of those items used in assessing SS2 students. Hence, this work assessed the effectiveness of the items of senior secondary two Chemistry Examination conducted by Akwa Ibom State Ministry of Education.

Research questions

The underlisted Research questions guided the studies

1. What is the difficulty index of each of the item of the 2014 and 2015 objective Chemistry tests?
2. What is the discrimination index of each of the 2014 and 2015 chemistry objective test items?
3. What is the point biserial correlation of items in the two tests?

Research Methodology

Research Design

The study adopted a descriptive design of survey research type. Population consisted of all the senior secondary school two (SS2) Chemistry students in Akwa Ibom State. A multistage sampling procedure was employed. The three Senatorial Districts in Akwa Ibom State were used for the study. First, simple random sampling was adopted to select two (2) Local Government Areas from each senatorial district making a total of six Local Government Areas. Secondly, simple random sampling technique was used to select three public secondary schools in each of the local government areas and thus a total of 18 secondary schools were selected and used for the study. Furthermore, intact class of senior secondary school two students offering Chemistry were used in each of the 18 schools selected. The total sample of the study was 700 respondents.

The instrument used in the study was adopted which was The Akwa Ibom State Ministry of

Education Mock Chemistry Examination for senior secondary two. The years of investigation were 2014 and 2015. The two objective items of each year was merged in one instrument and labelled section A & B respectively. The parallel reliability of the two-test obtained was 0.8. The two tests were valid in content as affirmed by the 3 chemistry experts involved in the study using test blueprint. The nature of the test (objective achievement test) informed the choice of test blueprint. The test blueprint ensures that expected cognition levels are covered.

Item analysis was carried out using SPSS and MS EXCEL packages respectively for the item properties (difficulty index, discrimination index and point biserial) discussed in the background and to answer each of the research questions raised. The data generated from the respondents was sorted by their overall scores in descending order to obtain the best 27% (upper scorers) and lower 27% (lower scorers) the total number that scored each item correctly among all the testees were obtained and denoted by item SCORE. The above estimations were used to calculate the following

1. . Item discrimination index = Diff/n; where diff = difference between the upper and lower scorers, n= total number of testees in either upper or lower group.

ii. Item difficulty index = item SCORE/N; where item SCORE is the total number of testees that answer each item correctly, N= addition of total number of testees in upper and lower group.

Results

Research Question 1: What is the index of difficulty of each of the items of2014 and 2015 multiple choice chemistry tests?

Table 1: Difficulty level (p-value) of 2014 and 2015 Akwa Ibom State Chemistry Test

Item	2014	2015	Item	2014	2015
1	.67	.88	26	.27	.22
2	.31	.89	27	.49	.36
3	.65	.18	25	.27	
4	.49	.45	29	.44	.29
5	.30	.19	30	.39	.31
6	.86	.16	31		.18
7	.63	.32	32	.21	.40
8	.34	.41	33	.33	.86
	.45	.41	34	.56	.26

10	.18	.32	35	.13	.32
11	.23	.70	36	.59	.31
12	.25	.51	37	.57	.45
13	.27	.35	38	.32	.35
14	.19		39	.35	.21
15	.29	.57	40	.24	.29
16	.26	.45	41	.23	.29
17	.18	.47	42	.25	.24
18	.49	.27	43	.46	.22
19	.26	.25	44	.30	.26
20	.57	.43	45	.35	.46
21	.51	.17		.19	.17
22	.18		47	.51	.21
23	.33	.59	48	.62	.25
24	.63	.50	49	.29	.18
25	.37		50	.30	.37

Table I shows the difficulty indices (p-value) of the 2014 and 2015 chemistry test respectively. The table reveals that different p-values exist for different items of each year represented. Based on Classical Test Theory, the difficulty index should range from O to l. This could be divided into three where items with p-value of 0.7and above is easy, 0.3 - 0.69 is moderate and 0.1 O .29 is difficult. Backhoff2012.

Table 1.1 This table shows a summary of spread of Items Based on Difficulty Indices based on how easy, moderate and more difficult the items were for both years.

Difficulty Index(p)	2014	2015
Easy items (p> 0.70)		
Moderate items (0.31 = 0.70)	30	25
Difficult items (p = 0.30)	19	20

From Table 1.1, 60% of the items of the 2014 Chemistry test were moderate, 2% of the items was very easy while 38% of the items were difficult. For the 2015 Chemistry test, the difficulty indices of the items revealed that 50% of the test items were moderate, 40%ofthe items were difficult while 5%were easy items. The percentage of difficult items in the two tests (38% and 40%) were high compared to Backhoff (2012) who stated that 5% of the question in a test is required to be very difficult for a balanced test.

Research Question 2: What is the discrimination power of each of the 2014 and 2015 Chemistry test items?

Table 2: Item Discrimination Indices

Item	2014	2015	Item	2014	2015
1	.36 .23 26 .16 18 .11 .10 27 .28 .41				
2		.39	.10	28	.17	.38
3		.33	.37	29	.41	.22
4	12 30 .22 .24 .03 .06 31 .30 .10 .41 .19 32 .07 .38 .39 .43 33 .31 22				
5					
6					
7					
8					
	.16	.24	34	.25	.16
10	.09	.13	35	.08	16
11	.06	.38	36	.42	.18
12	.26	.29	37		.49
13	18	.17	38		.20
14	.11	.47	39	.19	. 14
15	.13	.43	40	.11	.18
16	.00	.53	41	.01	16
17	.07		42	.03	14
18	.52	.27	43	.47	.10
19	.18	.07	44	.22	18
20	.28	.38	45	.26	.43
21	.42	.03	46		.02
22	.07	.49	47	.57	
23	.11	.47	48	.47	.20
24	.51	.52	49	.20	.08
25	.29	.17	50	.20	.36

Table 2.1 Distribution of Items Based on Discrimination Indices. A summary of the discrimination indices is grouped under Very good items, Reasonable items, Marginal items and Poor items as seen below.

Item Discrimination Value(d)	2014	2015

Very Good items (d = 0.40)	9	11
Reasonable items (0.30-0.39)	6	6
Marginal items (0.20-0.29)	11	9
Poor (d =0.19)	24	24

The summary of the result on table 2.1 reveals that the higher the discrimination index, the better the item. Such value indicates that the item discriminates in favour of the upper scoring group which should get more items correct Frisbie 1986.

Research Question 3: What is the point biserial correlation of items in the two tests?

Table 3: Point Biserial Correlation of 2014 Akwa Ibom State Chemistry Test

Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	18.28	20.590	.212	.515
18.28		20.590	.212	.515
2	18.68	21.447	.005	.535
3	18.34	20.478	.225	.513
4	18.58	20.544	.201	.516
5	18.69	21.294	.532	
6	18.10	21.690	-.038	.536
7	18.37	20.426	.231	.512
8	18.68	20.347	.270	.509
9	18.53	21.393	.008	.536
10	18.80	21.507	.005	.534
11	18.77	21.554	-.011	.536
12	18.78	20.850	.182	.519
13	18.71	21.263	.053	.531
14	18.78	21.469	.012	.534
15	18.72	21.310	.532	
16	18.73	21.816	-.080	.543
17	18.80	21.519	.003	.534
18	18.51	19.960	.329	.501
19	18.71	21.182	.073	.529
20	18.42	20.766	. 148	.521
21	18.47	20.423	.223	.513
22	18.82	21.385	.047	.530
23	18.68	21.394	.017	.534
24	18.36	20.143	.300	.505
25	18.63	20.871	.132	.523
26	18.73	21.274	.053	.531
27	18.54	20.712	.159	.520

28	18.73	21.231	.065	.529
29	18.57	20.433	.225	.513

30 18.61 20.825 .522 31 18.32 20.662 .186 .517 32 18.75 21.481 .005 .535 33 18.64 20.749 .163 .520

34 18.42 21.079 .078.529

35 18.86 21.432 .046 .530 36 18.39 20.469 .219 .514 37 18.45 20.986 .097 .527 38 18.66 21.296 .038 .532 39 18.63 21.173 .063 .530

40 18.74 .013 .534 41 18.73 21.877 -.095 .544

42 18.73 21.769 -.069 .542 43 18.53 20.293 .254 .510

44 18.67 21.070 .093 .527

3	18.32	26.707	.067	.644
4	18.06	25.755	.220	.634
5	18.32	26.516	.115	.642
6	18.32	27.004	-.010	.648
7	18.19	26.618	.058	.646
8	18.08	25.605	.252	.632
9	18.09	26.199	.132	
10	18.17	26.775	.023	
11	17.81	25.708	.246	.633
12	17.96	26.200	.127	

13			18.12		26.408		.093		.644			
14			18.03		25.327		.304		.628			
15			17.91		25.594		.252		.632			
	45	18.63	20.941	.116	.525	46	18.76	21.647	-.037	.538	47	18.50	19.892	.345	.499							
	48	18.38	20.164	.291	.506	49	18.71	21.212	.530													
	50		18.68		21.309		.037		.532			

Table 4 shows the point biserial correlation of 2014 Akwa Ibom State Chemistry Test. The figures in the table indicate the degree to which both data set that is, testees' overall examination scores and individual scores correlate. From the table, 54% of the items show very low point biserial, 34% shows acceptable point biserial while 12% of the items had negative point biserial.

Table 4: Point Biserial Correlation of 2015 Akwa Ibom State Chemistry Test

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	18.21	25.830	.234	.634
2	17.58	26.702	.105	.642
1	18.21	25.830	.234	
2	17.58	26.702	105	
16	18.07	25.235	.327	.626
17	17.99	25.409	.286	.629
18	18.21	26.156	.164	.639
	18.21	26.949	-.011	.650
20	18.08	25.642	.245	.633
21	18.31	26.970	-.003	
22	17.86	25.341	.312	.628
23	17.94	25.383	.293	.629
24	18.01	25.144	.625	
25	17.57	26.409	.638	
26	18.29	26.261	.167	.639
27	25.555	.273	.631	
28	18.15	25.799	.225	.634
29	18.20	26.215	.148	
30	18.17	26.198	.145	
31	18.31	26.840	.030	.646
32	18.13	25.729	.234	.634
33	17.57	26.242	.268	.635

				18.20		26.732		.035		.647		
35				18.17	26.430	.095						
36				18.18	26.276	.129						
37				18.03	25.282	.313	.627					
38				18.12	26.424	.090	.644					
39				18.30	26.636	.078	.644	40	18.18	26.583	.065	
41	18.16	26.809	.649									
42	18.23	26.699			.048	.646						
43	18.27	26.827	.647									
44	18.25	26.474	.102									
45	18.06	25.637	.243	.633								
46	18.30	27.133	.651									
47	18.26	26.985	-.013	.649								
48	18.24	26.788	.028	.647								
49	18.28	26.816	.029	.647	50	18.14	25.848	.212	.635			

Table 3 and 4 show the point biserial correlation of items for the two years. Point biserial values simply tells us if the right people got that item right. From Table 4, 50% of the items showed very low point biserial, 42% showed acceptable point biserial while 8% of the items had negative point biserial. For example, item 6 of the 2015 Chemistry mock examination showed a negative point biserial correlation of -0.010 this negativity could be due to its low difficulty index of 0.16 and its inability to discriminate (discrimination index of 0.06) between the bright and dull students. The Corrected point biserial indicates that the score from individual items did not form part of the total score before the correlation. This is a minor but very key detail because inclusion of the item score in the total score can automatically increases the point biserial value due to correlation of the item score with itself.

Discussion of findings

The results of the analysis indicated that the two tests had moderate reliability coefficient. The 2015 test had more difficult items than the 2014 test. 60% of items of the 2014 test were moderate, 2% of the items were very easy while 38% were very difficult. For the 2015 Chemistry test, 50% of the items were moderate, 5% of the items were easy while 40% of the items were difficult. Wood (1960) the greater the percentage of students scoring an item right, the easier the item. The greater the index of difficulty, the easier the item is to that group of testee. The discrimination level of items of 2014 test had very good items to be nine, reasonably good items were six, marginal items were Il and poor items were 24 while that of 2015 had very good items to be 10, reasonably good items were 6, marginal discrimination were 9 and poor items were 24. This implies that the higher the discrimination index, the better the item because the item discriminates in favour of the upper scoring group which should get more items right. The parallel reliability index of the two tests was good with reliability index of 0.572. The point biserial correlation of items for the 2014 test shows 54% of items having low point biserial, 34% shows acceptable point biserial while 12% had negative point biserial. For the 2015 test, 50% shows very low point biserial, 42% shows acceptable point biserial while 8% of the items had negative point biserial, Ebel and Frisbie (1986) state that biserial correlation (rbis) describes the relationship between scores of testees on a single test item (e.g "O" "l ") and

scores (e.g "O" "50") on the total test for all examinees.

The result in Table 2 reveals that the higher the index of item discrimination, the superior the item because the index point to the item discriminating in favour of the upper scoring group of testee which should definitely get more items correct. For example, item 6 of the 2015 Chemistry Test shows a negative point biserial correlation of-O.OI 0 this negativity could be due to its low difficulty index of(). 16 and its inability to discriminate (discrimination index of (.06) between the bright and dull students. The Corrected point biserial indicates that the item score did not form part of the total score before running the correlation. This is a minor but important fact to be considered as inclusion of the item score to the total score can automatically inflate the point biserial value due to correlation of the item score with itself. From the two years of investigation on tables 3 and 4 following the high percentage of items showing low point biserial, it reveals that the brilliant students or those in the upper scoring groups were getting the item wrong. This is in line with Fahmi and Naser 2013 who observed negative point biserial and therefore stated that negative point biserial means the candidate who scored highest in the whole test also scored zero on these questions than those who performed poorly on the whole examination. Another finding is that the items were too difficult which lead to guessing and should be eliminated or reviewed.

Conclusion

This study focused on assessment of items effectiveness of Akwa Ibom State Ministry of Education Chemistry Examination using the instruments used in the years 2014 and 2015 respectively. Information was sought on the difficulty level, discrimination power, reliability, point biserial correlation between the two tests. A major finding reveals that the items of the two years had bad items which needs to be revised. It is on this note that every educator must see item analysis as an important aspect when it comes to assessing how good the items are in the instrument prior to its administration.

Recommendations

On the basis of the findings of this study, the following recommendations were made.

- Trial testing of items developed should be carried out before the final testing and item analysis should be done to ensure the quality of the items.
- Good items for all forms of tests (items with high discrimination index and moderate difficulty) should be used in assessment of students.
- Problematic items identified should be substituted or eliminated completely from the test. Such item can be improved upon and stored in item bank especially if the same items are going to be administered in subsequent years based on the content it is testing.

References

Adeleke. J. O (2010). The basics of research and evaluation tools. Somerest Ventues, Ikeja Lagos.

Alonge, M. F., & Ojerinde, A. (1987). Guessing Tendency and Differential Performance of Secondary School Students on Mathematic Achievement Tests. Journal ofScience Teachers Association ofNigeria (JSTAN) 25 (1) and2 56-63.

Angelo, T. and Cross, P. (1993) Classroom assessment techniques: A handbook for college teachers. San Francisco: Jossey Bass.

Backhoff, E., Larrazolo, N & Rosas, M. (2012). The level of difficulty and Discrimination power of the Basic Knowledge and Skills Examination (EXHCOBA). Electronic Journal of Educational Research 2012;2.

Retrieved on Mar 2, 2012, from http://redie.uabc.mx/v012nol/contentsbackhoff.html.

Brennan, R. (2001). An essay on: The history and future of reliability from the perspective of replications. Journal of Educational Measurement, 38, 295-317.

Brown, J. D. (2001). Statistics Corner. Questions and answers about language testing statistics: point biserial correlation coefficient.Shiken: JAL T Testing & Evaluation SIG Newsletter, 5(3), 13-17. Retrieved June l , 201 7 from: http://jalt.org/test/bro_7.htm.

Cohen, R. J. and Swerdlik, M. E. (2010). Psychological testing and assessment: An introduction to test and measurement (7th Edition). Boston: McGraw-Hill Higher Education.

Cronbach, L. J. (2004). My current thoughts on Coefficient Alpha and successor procedures. . Educational and Psychological Measurement, 64, 391-418.

Ebel, R.L., &Frisbie, D.A. (1986). Essentials of educational measurement. Englewood Cliffs, NJ: Prentice-Hali.

Emmanuel, B. (2013). The place of Nigeria certificate in Education Chemistry teachers in UBE Basic Science Programme. STAN 54th Annual conference proceedings 177- 181 Englewood Cliffs, NJ: PrenticeHall.Examining Bodies for Senior Secondary Schools Mathematics.

Research Journal of Applied Sciences, 2(8): 913-915.

Fahmi, I. E and Naser, M. (2013). Analysis of use of a single best answer format in an Undergraduate Medical Examination Qatar medicaljournal.

Fred, N. K. and Howard, B. L. (2000). Foundations of Behavioural research. Fourth edition, California State University, Northridge. Earl Mcpeek purblishers.

Gronlund, N. E. and Linn, R. L. (1990). Measurement and Evaluation in Teaching. New York: Macmillan Publishing Company, 6th edition.

Gul-Ar Navi Khan, Nazia Ishrat, Abdul Q. Khan (2015). Using item analysis on essay types questions given in summative examination of medical college students: facility value, discrimination index vol 3, No,

Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991). Fundamental ofitem response theory. Newbury Park (CA): Sage.

Hatfield, S. (1997). "Assessment in the Major Tools and Tips for Getting Started." Paper presented at the 1997 Assessment

Conference in Indianapolis. Professor Hatfield is the Assessment Coordiator at Winona State University. http://www.merriam webster.com/dictionary/multiple%20choice.

Hunter, J. E. and Schmidt, F. L. (1990). Methods of meta-analysis: Correcting error and bias in research findings. Newsbury Park: Sage Publications.

Kahan, B. (2008). Review of Evaluation Frameworks. Saskatchewan Ministry of Education, Kael Consulting.

Kelly,T.L. (1989) The Selection of Upper and Lower Groups for the Validation of Test Items. Journal of Educational Psychology, 30.

Korau, Y. K. (2006). Educational Crises Facing Nigerian Secondary Schools and Possible

Solutions A paper presented at Faculty of Education, University of Ibadan.

Lawal, A. (2001). Evaluation of Students' Learning Outcomes l: Types and Uses of Tests. In I. O. Abimbola (Ed.). Fundamental Principles and Practice of Instruction. Ilorin: Belodan (Nig) Enterprises & Tunde Babs Printers.

Michigan University, Academic Technology Service, Introduction to Item Analysis. Retrieved May 07, 2017, from: http://www.scoring.msu.edu/itanhand.html

Nbina, J B (2012) Analysis of Poor Performance of Senior Secondary Students in Chemistry in Nigeria. An International Multidisciplinary Journal, Ethiopia Vol. 6

(4).SerialNo. 27(Pp 324-334

Odinko, M. (2014). Theory and Practice in Educational Evaluation p (50-60).

Oyejide,A.P (1991) Effects of' Confidence Scoring Procedures on the Psychometric Properties of Three Multiple Choice Test Formats.

Saage, O. (2009). Causes of Mass Failures in Mathematics Examination among Students a Commissioned Paper presented at Government Secondary School. Karu Abuja Science Day 1st March.

Suruji and Rana, S. R. (2014). "Test Item Analysis and Relationship Between Difficulty Level and Discrimination Index of Test Items in an Achievement Test in Biology". Paripex - Indian Journal of Research, Vol. 3(s6) pp.56-58.

Thorndike, R. L. (1971). Educational Measurement. Washington. DC: American Council on Education.

Varma S. (2012). Preliminary item statistics using point-biserial correlation and pvalues. Educational Data Systems 1 n http://www.eddata.com/resources/publications/ EDS Point_B iseri al .pdf. Accessed May Il, 2017.

Wood, D.A. (1960). Test construction: Development and interpretation of achievement tests. Columbus, OH: Charles E. Merrill Books, Inc.

Zurawski, Raymond M. (2009): Item Analysismaking the most of it, unpublished paper