***ISSN: 3281-3721***
***Volume 10, Number 2, 2023***
***International Journal of Education and Sustainable Development***

# FUNCTIONALITY OF ITEM DISTRACTERS OF SENIOR SECONDARY TWO MATHEMATICS ACHIEVEMENT TEST SCORES IN BENUE STATE, NIGERIA

**Amakiri, Hager Atisi Eremina and Omale Onuh**
**[1]Department of Educational Psychology, Guidance and Counselling, Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Rivers State, Nigeria.**
**[2]Department of Guidance and Counselling, Joseph Sarwuan Tarka University, Makurdi, Benue State, Nigeria.**

*[1]hager.amakiri@iaue.edu.ng, [2]omaleonuh@gmail.com*

## ABSTRACT

*This study examined the functionality of item distracters in multiple-choice Mathematics achievement tests for Senior Secondary Two (SS2) students in Benue State, Nigeria. The study aimed to determine the impact of functional and non-functional distracters on test reliability and student performance. A quasi-experimental design was employed, involving 501 SS2 students across 12 schools. Two Mathematics achievement test instruments: one containing more functional distracters (MATICMFD) and another with fewer functional distracters (MATICLFD) were administered. Data were analysed using descriptive and inferential statistics, including Item Response Theory (IRT), item difficulty index and t-test analysis. Findings revealed that both test instruments had acceptable item difficulty indices (0.30–0.90) and the mean achievement scores of students who took tests with more functional distracters (M = 14.63) and those with fewer functional distracters (M = 14.46) were not significantly different (t = 0.729, p > 0.05). The conclusion drawn from the study was that functional distracters do not significantly affect test scores but enhance item discrimination. It was recommended that Mathematics teachers undergo training in item analysis to ensure the development of high-quality multiple-choice tests that effectively measure student achievement.*

***Key words**: **Assessment, Reliability, Distractor, Functionality, Mathematics***

## INTRODUCTION

Globaly, assessment is carried out in every area of life including school where it is formally used. In Nigerian school system, it is a household word. It is an activity that goes on in all schools in the country. In every subject, the teacher is expected to conduct or carryout an assessment of his/her students. Every school registering students' for high stake examinations such as, the West African Senior School Certificate Examination (WASSCE) and National Examination Council (NECO) prepares and submits to the examination bodies the scores of her students on continuous assessment. The schools are also required to submit assessment scores to ministry or agency conducting the Junior Secondary School graduating examination, that is, Basic Education Certificate Examination (BECE) (Monday, 2009). At postsecondary school level, the Organization for Economic Co-operation and Development (OECD) launched the assessment for Higher Education Learning Outcome (AHELO) an initiative with the objective to assess whether it is possible to develop international measures of learning outcome in higher education. AHELO armed to complement institution-based assessment by providing a direct evaluation of students learning outcomes at the global level. The purpose of the AHELO feasibility study is to see whether it is practically and scientifically feasible to assess what students at higher education know and can do upon graduation within and across these diverse contexts (OECD, 2012). Assessment in the form of test is absolutely natural. It is a way of believing that the students have learnt or not learnt what they are taught (Department of Education, 2008). Test could be seen as a systematic procedure for observing a person's behaviour and describing it with the aid of numerical scale or category (Onuka & Durowoju, 2013; Ali, 2014). It is an experiment intended to show that something works or works well (Odinko, 2014). For Adikwu et al (2016), a test is an instrument

used for eliciting information from the testee or respondent.

A good test is that which is practicable, usable, reliable and valid. The purpose of testing (particularly for achievement) is for the teacher to obtain information about how much of the stated objectives have been acquired by the learner in order to decide what to teach next, or for the students to get information about what they know, so that they also have an awareness of what they need to learn or review. For convenience in terms of wider content coverage, ease of administration and marking, multiple choice questions are widely used by classroom teachers in various subjects/ areas, although evidence of validity and reliability are not proven due to their inability to analyse the test items hence ending up with psychometrically bad constructed testing instruments. In developing good quality test items to effectively measure students' achievement, it is pertinent that the best practices in test construction be employed by teachers. Construction of a good Teacher Made Test (TMT) with high psychometric properties will help students to achieve competence and it will provide guidelines for better performance when they are exposed to standard tests. In many instances, students are often examined with TMT items which have higher or lower item parameter(s) than the students' academic age/level even though such test items are drawn from the content of what has been taught to the students but without adherence to the procedures of test development. This is a practice among many Nigerian teachers who lack the knowledge of test construction to set their test items and use same to evaluate their students' academic achievement without analysis of item difficulty, discrimination and whether the distracters are functional or non-functional. Many of such test items fail to reflect the real abilities of the testee on the content measured hence, achievement is poorly measured.

Objective tests are types of test that require very short or brief answers. The requirement may be just a letter of the alphabet or other things which are very specific. They are presented in such a way that two or more equally able examiners scoring them will arrive at the same or nearly the same score since they are not subject to the influence of bias. Monday (2009), highlighted five (5) types of objective test which include short answer items, Completion, True/False, Matching tests and Multiple-Choice Items (MCI). This study is much more concerned with Multiple Choice Test Items due to it many distracters per item.

Multiple choice Items (MCI) are type of objective response in which testees are asked to select only correct answers from the choices offered as a list. Multiple choice items consist of a stem and several alternative answers called options which are usually at least three and at most five. The stem is the opening, a problem to be solved, a question asked, or an incomplete statement to be completed. The options are the possible answers that the testee can choose from, with only one correct answer called the *key* and the incorrect answers called *distracters*. Usually, a correct answer earns a set number of points toward the total mark, and an incorrect answer earns nothing.

Example 1: If 2x : (x+1) = 3 : 2, what is the value of x? A. $\frac{1}{2}$ B. 1 C. $1\frac{1}{2}$ D. 3. In the example above, only D bears the key (ie 2 x 3 : 3+1 yields 6 : 4 = 3 : 2 to the lowest form).

Example 2: If 33% of a number is 1386, find 42% of the number A. 1764 B. 1000 C. 2100 D. 1089. In example 2, the correct option is A which bears 1764

There are certain rules guiding the construction of both the stems and the response set of options in Multiple Choice Items. In the case of the stem, it should be written clearly and positively to avoid words such as except or not. If such words must be used, they should be bold or be in upper case letters or underlined. No information needed to answer the item should be left out so that a testee can answer it even without looking at the options and to keep the options brief. In the case of options, they should be arranged vertically and should contain only a correct answer (key). For numbers, the options should be arranged in order of magnitude or be almost equal in length and be logically arranged in the case of words.

Example 3: The following are quadratic expressions EXCEPT

A. $2n^2$ - 5n + 6

B. $y^2 = \frac{1}{4}$
C. (p+1)(2-p)
D. (5-x)(t+5)

Another concern about set of options is that the phrases 'all of the above' and 'none of the above' should not be so that a testee would not draw a conclusion that all of them are correct or none of them is correct. Since the primary aim of testing is for the testees to report self to the examiner about what has been learnt and the extent to which the material has been achieved, teachers should be knowledgeable to develop better multiple-choice tests in terms of functional distracters.

A distracter in a multiple-choice item, is the option that is not the key but appealing to some testees. It is plausible if it presents confusion to testees who are not sure of it. It is the option which appears to some examinees as the correct answer. A distracter is functional, effective or good if it has a response frequency of 5% or more where testees from the lower ability group are much attracted by it than those from the higher ability group and has a negative discriminative index (Juliana & Maria, 2017). That is, the distracters should attract only testees with partial knowledge of the trait measured. In this study, multiple choice test items with functioning distracter(s) discriminate among respondents with varying academic abilities since only examinees that possess the construct measured will score the item correctly hence, yielding scores that are almost the same if the instrument is tested on two or more groups or its equivalent form tested on the same group under similar and uniform conditions. The sameness in the several results obtains by the testees in several measures of the same characteristic using closely related instruments connotes the reliability of the instrument(s). Reliability (as one of the psychometric properties) of a test could not hold if the test items fail to discriminate examinees, whereas functioning distracters enhance discrimination among testees. Nitku and Brookhart, (2011); and Ali (2014) stated that a test score is said to be reliable when there is a reason for believing that the score is stable, un-fluctuating and trust-worthy. Stability and trust-worthiness depend upon the degree to which the score is an index of time-reliability' is free from chance error. Callans (2012) demonstrated for clarifications on the definition of reliability that, if an achievement test has been administered on Group 'A' and found a Mean score of 55. Again after 3 days the same test is administered on Group 'A' and found a mean score of 55. It indicates that the measuring instrument (Achievement test) is providing a dependable result.

Teachers in Secondary Schools are faced with the challenge of the inability to have well constructed achievement tests items hence the results of such tests are seldom appropriate. This is to say that there exist some flaws like validity and reliability in the teacher made instruments. One of the faulty features of some achievement tests made by class room teachers is low reliability of students' test scores in terms of discrimination or spread of scores among testees with varying abilities that come as a result of highly easiness or difficulty level of items resulting from non-functioning distracters or mis-keyed items. This is peculiar to all subjects including Mathematics and is fairly commonly observed in many schools due to diversified knowledge of teachers in terms of test constructions which a test is designed to suit the method or style of teaching. With regards to functional distracters as an index of item discrimination resulting to stable scores, the researcher felt the need to explore the functionality of item distracters in Teacher-Made Mathematics Multiple Choice Test (TMMMCT).

Mathematics teachers test their students in terms of logic reasoning with numbers, shapes and patterns or arrangement of objects where TMTs are usually given to determine their successes, the teacher needs to consider some other test qualities (in terms of multiple choice tests) like moderate p-value of between 0.30 and 0.70, discrimination index of the range 0.46 – 1.0 (April & Kristel, 2019) and functionality of item distracters which influence reliability coefficient for all classroom tests thereby ascertaining true variance through spread of scores which in turn helps to ensure the quality of measuring instrument as required for quality assessment in Education. The item Difficulty Index (P-value or **p**) is an appropriate choice for achievement tests

when the items are scored dichotomously (i.e correct and incorrect). It is the proportion of test takers who answer an item correctly. In the same light, it was averred that item difficulty values of between 0.30 and 0.70 are good for standardized achievement tests but 0.50 to 0.80 is the best for classroom tests (Tsegaye, (2019); April & Kristel (2019).

An effective distracter is an incorrect option in a multiple-choice item which by misconception attracts testees with partial knowledge of the trait measured (Rodriguez et al, 2014). Distracters constitute the most important element of a multiple-choice item. They also present the most challenging aspects of an item to the item writer. The difficulty of an item is most easily manipulated by the nature of the distracters, particularly in their similarity. Item-writing researchers have found that the plausibility and similarity of the distracters have a much greater impact on item difficulty than the characteristics of the stem do (Ascalon et al, 2007). Furthermore, Sudijono (in Mansour et al, 2012), stated that a distracter is functional if more than 5% of the respondents select it. A good functional distracter appeals to a higher proportion of low achieving examinees when compared with high-achieving testees. Distracters could be checked as $\frac{L-U}{N}$ where a positive valve indicates that more testees from the lower group choose the distracter than those from the upper group otherwise the distracter is chosen by more testees from the lower group. Some authors have pointed out that test item difficulty increases with increasing functioning distracters (Tarrant et al, (2009); Mozaffer & Farhan (2012); Nomdeo & Bandya (2015); Linnette et al, (2018); Aide & Loudon (2018); & Burud et al, (2019). Others contend that test items with less functioning distracters were more difficult (Pradip et al (2015). For Abdulghani et al (2014) though item difficulty increases with increasing functioning distracters, this is not always the case.

Due to these reasons, classroom teachers need to be given adequate orientation on Multiple Choice Items construction that would effectively measure a wide range of the objectives covered. In this study, the focus is on writing effective distracters which must be appealing to testees with partial knowledge of the attributes measured so as to discriminate among testees with diverse academic achievements. This study is based on the Item Response Theory (IRT). The earliest researchers who studied and developed the IRT were Fredrick, Larzarsfeld and Rasch in 1970s. The theory is mainly about the relationship between a testee and an item; ability desired of a testee to respond to an item on the test. In IRT it is believed that an examinee's response to a particular item is determined by an unobservable trait of the examinee. The trait is always hidden and hence called latent trait, denoted as theta ($\theta$). The IRT consists of three basic models; 1, 2 and 3-parameter models, each of which is used based on the nature of the test items. All the three models maintain the relationship that a test taker with a certain ability level can probably respond correctly to an item.

Therefore, this study seeks to assess the functionality of item distracters (in multiple choice items) on reliability of SS2 Mathematics achievement test scores using TMT in the content work of SS2. Multiple choice items were used because they are normally easy to administer, fast and easy to mark, test a wide content area, test a large number of testees objectively and above all, they are made up of distracters which their functionalities on reliability of the obtained scores would be assessed by the researcher.

## Statement of the Problem

Mathematics is a compulsory subject from primary to tertiary education levels in Nigeria education system. Students' enrolment is usually very high in mathematics classrooms in secondary schools because without at least a 'C' Grade in Senior Secondary Certificate Examination, it would be impossible for a student to gain admission into any tertiary institution in Nigeria. Multiple choice tests are a common method used by teachers for assessing students' achievement in mathematic in most secondary schools in Nigeria. However, the effectiveness of such assessment largely depends on the quality and functionality of their items distracters which is the incorrect answer options designed to challenge students and differentiate between varying

levels of cognition. The preliminary observations of the researchers reveal that most mathematics teachers utilise items drawn from text books or compiled by test vendors for assessing their students. This may be attributed to the challenging and time-consuming process of developing multiple-choice test items. This suggest that several distracters may be non-functional, meaning that they are rarely chosen by testees, which undermines the discriminatory power of the test items since the psychometric properties of the test items and item distracters were not determined by the subject teacher. This potential deficiency not only compromises the reliability and validity of the assessment scores but also raises concerns about the accuracy of the measurement of students' true ability in mathematics. Inaccurate test scores can lead to misinterpretation of students' performance. Considering that some Mathematics teachers brag about testees obtaining scores on all items in their multiple-choice test without recognising that a faulty constructed instrument could yield unreliable results. Hence, the researchers are asking, if the unreliability of such tests is due to non-functionality of item distracters? To answer the question, the researchers therefore assessed the functionality of item distracters on the reliability of scores in Senior Secondary Two (SS2) Mathematics achievement test scores in Benue North-East, using TMTs which shall help classroom teachers to gain better understanding of the need for item analysis in their Multiple-Choice Items construction.

## Aim and objectives

### Aim:

To examined the functionality of item distracters in multiple-choice Mathematics achievement tests for Senior Secondary Two (SS2) students in Benue State, Nigeria.

### Objectives:

To ascertain

1. Difficulty level of a test with items containing more functioning distracters and a test with items containing less functioning distracters respectively.
2. Mean achievement scores of students who take a test with items containing more functioning distracters.
3. Mean achievement scores of testees who take a test with items containing less functioning distracters.

### Research Questions:

1. What is the difficulty level of a test with items containing more functioning distracters (MATICMFD) and a test with items containing less functioning distracters (MATICLFD) respectively?
2. What is the Mean achievement score of students tested with items containing more functioning distracters (MATICMFD)?
3. What is the Mean achievement score of students tested with items containing less functioning distracters (MATICLFD)?

### Hypothesis

1. There is no significant difference in the Mean achievement scores of students who take a test with items containing more functioning distracters and those who take a test with items containing less functioning distracters

## METHODOLOGY

The study adopted a Non-equivalent Groups Design (quasi–experimental design). The population of the study covered 4896 Senior Secondary Two (SS2) students in all the 123 grant- aided Senior Secondary Schools in Benue North- East Education Zone in their intact classes, 2020/2021 academic session**.** The sample size for the study was 501 SS2 students in their intact classes

obtained from 12 different schools. Out of the 12 schools, 258 SS2 students responded to MATICMFD from six schools while 243, SS2 students responded to MATICLFD from the other six schools. Purposive sampling technique was adopted by the researcher to sample only three Local Government Areas (LGAs). Proportionate sampling technique was used to select the number of schools used in each of the three LGAs through sampling framed and two instruments were used for data collection in this study; Mathematics Achievement Test Items Containing More Functioning Distracters (MATICMFD) and Mathematics Achievement Test Items Containing Less Functioning Distracters (MATICLFD) each of which consisted of four options, the key and three distracters. The distracters were considered more functioning or less functioning based on the frequencies with which they were selected by the testees during trial testing as indicated here: Each of the instruments made-up of 20 items, having options A-D with one key for every item (making a total of 120 distracters and 40 keys for the two instruments since each of the 40 items had four options, 1 key and 3 distracters). Thus, a variation in the two instruments was just that the fourth option on MATICMFD and MATICLFD were not the same based on how they were chosen by testees during trial testing. Face and content validation of the instrument was done by five (5) Expert while Internal consistency of the instruments, MATICMFD and MATICLFD were established using split-half method. The reliability coefficient of MATICMFD was 0.969 and that of MATICLFD was 0.973. Descriptive and inferential statistics were used for data analysis. IRT item difficulty index was used to answer research question one. While mean scores were used to answer research questions 2 and 3. The t-test was used to test the hypothesis at 0.05 level of confidence,

## RESULTS

### Research Question one:

What is the difficulty level of a test with items containing more functioning distracters (MATICMFD) and a test with items containing less functioning distracters (MATICLFD) respectively?

**Table 1: IRT item Difficulty index of MATICMFD and MATICLFD**

| Item | MATICMFD | Remark | MATICLFD | Remark |
|---|---|---|---|---|
| **item1** | 0.75 | Good | 0.77 | Good |
| **Item2** | 0.80 | Good | 0.78 | Good |
| **Item3** | 0.75 | Good | 0.76 | Good |
| **Item4** | 0.75 | Good | 0.68 | Good |
| **Item5** | 0.68 | Good | 0.80 | Good |
| **Item6** | 0.78 | Good | 0.76 | Good |
| **Item7** | 0.77 | Good | 0.64 | Good |
| **Item8** | 0.66 | Good | 0.79 | Good |
| **Item9** | 0.73 | Good | 0.77 | Good |
| **Item10** | 0.76 | Good | 0.79 | Good |
| **Item11** | 0.70 | Good | 0.68 | Good |
| **Item12** | 0.74 | Good | 0.72 | Good |
| **Item13** | 0.80 | Good | 0.62 | Good |
| **Item14** | 0.73 | Good | 0.72 | Good |
| **Item15** | 0.66 | Good | 0.81 | Good |

| | | | | |
|---|---|---|---|---|
| **Item16** | 0.77 | Good | 0.65 | Good |
| **Item17** | 0.63 | Good | 0.73 | Good |
| **Item18** | 0.66 | Good | 0.63 | Good |
| **Item19** | 0.71 | Good | 0.63 | Good |
| **Item20** | 0.64 | Good | 0.62 | Good |

Key: Item Difficulty Index is good once it falls within the range of 0.30 to 0.90.

Table 1 showed that none of the 20 items in the two test-types was bad since the difficulty index of each of the items was above 0.30 and fall within the acceptable range of 0.30 to 0.90.

**Research Question Two:**
What is the Mean achievement score of students tested with items containing more functioning distracters (MATICMFD)?

**Table 2: Mean and Standard Deviation of MATICMFD**

| Variable | N | Mean | SD |
|---|---|---|---|
| MATICMFD | 258 | 14.63 | 2.69 |

The results in table 2 showed that the mean of MATICMFD was 14.63.

**Research Question Three:**
What is the Mean achievement score of students tested with items containing less functioning distracters (MATICLFD)?

**Table 3: Mean and Standard Deviation of MATICLFD**

| Variable | N | Mean | SD |
|---|---|---|---|
| MATICLFD | 243 | 14.46 | 2.48 |

The result in table 3 shows that the mean of MATICMFD was 14.46.

**Test of Hypothesis**

**Table 4: t-test analysis of Mean Achievement scores of MATICMFD and MATICLFD tests**

| **Group** | **N** | **$\bar{x}$** | **SD** | **Df** | **t-value** | **α** |
|---|---|---|---|---|---|---|
| **MATICMFD** | 258 | 14.63 | 2.69 | 499 | 0.729 | 0.05 |
| **MATICLFD** | 243 | 14.46 | 2.48 | | | |

**N = Number of testees that responded to each test, $\bar{x}$ = mean score for each test and S.D standard deviation of 20 items on KR-21**

Results in Table 4 showed that at df = 499, 0.729 > 0.05 (not significant) hence the null hypothesis is not rejected.

**Discussion of Findings**

The difficulty index of each item in the two instruments was within the range of 0.30 and 0.90 which indicated that the items were good. The findings are in line with those of April & Kristel (2019) and Tsegaye (2019) who stated that item difficulty values of between 0.30 and 0.70 are good for standardized achievement tests but 0.50 to 0.80 is the best for classroom tests. The results also show that the mean of MATICMFD was 14.63 while that of MATICLFD was 14.46 and

the mean difference between the two test types was 0.17. This result conformed to that of Pradip et al (2015) whose findings revealed that test items with less functioning distracters were more difficult. It is also similar to part of the findings by Abdulghani et al (2014) which stated that though item difficulty increases with increasing functioning distracters but not always. The result however opposed those of Tarrant et al (2009), Mozaffer and Farhan (2012), Nomdeo and Bandya (2015), Linnette et al (2018), Aide and Loudon (2018) and Burud et al (2019) who in their separate studies pointed out that test item difficulty increases with increasing functioning distracters. The variation in this study and the reviewed studies may be due to the different sample sizes used in the studies or quality of items used or both. There was no significant difference between the mean score of testees on MATICMFD and MATICLFD. The null hypothesis which states that there is no significant difference in the Mean achievement score of testees who sat for MATICMFD and those who sat for MATICLFD was retained. This result could be different from that of other researchers in related studies due to test item analysis before its administration.

## CONCLUSION

Based on the findings of the study, it was concluded that there was no statistically significant difference between the mean achievement scores of testees who responded to the two tests, (MATICMFD and MATICLFD), difficulty indices of MATICMFD and MATICLFD were not significantly related.

## RECOMMENDATIONS

To further improve the psychometric properties of teacher made test, the researchers recommended that:

1. Mathematics teachers in Benue North-East should be trained by the State Government/ stakeholders to carryout test item analysis in the development of instruments in order to effectively measure students' achievement.
2. School administrators should ensure and encourage regular updates of test items by mathematics teachers through item analysis particularly in terms of Multiple-Choice Items using IRT.

## REFERENCES

Abdulghani, H.M., Ahmad, F., Ponnamperuma, G.G., Khalil .S. & Aldrees, A. (2014): The relationship between non-functioning distracters and item difficulty of multiple choice questions. J Health Spec 2014; 2(1),48-51.

Adebule, S.O. (2009). Reliability and level of difficulty of objective test items in Mathematics achievement test: *Educational Research and Review, 4 (11), 585-587.*

Adikwu, O., Aduloju, M.O. & Agi, C.I. (2016): Measurement and Evaluation in Education. Shibboleth Prints.

Aide, M. & Loudon, C, (2018). Item statistics derived from three- option versions of multiple - choice questions are usually as robust as four or five option versions.

Ali, H. (2014). Characteristics of a good test: Published Oct. 13, 2014.

April, M. O. & Kristel, A M. (2019). Analysis of multiple-choice questions (MCQs): Item and test statistics from the 2nd year Nursing Qualifying Exam in a University in Cavite, Philippines.

Ascalon, M.E., Meyers, L.S., Davis, B .W. & Smits, N. (2007) Distracters similarity and item-stem

structure: Effects on item difficulty. Applied Measurement in Education, 20,153-163

Burud, I; Nagandla, K & Agarwal, P. (2019). Impact of distracters in item analysis of multiple choice questions. *International Journal of Research in Medical Sciences*

Callans, M. (2012). Validity and Reliability in Testing: Wonderlic Inc.

Department of Education. (2008). Plan and conduct Assessment. http://www.education.gov.za

Juliana, L.D. & Maria, L.V. (2017). Analysis of multiple choice questions: Item difficulty, discrimination index and distracter efficiency. DOI Number: 10.5958/097-

Linnette, J. D., Maria, L. V. & Harbi, F.A. (2018). Relationship between item difficulty and Non-functioning distracters of Multiple Choice Questions.

Mansour, S., Zarichehr, V., Negin, M. A., Manzardokht, B., Alireza, S., Mohammad, A. & Zohreh, A. (2012). An item analysis of written multiple-choice questions. https://doi.org.10.21010/ajtcam.v14i4.13

Monday, T. J. (2009). *Fundamentals of test and measurement in education (3rd Edition).* The University of Calabar Press.

Mozaffer, R. H., & Farhan, J. (2012). Analysis of one-best multiple-choice questions. *Journal of Pakistan Medical Association, 62, 142–147.*

Nitku, A.J. & Brookhart. S.M. (2011). *Educational assessment of students (6th Edition*). Boston, MA: Pearson.

Nomdeo, S. K. & Bandya, S. (2015). Item analysis of multiple choice questions from an assessment of medical students in Bhubaneswar, India

Odinko, N. (2014). *Evaluation, research theory and practice*: Nigeria: Ibadan: Giraffe.

OECD (2012): The Assessment of Higher Education Learning Outcomes http://www.oecd.org/education/imhe/theassessmentofhigher

Onuka, A. O. U. & Durowoju, E. O. (2013). Stakeholders' role in curbing examination malpractice in Nigeria. *International Journal of Economy Management and Social Sciences, 2(6)*

Pradip, C., Girish, R. C., Bhoomika, R. C., Jayesh, V. V. & Suresh, P. R. (2015). Relationship between difficulty index and distracter effectiveness in single best-answer stem type multiple choice questions. *International Journal of Anatomy Research*.

Rodriguez, M. C., Ketller, R. J. & Elliott, S. N. (2014). Distracter functioning in modified items for test accessibility. First Published October 7, 2014

Tarrant, M., Ware, J., & Mohammed, A.M. (2009). An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. BMC Medical Education, 9, Article 40. https://doi.org/10.1186/1472-6920-9-40

Tsegaye, K.N. (2019). Post-exam analysis: Implication for intervention. http://dx.doi.org/10.1101/510081 University of Virginia 2019