

## **Item Analysis in Teacher-Made Test: Key to Assurance of Quality in Nigerian Educational System**

**Aliyu A. Zakariyya, Ph.D., Abubakar S. Ndagara,**  
Department of Mathematics,  
Niger State College of Education, Minna, Niger State.

**&**

**Hadijat O. Sadiq,**  
Department of Science Education, Ibrahim Badamasi Babangida  
University, Lapai, Niger State.

### **Abstract**

One of the methods of assessing the quality of educational provisions is the performance of learners in schools. Test or examination is one of the major ways of evaluating the learners' performance. Tests play a very important role in the educational system as they are used as tools in measurement and evaluation processes. Test as a tool of measurement process must then be qualitative. The quality of test itself can be determined through test item analysis. An item analysis provides three kinds of important information about the quality of test items. Item difficulty: A measure of whether an item was too easy or too hard. Item discrimination: A measure of whether an item discriminated between learners who knew the material well and learners who did not. Effectiveness of alternatives: Determination of whether distractors (incorrect but plausible answers) tend to be marked by the less able students and not by the more able students. The world is regarded as a global village. As such within the teaching profession, the continent and the entire world are equally forging common fronts to see that the teaching profession has common benchmark, so that teachers' made test and competencies among others are comparable across countries. This paper therefore, discussed concept of item analysis and how indices such as difficulty index, discriminating index and distractors index can be calculated. The paper finally recommended among others, that teachers should not only strive to provide teaching material and to test the students' comprehension about it, but also be able to analyse the quality of their test.

### **Introduction**

Teachers at all levels of education are confronted with multifarious tasks. These tasks include planning of lesson and presentation, classroom management, assessment of learning and reinforcement of learning which is all inherent to the teaching profession. Among these many tasks that a teacher need to perform proficiently in the classroom is the assessment of students' learning in the form of teacher-made test. Classroom assessment plays an important role in determining the quality of the input process and output elements of education. They can be used to reach important decisions about the students' grades or class of degree. Classroom assessment tools constructed by many teachers are either formative (periodic tests) or summative tests (end of the term, session or semester examinations).

Teacher-made tests play a very significant role in that they are part of the teaching and learning process. Information from the test or examination data can be employed in placing students, formative evaluation, diagnostic evaluation, making selection decisions, arriving at curricular decisions, personal decision making and award of certificate. Teacher-made tests help the teacher to identify the content (knowledge or skills) which has been mastered by learners and the teacher knows through the results from his/her tests the areas where the learners have difficulties and thus, finds ways of overcoming the difficulties so that these learners can do better (Muzenda, 2017). Results from teacher-made tests can also enable the teacher to assess his/her strengths and weaknesses. Kiragu and Luke(2014) argue that teacher-made tests dominate the assessments used by teachers, regardless of test purpose, grade level, or subject area.

Development of valid and reliable test give teachers the necessary confidence that the decisions they made as a result of the tests are indeed pictures of the realities of their students and their classroom instructions. The results of the tests eventually describe actual competencies and mastery of knowledge of the learners. This parameter give teachers a feeling of confidence that the grades they give to their students

are reflections of their learned competencies since these are measured by valid and reliable tests. Now, the vital questions are how well does teacher-made tests measure what they intend to measure and how consistent they will remain if teachers do not apply item analysis and determine the validity and reliability of tests?

In most cases, item analysis, which includes determination of item difficulty and discrimination, as well as distractor analysis are not done because it is time consuming and demanding if done manually. The above view was also shared by Lebagi&Sudjoko(2017) when they said that the quality of test itself can be determined through test item analysis. They further observed that in many occasions, teachers tend to ignore test item analysis because of time limitation and other responsibilities. Other reasons may include the large students' population and in some cases lack of knowledge of the processes involved. With reference to this problem, the paper aimed to discuss the concept of test items analysis and the procedure involve in the determination of difficulty index, the discrimination index, and the distractor index as key indicators for ensuring quality in teacher-made tests.

### **Concept of Quality**

All over the world there is an increasing interest in quality and standards of education. If Nigeria is to achieve its aspiration to be one of the most dynamic and knowledge-based economy in the world, then Nigerian educational system will need to demonstrate that it takes the quality of its educational programmes and awards of certificates seriously and should be willing to put into place the means of assuring and demonstrating that quality. Quality can be seen as the standard of something when it is compared to other things like it, to see how well or bad the something is.

According to Thom-Otuya, and Inko-Tariah (2016) quality can be defined in terms of excellence, perfection, value for money and fitness to purpose.

Quality education entails among others the learning resources, technology, program enrolled, modules done,

lecturing methodology, learning environment, teachers or lecturers qualifications and dedication, performance or certificate awards and assessment procedure. According to Obasi (2010) the indicators of quality and functional education among others include: Effective and efficient performance of graduates in society, industries and other work places; International transferability/admissibility of graduates/ students for higher studies without remedial conditions. However, the quality of education provided by institutions may vary based on factors such as, the teaching system, programs provided, quality of lecturers, and the learning environment under which the education is being provided.

It should be noted that education produces all kinds of manpower such as: engineers, teachers, lawyers, medical doctors, architects, soldiers and scientist needed for national development. It is the product of education that transforms the society. Therefore, the quality of education determines the quality of manpower and their products. It has been argued earlier that assessment plays an important role in determining the quality of the input and output elements of education, and test plays important role as the tool in measuring the success of teaching and learning activities.

### **Concept of Item Analysis**

The procedure in teaching and test administration requires for a good teacher to provide teaching material, assess students' comprehension and also to analyse the quality of test items. To know the quality of test items, there is the need to conduct test item analysis in order to provide information about students' strengths and weaknesses, and also to determine the quality of the item itself. Item analysis is a process which examines students' responses to individual test items (questions) in order to assess the quality of those items and of the test as a whole. Anna and Michela (2011) defined item analysis as a process in which both students' answers and test questions are examined in order to assess the quality and quantity of the items and the test as a whole. Item analysis is especially not only valuable in improving items which will be used again in later tests, but it

can also be used to eliminate ambiguous or misleading items in a single test administration.

Item analysis can be either qualitative or quantitative analysis. Qualitative item analysis includes going over and analysing the test by the teacher or expert prior to its administration in order to locate any typographic or similar errors, any grammatical hints that may be involuntarily hidden in the text of the test and evaluation of text for its reading level and the appropriateness of its use. Qualitative analysis can also include consideration of content validity, as well as the evaluation of items in terms of effective item writing procedures (Evrero, 2015). Quantitative item analysis is generally based on the factors obtained statistically. Quantitative analysis includes principally the measurement of item difficulty, item discrimination and effectiveness of the distractors. This paper focused on the second dimension of item analysis namely the quantitative analysis of test.

### **Relevance of Item Analysis to Teachers**

In several occasions, there are some teachers who can provide good teaching material as well as construct good test but they do not possess the knowledge to analyse the quality of the test items so as to assess its level of difficulty, discrimination power, and the functionality of the distractors. Anna and Michela (2011) posited that lack of awareness about the importance of test item analysis was responsible for non-adherence to its practice. However, Sanju, Rashmi, Pradeep and Manish (2014); Evrero (2015); Sahoo and Singth (2017) have found the following as uses of item analysis to teachers:

- The result of test analysis can be a stepping stone for them in improving their test items as well as informing them about the quality of their teaching and learning.
- Item analysis allowed teachers to observe the item characteristics, and to improve the quality of their tests. Item analysis provides diagnostic information for determining the quality of the test terms.
- Item analysis is valuable for increasing teachers' skills in test construction, and identifying specific areas of course

- content which need greater emphasis or clarity.
- Item analysis data can provide a basis for efficient class discussions of the test results. When teachers know how effectively each item functioned in measuring achievement, it will enable them to confine the discussion to those areas which will be most helpful to learners.
  - Item analysis permits teachers to identify test items that are too difficult or too easy, and to differentiate between students who have learned the content and those who have not.
  - Improving the skills in the test through item analysis can save teachers' time and energy.

### **Problems of Teacher-made Test**

Teacher-made tests are usually criterion referenced tests that are designed to assess students' mastery of a specific body of knowledge. According to Omoruan (2018) teacher-made test is a test instrument constructed by the classroom teacher to measure the extent of students' achievement of a certain class based on some specific objective. Teacher-made tests are normally prepared and administered for testing class-room achievement of students, evaluating the method of teaching adopted by the teacher and other curricular programmes of the school or institution.

Research indicates that teacher-made tests dominate the assessments used by teachers, regardless of test purpose, grade level, or subject area. However, this most valuable instrument in the hands of the teachers suffered some flaws. These include according to studies of Lebagi and Sudjoko (2017); Omoruan (2018); Hartell and Strimel (2019):

- i. Ambiguity in the words and sentence structure of some items
- ii. Improper arrangement of the items
- iii. Lack of strict adherence to the principles of test construction, test administration and analyses
- iv. The same person served as instructor, test writer, and evaluator

- v. The quality of test items, the reliability of test measures, the procedures for administering and scoring and the interpretation of scores are questionable
- vi. They usually measure only a limited part of a subject area, they do not cover a broad range of abilities and they rely too heavily on memorized facts and procedures.

### **Indices of Item Analysis**

An item analysis provides three kinds of important information about the quality of test items, these includes:

- Item difficulty: A measure of whether an item was too easy or too hard.
- Item discrimination: A measure of whether an item discriminated between students who knew the material well and students who did not.
- Effectiveness of alternatives: Determination of whether distractors (incorrect but plausible answers) tend to be marked by the less able students and not by the more able students (Ibrahim & Zakariyya, 2013 & Baji, & Wachiko, 2015).

### **Optimal item difficulty/ Difficulty index**

Item difficulty index is the percentage of the total group that got the item correct. Item difficulty is important because it reveals whether an item is too easy or too difficult. In either case, the item may add to the unreliability of the test because it does not aid in differentiating between those students who know the material and those who do not. For example, an item answered correctly by everyone does nothing to aid in the assignment of grades. The same is true for items that no one answers correctly. If the difficulty index of an item is high, it shows that the item is easy; however, if the difficulty index of an item is low, it shows that the item is difficult.

**Difficulty index=**

$$\frac{\text{number of testees who answered an item correctly}}{\text{Total number of testees}} \quad 100$$

The ideal difficulty index is 50% however, indices between 30%-70% are accepted to be good. Items with difficult indices above 70% are considered too easy, while those below 30% are too difficult (Ibrahim & Zakariyya, 2013; Musa, Shaheen, Elmardi, & Ahmed, 2018; Denis, Cherly, Paul & Lester, 2019).

### **Item Discrimination/Discrimination Index**

Item discrimination compares the number of high scorers and low scorers who answer an item correctly. It is the extent to which items discriminate among testees in the high and low groups. High performers should be more likely to answer a good item correctly, and low performers more likely to answer it incorrectly. The discrimination index of an item is the ability to distinguish high and low scoring learners. The closer this value is to 1, the better the item distinguishes the learners who get a high score from those who get a low score.

Item discrimination or the discriminating power of a test item refers to the degree to which success or failure on an item indicates possession of the ability being measured. It determines the extent to which the given item discriminates among examinees in the function or ability

measured by the item. This value ranges between -1.0 and 1.00. The higher the value, the more discriminating the item is. A highly discriminating item indicates that the students who had high tests scores got the item correct whereas students who had low test scores got the item incorrect.

$$\text{Discrimination index} = \frac{GU - GL}{T}$$

where

GU= number of testee from the upper half who got the item correctly

GL= number of testee from the lower half who got the item correctly

T= total number of the testees (Ibrahim & Zakariyya, 2013).

Boopathiraj and Chellamani (2013) suggested identification of upper 27% and lower 27% examinees having highest and lowest scores in rank order respectively on the total test. Discrimination



index is classified according to (Sahoo& Singh,2017) as:

- 0.40 and above: very good items,
- 0.30–0.39: reasonably good,
- 0.20–0.29: marginal items (i.e. subject to improvement), and
- 0.19 or less is poor items (i.e. to be rejected or improved by revision).

When an item is discriminating negatively, it shows that the more knowledgeable examinees are getting the item wrong and the less knowledgeable examinees are getting the item right. A negative discrimination index may indicate that the item is measuring something other than what the rest of the test is measuring. If an item has discrimination below 0.0, that is negative value, it suggests a problem. More often, it is a sign that the item has been mis-keyed (Boopathiraj & Chellamani, 2013).

### **Distractor Index**

A standard multiple-choice test item consists of two basic parts: a problem (stem) and a list of suggested solutions (alternatives). The stem may be in the form of either a question or an incomplete statement, and the list of alternatives that contain one correct or best alternative (answer) and a number of incorrect or inferior alternatives (distractors). The purpose of the distractor is to appear as plausible solutions to the problem for those students who have not achieved the objective being measured by the test item. Conversely, the distractors must appear as implausible solutions for those students who have achieved the objective (Sanju, Rashmi, Pradeep& Manish, 2014). The quality of a multiple-choice question is heavily dependent on its distractors. It is, therefore, important for teachers to observe how many students select each distractor and to revise those that draw little or no attention.

### **To Calculate Distractor Index:**

First find the total scores for every testee. Second divide the testees in to three groups in rank order, the upper 27% and the lower 27% and the middle 46%. Third calculate the percentage of testees who choose each alternative (a,b,c and d)

in the upper and lower groups. Non-functional distractor (NFD) in an item is option (other than key) selected by <5% of testees; alternatively, functional or effective distractors are those selected by 5% or more participants (Hingorjo&Jaleel, 2012; Mahjabeen,etal, 2017).

If there is distractor that no one chooses in the lower group, it should be omitted. If there is distractor the upper group chooses more than the lower group, it should be omitted also. If there is distractor the upper group chooses more than the right alternative, the answer key should be revised, if it is right this distractor should be omitted. Whenever the proportion of examinees who selected a distractor is greater than the proportion of examinees who selected the key, the item should be examined to determine if it has been mis-keyed or double-keyed (Boopathiraj & Chellamani, 2013).

In addition to being clearly incorrect, the distractors must also be plausible. That is, the distractors should seem likely or reasonable to an examinee that is not sufficiently knowledgeable in the content area. Mark, Okan, Guo& Zhang (2017) viewed that distractors create an important part of the context required to solve a multiple-choice item that can affect item quality and learning outcomes. Within this context, a complex relationship exists between the correct and incorrect options due to the fact that students are required to make a distinction among response options in order to select the correct response. Therefore, distractors should share important information with the correct option.

### **Conclusion**

Teachers should learn to construct tests which are valid and reliable so that the grades they give to their students will speak of their actual performance in class. By looking at the pattern of responses to distractors, teachers can often determine how to improve the test. Construction of items in a test should be made with care such that it won't be too easy or too difficult for the students to answer. Also, options to be given in a multiple choice test should be made plausible to identify those who are studying and guessing during the test. A complex relationship

should also exist between the key and distractors. Identifying test items which are retained, revised and rejected will be very helpful for teachers to come up with a test bank. These test items may be modified or used for other school years because these are good items. Failing to use qualitative and quantitative item analysis may lead to the invalid assessment results, influencing students' academic success. Finally, teachers are enjoined to do item analysis of their tests in order to improve their quality. This ensures all the students have equal chances for obtaining the highest score on the test according to their knowledge and skills and unbiased by any subjective factors.

### **Recommendations**

In view of the relevance of item analysis in ensuring quality assurance in teacher-made test, the following recommendations are offered:

1. Schools, educational institutions and examination bodies should organise trainings, workshops and sensitization for teachers and major stakeholders in the education sector on assessment, test construction and item analysis for efficiency in teaching and learning process.
2. Teachers should combine qualitative and quantitative item analysis to allow efficiency in teaching and learning.
3. With the increase on Computer Based Test (CBT) examinations by various examination bodies and tertiary institutions in the country, item analysis software should be used to facilitate the analysis.
4. Where the item analysis is done by somebody other than the subject teacher or course lecturer, its result should be communicated to the individual teacher or lecturer concern for improvement in future tests.

## References

- Anna, S. & Michela, F. (2011). The use of item analysis for the improvement of objective examinations. *Procedia - Social and Behavioural Sciences*. 29 (1), 188 – 197.
- Baji, M. I. & Wachiko, B. (2015). *Elements of measurement, evaluation and statistics*. Minna: Havilah Kingdom Press.
- Boopathiraj, C. & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International Journal of Social Science & Interdisciplinary Research*. 2 (2), 189-193.
- Denis, A. T., Cherly, C. C., Paul, B. S., & Lester, B. S. (2019). Development of valid and reliable teacher-made tests for grade 10 Mathematics. *International Journal of English and Education* 8(1), 62-83.
- Evrero, E. S. (2015). Item analysis of test of number operations. *Asian Journal of Educational Research*, 3(1), 18-25.
- Hartell, E. & Strimel, G. J (2019). What is it called and how does it work: examining content validity and item design of teacher made tests. *International Journal of Technology and Design Education*, 29:781–802
- Hingorjo, M. R. & Jaleel, F. (2012). Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc*. 62(2), 142-147.
- Ibrahim, M.O. & Zakariyya, A. A. (2013). *Measurement and evaluation in education*, Zaria, Nigeria: Stevano Printing Press.
- Kiragu, K. & Luke, O. (2014). Validity and reliability of teacher-made tests: Case study of year 11 physics in Nyahururu District of Kenya. *African Educational Research Journal* 2(2), 61-71.
- Lebagi, D. & Sudjoko, S. (2017). The quality of teacher-made test in EFL classroom at the elementary school and its wash back in the learning. *Journal of English Education*, 2(2), 97-104.
- Mahjabeen, W., Saeed, A., Hassan, U., Zafari, T., Bult, R., Sadaf, K., & Rizvi, M. (2017). Difficulty index, discrimination index and distractor efficiency in multiple choice questions. *Pak. Inst. Medical Science*, 310-315.
- Mark J. G., Okan B., Guo, Q., & Zhang, X. (2017). Developing, analysing and using distractors for multiple-choice tests in education: A comprehensive review. *Review of Educational Research*. 87(6), 1082 – 1116.
- Musa, A., Shaheen, S., Elmardi, A., & Ahmed, A. (2018). Item difficulty & item discrimination as quality indicators of physiology MCQ examinations at the Faculty of Medicine Khartoum University. *Khartoum Medical Journal*. 11(2), 1477 – 1486.

- Muzenda, D. (2017). Effectiveness of teacher-made tests in Goromonzi district primary Schools. *IJRDO-Journal of Educational Research*, 2(2), 53-65.
- Obasi, O. (2010). Qualitative and functional education as a social imperative for authentic National development: The Nigerian situation. *School of Business and Management Technology Conference Journal*, 1(1): 5-14.
- Omoruan, B. E. (2018). Valid teacher-made tests: It's implication on students' achievement in Business education. *ATBU, Journal of Science, Technology and Education*, 6(2), 65-77
- Sahoo, D. P. & Singh, R. (2017). Item and distracter analysis of multiple choice questions (MCQs) from a preliminary examination of undergraduate medical students. *International Journal of Research in Medical Sciences*. 12(5), 5351-5355.
- Sanju G., Rashmi S., Pradeep K. & Manish, R. (2014). Item and test analysis to identify multiple choice questions (MCQs) from an assessment of Medical Students of Ahmedabad, Gujarat. *Indian Journal of Community Medicine*, 39(1), 17-20.
- Thom-Otuya, B. E. N. & Inko-Tariah, D. C. (2016). Quality education for national development: The Nigerian experience. *African Educational Research Journal*, 4(3): 101-108.