Comparism of Item Difficulty and Discrimination of Pre And Post University Entrance Examinations in Nigeria

Adamu Chidubem Deborah [1], Babatimehin Temitope [2], Adeoye Oluseyi Peter [3]

ARTICLE INFO	**ABSTRACT**

The study compared the difficulty and discrimination indices of the Pre-University Tertiary Matriculation Examination (Pre-UTME) and Post-University Tertiary Matriculation Examination (Post-UTME) in the related subject areas of Arts, Commercial and Science in Obafemi Awolowo University (O.A.U.), Ile-Ife, Nigeria. The study adopted the exploratory survey research design. The population for the study comprised all part one undergraduate students of the O.A.U., Ile-Ife in 2013/2014 session. While simple random sampling technique was used to select 40 departments, convenience sampling technique was used in drawing 20 part one undergraduate students from each of the 40 departments. A 20-item achievement test titled: 2013/2014 Pre-UTME and its corresponding 2013/2014 O.A.U. Post-UTME were instruments used for collecting data for this study. Data collected were subjected to item analysis. Results revealed that the items of the O.A.U.'s 2013/2014 in the subject areas of Arts, Commercial and Science (with average difficulty indices, p, of 0.53, 0.46 and 0.55 respectively) were more difficult than items of the 2013/2014 Pre-UTME in the same subject areas (having average p of 0.62, 0.54 and 0.60 respectively). Finally, the results revealed that the items of the 2013/2014 Pre-UTME and O.A.U.'s Post-UTME in the subject areas of Arts, Commercial and Science discriminated appropriately between students in the high scoring group and those in the low scoring group with average discrimination indices of 0.45, 0.43, 0.46 and 0.46, 0.43, 0.42 respectively.

INTRODUCTION

Examinations could be internal or external. It could be oral, written or both (Ethe and Odjegba, 2019, p. 194-205). Internal examinations include continuous assessment tests, terminal, semester and annual or promotion examinations (Osunde, 2009). Examples of external or public examinations conducted in Nigeria are secondary school certificate examinations (internal and external) conducted by the West African Examination Council (WAEC) and National Examination Council (NECO) (Falaye and Eze, 2015, p. 1-9). NECO conducts other examinations such as National Common Entrance Examination (NCEE) for admission into secondary schools, Junior School Certificate Examination (JSCE), the Gifted Examination into Federal Academy, Suleja (Olatgoye, n. d.). Other public examinations in the country include examination into the higher institutions called Pre-Unified Tertiary Matriculation Examination (Pre-UTME) conducted by the Joint Admissions and Matriculation Board (JAMB) and Post-University tertiary matriculation Examination (Post-UTME) conducted by federal, state and private universities in the country. National Teachers' Institute (NTI) conducts admission tests into tertiary institutions while the National Business and Technical Examination Board (NABTEB) conducts professional examinations for teachers and technicians respectively.

Also, Teachers' Registration Council of Nigeria (TRCN) is saddled with the responsibility of registering and conducting professional and qualifying examination to teachers in Nigeria and issuing teaching license to them which is renewable in every two years. For any candidate who sits and passes the Pre-UTME and obtains the cut-off mark or above the cut-off mark of his or her choice of the federal, state or private university registered for in the JAMB will be considered and invited to write the Post-UTME (i.e, screening examination conducted by individual universities in Nigeria before the offer of tentative admission for those who qualify or pass) (Falaye and Eze, 2015, p. 1-9). The Pre-university entrance examination in Nigeria is conducted by the Joint Admissions and Matriculation Board (JAMB) (JAMB, 2013). The JAMB is an examination body saddled with the responsibility of registering and conducting entrance examination into the federal, state and private tertiary institutions in Nigeria (Nnaka and Anaekwe, 2011, p. 86-91). Earlier than 2005 when the Post-UTME was introduced by the federal and state universities in Nigeria (private universities joined in the screening exercise later), it was a difficult task buying JAMB forms and registering for the examination as well as frustrating and tiring to get the results when they were ready (Chika and John, 2010, p. 101-106). This difficulty was accounted for in the late release of results by the JAMB which compelled some universities to allow their first year students to resume lectures very late than planned in their school calendars. The quality of education fell drastically due to the production of low quality graduates by the universities in Nigeria. Majority of these graduates participated in examination

1 dubem_deby@yahoo.com, Department of Educational Foundations and Counselling, Faculty of Education, Obafemi Awolowo University, orcid.org/0000-0002-4003-0808
2
3 seyipeter680@gmail.com, Department of Educational Foundations and Counselling, Faculty of Education, Obafemi Awolowo University, orcid.org/0000-0002-5387-8031

malpractices throughout their stay in the universities. The rate of examination malpractice, failure and low productivity among these undergraduates and graduates became alarming such that the former Nigerian President, Obasanjo accuses JAMB of corrupt practices (Ade, 2006, p. 7), which he said had affected the standard of education in Nigeria. As a result, the UTME ceased to be the only yardstick to be relied upon for the true reflection of candidates' performance for entry into the Nigerian universities. This led the former President to recommend further screening for candidates' who wrote and passed (i.e., for those who obtained the cut-off point set by the JAMB) the UTME conducted by the JAMB. This further screening is called Post-UTME. The P-UTME is a screening or an internal examination organized by the individual universities after JAMB had selected the candidates for admission. According to the then Nigerian President, the P-UTME would restore the past glory of tertiary education in the country and would make university education only for those who want it and use it. This is based on the belief of Ebiri (2006, p. 49) that a nation without a strong educational base is bereft of hope and future. All the individual universities in Nigeria under the direction of the federal government organized their first admission exercise for the students who were admitted in the 2005/2006 session. The P-UTME exercise was faced with many oppositions. These oppositions came from the students, parents and wards who could not guarantee their admissions. The P-UTME exercise was also opposed by the JAMB executives as they feared that their functions were being threatened and undermined. Opposing the P-UTME exercise was also the JAMB workers who felt their avenue of making quick money was being hampered. These complaints and oppositions still continue and on the high side till date as majority of the stakeholders in education have failed to discriminate between candidates admitted through the UTME (JAMB) only and the ones admitted through both the Pre-UTME (JAMB) and Post-UTME (screening examination). For example, students, wards and parents envisage the Post-UTME as an additional burden on parents to pay for another admission exercise after paying for the Pre-UTME and also as more stress on the students seeking admission into the universities (Amatareotubo, 2006).

Although Post-UTME was introduced into Nigerian universities to help ameliorate lots of problems in the system, several problems were identified by stakeholders in the country's university education (Ado, Hadiza and Samira, 2016, p. 261-270). Several studies have confirmed that the contents of Pre-UTME and Post-UTME items are not similar and that items of the Post-UTME are more difficult than items of the Pre-UTME (Faleye and Eze, 2015, p. 1-9). Item difficulty determines how difficult or easy an item is (Aduloju and Kpum, 2019, p. 139-151). Item difficulty is the percentage of learners who responded correctly to an item and ranges from 0.0 to 1.0. The closer the difficulty of an item approaches zero, the more difficult that item is (Aduloju and Kpum, 2019, p. 139-151). Emaikwu (2011) submitted that in practice, item difficulty index within the range of 0.3 to 0.7 is acceptable. Item difficulty can and should be calculated and analyzed for all types of questions, not just true-false and other forced-choice items (Afolabi and Dibu-Ojerinde, 2012). To compute the item difficulty, divide the number of people answering the item correctly by the total number of persons answering the item. The proportion for the item is usually denoted as p and is called item difficulty (Crocker and Algina, 1986). An item answered correctly by 75% of the examinees would have an item difficulty or p value of 0.75 where as an item answered correctly by 50% of the examinees would have a lower item difficulty or p value of 0.5 (Afolabi and Dibu-Ojerinde, 2012).

A p value is basically a behavioural measure. Item difficulty has a profound effect on both the variability of test scores and the precision with which test scores discriminate among different groups of examinees (Thorndike and Hagen, 1977). When all of the test items are extremely difficult, the great majority of the test scores will be extremely high. In either case, test scores will show very little variability. Thus, extreme p values directly restrict the variability of test scores (Iweka, 2014). In the example below, five true-false questions were administered to a class of 40 students. For each question, the number of students answering correctly was determined and then converted to the percentage of students answering correctly.

Question	Correct Responses	Process	Item Difficulty	Item Difficulty Index
1	30	30/40	75%	0.75
2	34	34/40	85%	0.85
3	12	12/40	30%	0.30
4	26	26/40	65%	0.65
5	40	40/40	100%	1.00

In practice, approximately half of the students should answer true-false questions correctly even if they do not know the material because each student has a 50/50 chance of getting the correct answer by guessing (Suruchi and Rana, 2014, p. 56-58). If the students simply did not know the material, approximately half would select the correct answer by chance alone, resulting in an item difficulty close to 50%. The difficulty index of the five true-false questions above can be computed by dividing the number of total students (40) who participated in the test. Using this formula, the difficulty of question 1 is 30/40 which is equal to p value of 0.75. A rough "rule of thumb" is that if the item has p value of 0.70 and above it is an easy item, if the difficulty is 0.30 and below, it is a difficult item. Given these parameters, questions 1 and 2 are regarded as easy items hence, 75% and 85% of the students answered them correctly respectively. Item 4 is moderately easy with 65% (26/40 = 0.65 p value) of the students answering it correctly. Item 5 is not a very good item. It is extremely easy as all the students (100% = 40/40 with p value of 1.0) answered it correctly and should be discarded hence, it shows mastery of what the teacher expected from the students. Item 3 is extremely difficult as only 30% (12/40 = 0.3 p value) of the students answered it correctly.

Since only 30% of the students answered question 3 correctly in the example above, there is an indication of a problem. It could be that the question was not well written making misleading and encouraged students to select the incorrect answer. Probably also, the teacher may not have taught or covered the material as thoroughly as she or he had wanted to do (Afolabi and Dibu-Ojerinde, 2012). Discussion of the question with the students can probably clarify the situation. If students do not know the content, some re-teaching or review is probably in order. For questions in which students construct their responses (short answer, completion, essay, etc.) examining the incorrect answers given by students may also be helpful in determining the problem.

The discrimination index of an item is the ability of the item to distinguish or discriminate between high and low scoring learners. Computing the item discrimination index enables a teacher to determine whether the question discriminates appropriately between lower scoring and higher scoring students. When students who earn high scores are compared with those who earn low scores, we would expect to find more students in the high scoring group answering a question correctly than students from the low scoring group. In the case of very difficult items which no one in either groups answered correctly or fairly easy questions which even the students in the low group answered correctly, the number of correct answers might be equal for the two groups. What one would not expect to find is a case on which the low scoring students answered correctly more frequently than students in the high group (Faleye and Eze, 2015, p.1-9).

In computing the discrimination index, D, first score each student's test and rank order the test scores. Next, the 27% of the students at the top and the 27% at the bottom are separated for the analysis. Wiersma and Jurs (1990) stated that "27%" is used because it has shown that this value will maximize differences in normal distributions while providing enough cases for analysis.

There need to be as many students as possible in each group to promote stability, at the same time it is desirable to have the two groups be as different as possible to make the discriminations clearer. According to Kelly (as cited in Popham, (1981) the use of 27% maximizes these two characteristics. Nunally (1972) suggested 25%. The discrimination index, D, is the number of people in the upper group who answered the item correctly, divided by the number of people in the largest of the two groups. According to Wood (1980) when more students in the lower group than in the upper group select the right answer to a question, the question actually has negative validity. Assuming that the criterion itself has validity, the item is not only useless but is actually serving to decrease the validity of the test.

The item discrimination index, D can vary from -1.00 to 1.00. A negative discrimination index (between -1.00 and zero) results when more students in the low group answered correctly than students in the high group. A discrimination index of zero means equal number of high and low students answered correctly, so the item did not discriminate between groups. The closer the value of the discrimination index is to 1, the better the item distinguishes the learners who get a high score from those who get a low score. Hotiu (2006) stated that, items with discrimination indices of 0.3 and above are considered good items while items with discrimination indices less than 0.3 are considered poor items and subject to revision. A positive discrimination index occurs when more students in the high group answer correctly than the low group. A positive discrimination index occurs when students in the high group answer correctly than the low group. If the students in the class are fairly homogeneous in ability and achievement, their test performance is also likely to be similar, resulting in little discrimination between high and low groups. Questions having an item difficulty index (not item discrimination) of 1.00 or 0.00 should not be included when calculating item discrimination indices (Afolabi and Dibu-Ojerinde, 2012). An item discrimination of 1.00 indicates that everyone answered correctly while 0.00 means no one answered correctly. Therefore, neither type of item discriminates between students. Such questions serve no purpose because they simply provide evidence of what the teacher probably already knows about the students. The higher the difficulty index, the better the item because such a value indicates that the item discriminates in favour of the upper group, which should get more items correct (Boopathiraj & Chellamani, 2013, p. 189-193).

A negative discrimination index is most likely to occur when an item covers a complex material written in such a way that it is possible to select the correct response without any real understanding of what is being assessed (Afolabi and Dibu-Ojerinde, 2012). A poor student may make a guess, select that response, and come up with the correct answer. Good students may be suspicious of a question that looks too easy, may take the harder path to solving the problem, read too much into the question, and may end up being less successful than those who guessed. As a rule of thumb, in terms of discrimination index, 0.40 and greater are very good items, 0.30 to 0.39 are reasonably good but possibly subject to improvement, 0.20 to 0.29 are marginal items and need some major revision or should be eliminated (Ebel and Frisbie, 1986).

In the example below, there are 16 students in the high group and 16 in the low group (with 24 between the two groups which are not represented). For question 1, all 16 students in the high group answered correctly, while only 8 in the low group answered correctly, thus H–L (16–8) = +8. The last step is to divide the +8 by the number in each score group (16) to yield +0.5. Simply looking at the number of correct answers in the two groups (16 in the high group and 8 in the low group) indicates that the question discriminates in the desired direction, and the positive discrimination index (+0.5) verifies this (Afolabi and Dibu-Ojerinde, 2012).

Item	Number in high answering correctly	Number in low answering correctly	Process (H–L)/16	Discrimination
1	16	8	(16–8)/16	0.5
2	15	5	(15–5)/16	0.625
3	10	12	(10–12)/16	–0.125

Questions 1 and 2 provide results in the expected direction. One would expect more of the students with high scores to answer correctly than students with low scores. When this situation occurs we have a positive discrimination index. Question 3, however, shows an unexpected result, with more of the low scorers getting the question right than high scorers. For question 3 in the example above, subtracting L from H gives a negative number. Dividing by the number in the groups gives a negative index. This is a cue that there may be a problem with the way the question was presented in the test or the way the material was taught (or not taught) (Quaigrain and Arhin, 2017, p. 1-11).

Analysis of each item by calculating difficulty and discrimination indices provides feedback on what the learners have learned and enables instructors (teachers) to determine and correct the faulty items. Ado,

Hadiza and Samira (2016, p. 261-270) argued that if there is a high correlation between the difficulty index of an item and its discrimination index, the item is considered to be good. This means that the two parameters provide estimates that can be used in deciding the item quality of a test. In other words, this contributes to increasing the validity and reliability of the tests by revealing whether the items are working well or not. In this wise, lecturers in the federal, state and private universities in Nigeria can remove faulty items from the pool of Post-UTME items or change the items, that is, if their difficulty and discrimination indices do not correspond with that of the Pre-UTME (JAMB). However, this study seeks to compare the item difficulty level and discrimination strengths of the Pre-UTME (JAMB) and Post-UTME (screening examination conducted by the individual universities in Nigeria).

Situation of the Problem

Over the years till date, parents, students and the JAMB have raised critical issues and oppositions on the adequacy, cost, convenience and quality of the Post-UTME in Nigeria. Most of these oppositions are geared towards the abolition of the Post-UTME. Of all these critical issues and oppositions, this article is however only interested in comparing the level of item difficulty and discrimination strength of the Pre-UTME and Post-UTME entrance examinations into the Nigerian universities.

Aim of the Study

Whereas some reports revealed that the item difficulty level and discrimination strength of the Pre-UTME (JAMB) correspond with the Post-UTME, some studies have revealed that they do not correspond. The purpose of this study therefore, is to compare the difficulty level and discrimination strength of the Pre-UTME (conducted by the JAMB) and Post-UTME (screening examination conducted by individual universities) entrance examinations in Nigeria. Two research questions were raised to guide the study. They are: 1) What is the difficulty level of the 2013/2014 Pre-UTME and its corresponding OAU's Post-UTME? 2) What is the discrimination strength of the 2013/2014 Pre-UTME and its corresponding

OAU's Post-UTME?

METHOD

This study employed the exploratory survey design. It involved a descriptive survey that is exploratory in nature. This is because the study was only interested in comparing the Pre-UTME and the O. A. U's Post-UTME entrance examinations. The survey was conducted on the basis of providing information on the comparism of both the pre and post university entrance examinations in Nigeria. In doing this, the item difficulty and discrimination of both tests in the same academic session were compared.

Study Group

All part one undergraduate students of the O.A.U., Ile-Ife in 2013/2014 academic session made up the population for the study. Ten faculties were randomly selected from the thirteen faculties in the O.A.U. Four departments were randomly selected from each of the ten faculties making a total of forty departments. A sample of 20 part one undergraduate students was conveniently (because the selected part one undergraduate students were those accessible and at the same time willing to participate in the study as of the time data were being collected) selected from each of the forty departments making a total of 800 undergraduate students of the O.A.U. The sampling procedure that was adopted in this study was the multistage technique.

Material

The instruments used for collecting data for this study were a 20-item achievement test titled: 2013/2014 Pre-UTME and its corresponding 2013/2014 O.A.U. Post-UTME. The type of question format that was employed in this study was the four-option multiple-choice format with response options labelled A, B, C and D. This is because it is easier and demands less time to complete. The Post-UTME served as an alternative confirmatory test to the Pre-UTME. The instruments for the study were validated by three experts in Educational Tests and Measurement and two subject specialists. The reliability indices of the instruments were determined using Kuder-Richardson formula 20 (KR20) and the reliability coefficient was 0.80. Data collection for this study lasted for a period of six weeks. This is because the 2013/2014 Pre-UTME was administered on the undergraduate students selected for this study in the first instance followed by the

2013/2014 O.A.U. Post-UTME after a period of two weeks. The responses of the examinees to the test items were scored and coded as "0" for not answering an item correctly and "1" for answering a question correctly.

Data Analyses

Data collected from the two instruments after coding were used for item analyses. For the difficulty indices, items with item facility ranging from 0.3–0.7 ($0.30 \leq p \leq 0.70$) were acceptable as being moderately difficult (good items) while items having item facility less than 0.3 and above 0.7 were considered faulty (difficult and easy) items. For the discrimination indices, items with indices of 0.3 – 0.8 were considered good items while items with discrimination indices less than 0.3 were considered faulty items.

FINDINGS

Research Question 1: What is the difficulty level of the 2013/2014 Pre-UTME and its corresponding O.A.U's Post-UTME?

To answer this question, the difficulty level of each of the 20 items that made up the Pre-UTME and Post-UTME in the three subject areas of Arts, Commercial and Science were determined after which the average difficulty level of the two tests were calculated for the three subject areas. This is because difficulty is item specific and the average difficulty level shows how the entire test functioned. The result is as presented in Table 1.

Table 1: Item Analysis between Difficulty Level of Pre-UTME and O.A.U's Post-UTME

Items	Pre-UTME Arts	Commercial	Science	Post-UTME Arts	Commercial	Science
1	56.67	46.67	63.33	36.67	43.33	76.67
2	53.33	63.33	53.33	60.00	53.33	56.67
3	76.67	10.00	40.00	63.33	16.67	33.33
4	63.33	93.33	33.33	60.00	66.67	30.00
5	36.67	76.67	46.67	33.33	53.33	46.67
6	70.00	63.33	52.94	63.33	60.00	76.67
7	53.33	36.67	86.67	43.33	36.67	76.67
8	73.33	26.67	53.33	46.67	6.67	50.00
9	53.33	76.67	50.00	53.33	60.00	40.00
10	66.67	76.67	73.33	40.00	46.67	60.00
11	53.33	46.67	56.67	73.33	60.00	50.00
12	73.33	33.33	70.00	60.00	43.33	50.00
13	60.00	56.67	53.33	73.33	63.33	43.33
14	53.33	36.67	86.67	43.33	43.33	86.67
15	60.00	33.33	73.33	43.33	16.67	76.67
16	66.66	56.67	60.00	56.67	60.00	56.67
17	73.33	60.00	66.67	60.00	60.00	43.33
18	72.59	68.23	84.92	69.31	65.47	78.67
19	56.67	36.67	53.33	70.00	36.67	50.00
20	73.33	93.33	70.00	50.00	50.00	33.33
Mean	61.50	53.67	59.65	53.00	46.33	54.83

From Table 1, it could be observed that while the average difficulty level of Arts for Pre-UTME and Post-UTME was ($p = 61.50$ and 53.00) respectively, the average difficulty level of commercial and science was ($p = 53.67$ and 46.33) and ($p = 59.65$ and 54.83) respectively. This showed that all the items of the Pre-UTME

and Post-UTME in the subject areas of Arts, Commercial and Science were moderately difficult except item 3 of the Commercial Pre-UTME and post-UTME that was difficult. Therefore, items of the Pre-UTME and Post-UTME were both moderately difficult.

Research Question 2: What is the discrimination strength of Pre-UTME and O.A.U's Post-UTME?

To answer this question, the discrimination index of each of the 20 items that made up the Pre-UTME and post-UTME of the O.A.U. in the three subject areas where determined after which the average discrimination index of the two tests were calculated for the three subject areas. The result is as presented in Table 2.

Table 2: Item Analysis between Discrimination Strength of Pre-UTME and O.A.U's Post-UTME

Items	Pre-UTME Arts	Commercial	Science	Post-UTME Arts	Commercial	Science
1	.47	.53	.43	.42	.47	.47
2	.53	.39	.46	.48	.49	.73
3	.47	.45	.67	.73	.39	.53
4	.40	.41	.40	.40	.40	.47
5	.41	.42	.40	..40	.40	.40
6	.47	.43	.49	.40	.43	.40
7	.47	.41	.41	.47	.33	.47
8	.47	.40	.43	.53	.41	.46
9	.40	.41	.47	.40	.43	.47
10	.43	.42	.44	.40	.44	.43
11	.47	.40	.47	.44	.43	.47
12	.40	.49	.47	.47	.40	.45
13	.47	.50	.53	.53	.40	.40
14	.47	.43	.47	.40	.43	.47
15	.48	.43	.40	.47	.44	.47
16	.49	.43	,44	.47	.49	.47
17	.40	.45	.46	.43	.50	.43
18	,40	.42	,47	.47	.47	.40
19	.43	.45	.49	.47	.40	.46
20	.47	.42	.40	.48	.49	.42
Mean	0.45	0.43	0.46	0.46	0.43	0.42

From Table 2, it could be observed that while the average discriminating strengths of Arts for Pre-UTME and Post-UTME were (D = 0.45 and 0.46) respectively, the average discriminating indices of Commercial and Science were (D = 0.43 and 0.43) and (D = 0.46 and 0.42) respectively.

The table further revealed that the discrimination indices of all the items of Arts, Commercial and Science subject areas of the Pre-UTME and Post-UTME were very good. From these result, it is seen that items of the Pre-UTME and Post-UTME in the Arts, Commercial and Science subject areas discriminated appropriately between low and high scoring students.

RESULT, DISCUSSION AND SUGGESTIONS

As revealed by the outcome of the difficulty investigation, almost all the items of the Post-UTME in Arts, Commercial and Science subject areas were more difficult than the items of the Pre-UTME. Question 3 in the Commercial subject area of both the Pre-UTME (having p value of

0.01) and Post-UTME (0.67 p value) was much difficult. Result therefore, revealed that the items of the 2013/2014 Post-UTME were more difficult than items of the Pre-UTME in the same session. As a result of

this, the percentage of students taking the Pre-UTME who answered the items correctly was higher than those taking the Post-UTME who answered the items correctly (Afolabi and Dibu-Ojerinde, 2012).

The discrimination strengths of the Pre-UTME and Post-UTME were presented for item analysis. The individual items in the subject areas of Pre-UTME and Post-UTME of the OAU on the average have equal mean scores. Because items of the Pre-UTME and Post-UTME of the OAU have equal mean score, result has shown that students who participated in the two tests of the same year achieved equal mastery of the items in the related subject areas.

In summary, items of the 2013/2014 Post-UTME were more difficult than items of the Pre-UTME of the 2013/2014. Items of the 2013/2014 Pre-UTME and Post-UTME of the O.A.U. in the subject areas of Arts, Commercial and Science discriminated appropriately between students in the high scoring and those in the low scoring groups. It then means that students with superior ability answered the questions in the two tests correctly more frequently than students with inferior ability. Therefore, the percentage of students taking the Pre-UTME who answered the questions correctly was higher than those taking the Post-UTME who answered the items correctly.

Suggestions

Based on the findings, results and discussion of this study, the following suggestions were made: In order to develop more quality test that can serve as a true screening examination to the Pre-UTME, it is necessary that lecturers perform a more robust item analysis on the Post-UTME items just like the JAMB does on the items of the Pre-UTME. 2) The Federal Government, JAMB and university lecturers should look at a way of merging the two examinations into one so as to ensure that only one scale of measurement is used for screening candidates for admission.

REFERENCES

Ado, A. B., Hadiza, H. & Samira, A. B. (2016). Evaluation of Northwest University, Kano Post UTME test items using item response theory. *International Journal of Evaluation and Research in Education (IJERE). 5, 261-270.*

Aduloju, M. O. & Kpum M. M. (2019). Item analysis and reliability of multiple-choice items of physics teacher-made test in Federal University of Agriculture, Makurdi, Benue State. *Nigerian Journal of Educational Research and Evaluation. 18(2), 139-151.*

Afolabi, E. R. I., & Dibu-Ojerinde, O. O. (2012). Educational tests & measurement. IleIfe, Nigeria: Obafemi Awolowo University Press.

Amatereotubo, M. (2006). "Post-UTME screening matters arising. *Retrieved from http://www.onlinenigeria.com/articles/*

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 and Interdisciplinary Research. 2(2), 189-193.*

Chika, J. A., & John, C. I. (2010). An assessment of post university tertiary matriculation examination: A case study of university of Benin. *Journal of Social Sciences. 22(2), 101-106.*

Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice Hall.

Ethe, N. & Odjegba, O. G. (2019). Comparism of WAEC and NECO SSCE mathematics multiple-choice test items parameters across school location: Application of item response theory. *ASSEREN Journal of Evaluation. 4(1), 194-205.*

Faleye, B. A. & Eze, C. D. (2015). Estimation of parallelism of UTME and Post-UTME among undergraduate students of Obafemi Awolowo University, Ile-Ife. *Ife Journal of Behavioural Research. 7(1 & 2), 1-9.*

Hotiu, A. (2006). The relationship between item difficulty and discrimination indices in multiple-choice tests in a physical science course. (Unpublished master's degree project) Florida Atlantic university, Florida. *Retrieved from: http://www.physics.fau.edu/research/education/A.Hotiu_thesis.pdf*

Iweka, F. (2014). Comprehensive guide to test construction and administration. Omoku: Chifas Nig.

Nnaka, C. V. & Anaekwe, M. C. (2011). Admission policy in science at tertiary education: Extent of goal attainment and imperatives for reforms in stem education. *52nd Annual STAN Conference Proceedings (p. 86-91).*

Nunnaly, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Osunde, A. U. (2009). Essay and multiple-choice tests: Bridging the gap. *Workshop papers on Multiple-choice Test item writing procedures for academic staff. University of Ilorin.*

Popham, W. J. (1981). *Modern educational measurement.* Englewood Cliff, NJ: Prentice-Hall.

Quqigrain, K. & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher developed test in educational measurement and evaluation. *Cogent Education, 4(2), 1-11.*

Suruchi, S., &Rana, S. R. (2014). Test item analysis and relationship between difficulty level and discrimination of test items in an achievement test in Biology. *Paripex-Indian Journal of Research. 3(6), 56-58.*

Thorndike, R. L. & Hagen, E. P. (1977). *Measurement and evaluation in psychology and education.* New York: John Wiley & Sons Press.

Wiersma, W. & Jurs, S. G. (1990). *Educational measurement and testing* (2nd ed.). Boston: Allyn and Bacon Press.