Psychometric Measurement of Accounting Multiple Choice Questions Achievement Test for the assessment of students with Special Needs

Basil C. E. Oguguo[1]
Lotobi R. Awele[2]
Clifford O. Ugorji[3]
Roseline N. Azubuike[4]
Catherine .U. Ene[1]
Felicia C. Ugwu[1]
Chinonso N. Ojinnaka[3]
Judith .N. Eke[5]
Chukwuka, A. Ihediohanma[6]
Zephyrinus I. Aririguzo[7]
Roseline I. Ezechukwu[8]

[1] Department of Science Education, University of Nigeria, Nsukka, Nigeria
[1] University of Benin, Edo State
[1] Imo State University, Owerri, Nigeria
[1] Department of Business Education, University of Nigeria, Nsukka, Nigeria
[1] Federal University Lokoja, Kogi State
[1] Department of Educational Psychology, Federal College of Education, Eha Amufu,
[1] Department of Political Science, Federal University of Kashere, Nigeria
[1] Department of Educational Psychology, Federal College of Education, Rivers State, Nigeria

Corresponding Author: Basil C. E. Oguguo Email: basil.oguguo@unn.edu.ng

Abstract
This study deterimned the psychometric properties of accounting multiple choice questions achievement test that will be used in the assessment of students with special needs. The study was an instrumentation research design. 200 students offering accounting participated in the study and were drawn using simple random sampling technique. Accounting Multiple Choice Questions Achievement Test (AMCQAT) was the instrument used for data collection. The instrument covers all the levels of behavoural objectives with 100 multiple choice test items developed by the researchers. Item analysis was carried out using Statistical Package for Social Sciences (SPSS version 23). The result showed that the achievement test is highly valid hence four out of the seven experts agreed that the instrument passed content validity while the rest of the three strongly agreed that the instrument is valid. This study is an improvement on similar instruments that are currently presenting challenges to students with special needs. The instrument developed from this study may be used to collect information about how students with special needs are performing and thus contribute to the description of educational phenomenon.

Keywords: Psychometric measurement**,** test item characteristics, Validation**,** Accounting achievement test, special education

DOI: 10.9756/INT-JECSE/V14I3.340

Introduction

Assessment of the accounting students' with special needs performance is an important part of the educational framework in Secondary Education in Nigeria. Assessment scores can be used for motivation, promotion, and feedback for areas of weakness in students (Black & William, 2009). Generally, assessments can be designed to measure three different levels of learning outcomes: knowledge, skills and attitudes. In this respect, the most common methods to assess accounting knowledge are: Multiple Choice Questions (MCQs), extended matching questions (EMQs) and short essay questions (SEQ). But the one used in this study is multiple choice questions constructed in Accounting. Accounting Multiple Choice Questions Achievement Test (AMCQAT) is an important tool for formative and summative assessment that can measure and assess the knowledge of learners with special needs in senior secondary school education.

Education has varying definitions by different educators; In fact, there are probably as many definitions as there are people who write about it. The earliest writers on education; Plato and Aristotle as cited in Onwuka and Duruwoju (2011), had seen education as an instrument for making good citizens. To Plato, education was nothing but the development in the citizens of the virtues which his nature made him capable of; that is the realization of one's potentialities. To Aristotle, "Education is the creation of sound mind in a sound body." Udeozor (2004) asserted that, Education, whether African or western is the sum total of all necessary processes by means of which an individual develops abilities, attitudes and other forms of behaviours of positive value in society. Education is the best instrument for achieving the above mentioned goals. Subsequently, many subjects are studied at different levels of educational institutions in Nigeria toward realizing the goal of education as an instrument for effective national development. Accounting is one of the subjects studied at the senior secondary school level as prescribed by the Federal Republic of Nigeria (FRN, 2014), in her National Policy on Education. Education of students with special needs is designed to facilitate the learning of individuals who, for a wide variety of reasons, require additional support and adaptive pedagogical methods in order to participate and meet learning objectives in an educational programme.

Accounting is a generic term covering both the book-keeping and accounting aspect of an economic entity (Eze, 2014.). According to American Institute of Certified Public Accountants (AICPA) as cited in Agbo (2007), accounting is the art of recording, classifying, summarizing in terms of money, transactions and events which are in part at least of financial character, and interpreting the result thereof. Hence, Accounting plays a very important role in the economic advantage of any Nation, including Nigeria. The senior secondary school level is the building block where students need to be acquainted with the basic Accounting skills, through laying a sound and solid foundation for their formative years in accounting related discipline. This will help and lead them to the acquisitions of knowledge that will help them function effectively as catalyst for development. The students start the formal Accounting Education when they are in Junior Secondary School as Business Studies and after that as Principle of Accounting in Senior Secondary School. The students will study Accounting for three years before sitting for the West African Senior Secondary Certificate Examination (WASSCE) at the Senior Secondary three (SS3). For instance, the transition in which learners undergo from primary level to secondary level and tertiary level require some test of achievement.

An achievement test is one designed to measure a student's knowledge of his proficiency in certain skills (Lovely Professional University, 2012). The type of maximum performance test that describes what a person has learned to do is called achievement test (Thorndike & Thorndike-Christ, 2010). Achievement test according to Ibe (2006) is viewed basically as the competence a person has in an area of content. This competence is the result of many intellectual and nonintellectual variables. Achievement can be referred to as acquisition, learning, or knowledge representation. According to Flateby (2014), Achievement test can be written to ascertain students' level of learning a subject matter course across their entire programme, measuring an individual's achievement of instruction and course objective, assess the group's performance, evaluates the test and promotes learning. Test construction is the various ways that items in psychological measures are created and decided upon. It is also the cultivation of a test, generally with a concise or obvious goal to meet the typical standards of validity, dependability, and other aspects of test standardization. The development of any good test must follow the following steps: Planning the test, item writing, and item analysis.

The test instructor who plunges directly into item writing is likely to produce a lopsided test. To guard against fortuitous imbalances of item coverage, table of specification should be drawn up before any item are written. But how many teachers make out time to plan a test using table of specification? How many teachers cover a large content area during test construction? Secondary Education teachers lack the competence and expertise in test construction and as a result tests constructed by them may not be reliable and valid. Validation according to Gronlund, (1981) refers to the degree of accuracy with which a test measures what it is supposed/intended to measure. However, assessment instrument must be both reliable and valid for the study result to be credible.

DOI: 10.9756/INT-JECSE/V14I3.340

Teachers also resort to constructing tests that cover only the few content areas they have covered in a term (Nworgu, 2015). The problem of this study is that teachers in secondary schools rely so much on teacher made tests. Could the psychometric properties of the teacher made test be established, by establishing the norms and validity and determining the various levels of items analysis of the test? The use of test items in evaluating student outcome in accounting will enhance students' achievement. It will also enhance to an extent, the transfer of knowledge, skills and sentiments acquired in accounting to true life situation. Also, according to FME (1985), Continuous Assessment stipulates that test items should be used for the evaluation of the achievement of the Secondary School Students in Accounting and other subjects. Popham (2008), states that evaluation is the appraisal or determination of the value of an object or characteristics of an individual by justification saying that it is good or bad; better or worse necessary or unnecessary; desirable or un desirable; poor or excellent; successful or unsuccessful, when the objects or the particular character of the individual is placed against a certain criterion. Asuru (2012) considers evaluation as the systematic process of passing valued judgment as to the worth of a thing, object, programme etc. Evaluation is an integral part of teaching and learning process, it is a relatively inclusive concept as applied to education generally and to science, technology and mathematics (Ibe,2006).

Measurement is the harnessing of responses to test items or other stimuli, or the collection and analysis of expert or examinee judgements for the purpose of making inferences and, ultimately, to arrive at decisions based on those inferences (Secolsky & Denison, 2018). According to Onunkwo (2002), Educational Measurement is the Quantitative representation of students' characteristics/traits. Measurement is situating data from an assessment in a quantitative framework, to characterize the evidence the observations provide for the interpretations and inferences the assessment is meant to support (Mislevy, 2018). Also, Evaluation is about making judgment from the measure of student's traits. Such instruments include test, observation, techniques, interviews, sociograms, and inventories etc. In a study carried out by Ezechukwu, et al, (2020) to determine the psychometric properties of the Economics Achievement Test (EAT) using Item Response Theory (IRT). The result of the study shows that the range of item difficulty obtained using one parameter logistic model (1PLM) is -2.706 – 1.571, while that of two parameter logistics model (2PLM) is -4.00 – 1.338. The finding of the study also revealed that there is a significant different between the item difficulty estimated using 1PLM and 2PLM. Obinne (2008) examined the psychometric properties of the items of the Biology examinations conducted by the National Examination Council (NECO), and the West African Examination Council (WAEC) using the Item Response Theory (IRT). It was found that the Biology examination items from the two examination bodies were reliable and valid. Biology items in the NECO-conducted examination for 2001 were more difficult than those of WAEC of the same year. WEAC items were more prone to guessing than those of NECO items. The author's study further reveals that negative difficulty estimates indicate that the items are easy while positive difficulty estimates indicate that the items are hard. The findings of Chong (2013) reveal that difficulty parameter or the threshold parameter value tells how easy or how difficult an item is. The findings which revealed that the items were selected based on the b-value range of -3 to +3 corresponds with (Baker, 2001) that theoretically, difficulty values can range from - 00 to + 00, in practice, difficulty values usually are in the range of - 3 to + 3. Baker also described the range of values for item discrimination as follows: very low, 01 - .34, Low, 35 - .64, moderate, 65 - 1.34 High, 1.35 - 1.69 and Very high, 1.70 and above.

Over the years, students' with special needs poor achievement in accounting in the public secondary schools have been attributed to many factors, such as, lack of valid test construction by teachers, teacher's reliance on teacher-made test which are designed to measure the achievement of students only on the cognitive domain and which are not reliable as well as lacks experts' validation (Ibe, 2006). Also, some secondary school teachers are not competent in effective test construction, because they are not exposed to workshops, seminars and conferences that will help them know how to construct an effective test item. The procedures adopted in the development of the teacher-made test are haphazard and non-elaborate; and their qualitative item analysis are not determined. The achievement of students was very poor in accounting because of bad planning and negligence by the teachers. However, lack of well-constructed, valid and reliable instrument for assessing the achievement of secondary school students in accounting has stimulated the study. The purpose of this study was to determine the psychometric characteristics of Accounting Multiple Choice Questions Achievement Test for the assessment of students with special needs through expert's judgement. The following questions were addressed:

1. To what extent do experts agree that the instrument constructed has content validity?
2. What are the ranges of difficulty indices of the items of AMCQAT?
3. What are the ranges of discrimination indices of the items of AMCQAT?
4. What are the respective distracter indices of the instruments?

DOI: 10.9756/INT-JECSE/V14I3.340

Materials and Methods

Research Design

The researchers adopted instrumentation research design in conducting the study. The reason for using the design is that it aims at determining the psychometric properties of an instrument which involves validation and production of a test for teachers and others to use in assessing students' achievement in accounting.

Study Participants

A total of 200 students with special needs offering accounting in Senior Secondary School (SS3) were randomly sampled from Public Secondary Schools in South-East, Nigeria. The recruitment of the participants was done through the school authorities.

Research Instrument

The instrument used for data collection was the Accounting Multiple choice questions Achievement Test (AMCQAT). The test comprised of 100 multiple choice tests in accounting, the items were drawn based on the three levels of the educational domain. The AMCQAT was developed by the researchers under all the topics in accounting curriculum. The items were drawn from the table of specification based on the educational objectives of bloom taxonomy of education; each of these multiple-choice test items was configured to consist of a question stem followed by five answer options, lettered A-E, the correct answer (key) and distracters. The special need students respond by choosing one option from the letters A-E for each item.

Validity and Reliability of the Instrument

The AMCQAT was validated in both content using the table of specification and face validation, by giving it to three experts, one Accounting Education expert and two measurement and evaluation experts. Based on the suggestions of the experts, the items were reviewed accordingly. The reliability coefficient of the instrument was determined using Kuder-Richardson formula (K-R_{20}). The reliability coefficient of temporal stability over time and that of internal consistency of the instrument were determined with 25 items of AMCQAT on 30 students outside the main study area. Internal consistency reliability of AMCQAT was estimated to be 0.70 using Kuder-Richardson Formula (K-R_{20}) while temporal stability of AMCQAT was determined to be 0.96 using Pearson Product-moment correlation.

Test Administration

The instrument was administered on all the participants (Students with special needs that are offering accounting). The accounting teachers assisted in administering the test to their students; at the end of the test administration, the scripts of the students were collected by their teachers and handed over to the researchers. The students' scripts were marked and scored by the researchers.

Data Analysis

The data collected from the administration of the AMCQAT were analysed using SPSS version 23.00. The analysis involves percentage, mean and standard deviation, item difficulty (U-L/N), item discrimination (D = U-L/ ½ N), and item distracter (L-U/ ½).

Ethical Approval

The ethics committee at the schools where the research was conducted granted ethical approval. The informed consent forms were properly filled and signed by the participants.

Results

Table 1a shows the seven (7) experts Content validity rating of AMCQAT. The result shows that out of the seven experts who rated the content validity of the instrument, four (4) of the experts agreed that the instrument passed content validity while the rest of the three (3) strongly agreed. Table 1b also shows the frequency distribution of seven (7) experts on the content validity rating of the instrument. The result shows that only two (2) experts being about 28.6% of the respondents rated that the psychomotor level of the instrument was weak but in overall reported that the instrument passed the test of validity as the mean reports 3.71. The result reports a grand mean of 4.37 indicating a total agreement by the respondents.

DOI: 10.9756/INT-JECSE/V14I3.340

Table 1a. The Seven (7) Experts Content Validity Rating

S/No.	Cognitive	Affective	Psychomotor	Content Validity	Report
1	4	4	4	4	Agreed
2	5	5	4	5	Strong Agreement
3	4	4	5	4	Agreed
4	5	5	4	5	Strong Agreement
5	5	5	5	5	Strong Agreement
6	5	5	2	4	Agreed
7	5	5	2	4	Agreed

Rating Scale: 5.0 =Strong agreement, 4.0 =Agreement, 3.0 =Neutral, 2.0 =Weak agreement, 1.0 = No agreement

Table 1b. Frequency Distribution of the Experts

S/No.	SA	A	N	WA	NA	Mean	Standard Dev.	Total
Cognitive	5 (71.4)	2 (28.6)	0	0	0	4.71	0.488	7
Affective	5 (71.4)	2 (28.6)	0	0	0	4.71	0.488	7
Psychomotor	2 (28.6)	3 (42.9)	0	2 (28.6)	0	3.71	1.254	7
	Grand Mean					4.377		

Table 2 shows the range of difficulty indices of the items of AMCQAT and its proportion. The result shows a higher proportion in the difficulty level with 0.70-0.89 range for the difficulty index of the individual scores of the items of AMCQAT. The result shows that there is a proportion of 0.72 difficulty index meaning that the majority of the items which ranged from 0.70 – 0.89 is moderately difficult. Only a proportion of 0.19 out of the entire items which ranged from 0.90 – 1.00 was reported very easy while a proportion of 0.08 ranged from 0.00 – 0.49 which is reported very difficult. Only 1% of the item ranging from 0.50 – 0.69 is reported fairly difficult. See figure 1 for a graphical illustration of the result.

Table 2. Range of difficulty indices of AMCQAT

Difficulty Index Level			
Ranges	F	(proportion)	Report
0.00-0.49	8	(0.08)	Very Difficult
0.50-0.69	1	(0.01)	Fairly Difficult
0.70-0.89	72	(0.72)	Moderately Difficult
0.90-1.00	19	(0.19)	Very Easy
Total	100		

DOI: 10.9756/INT-JECSE/V14I3.340

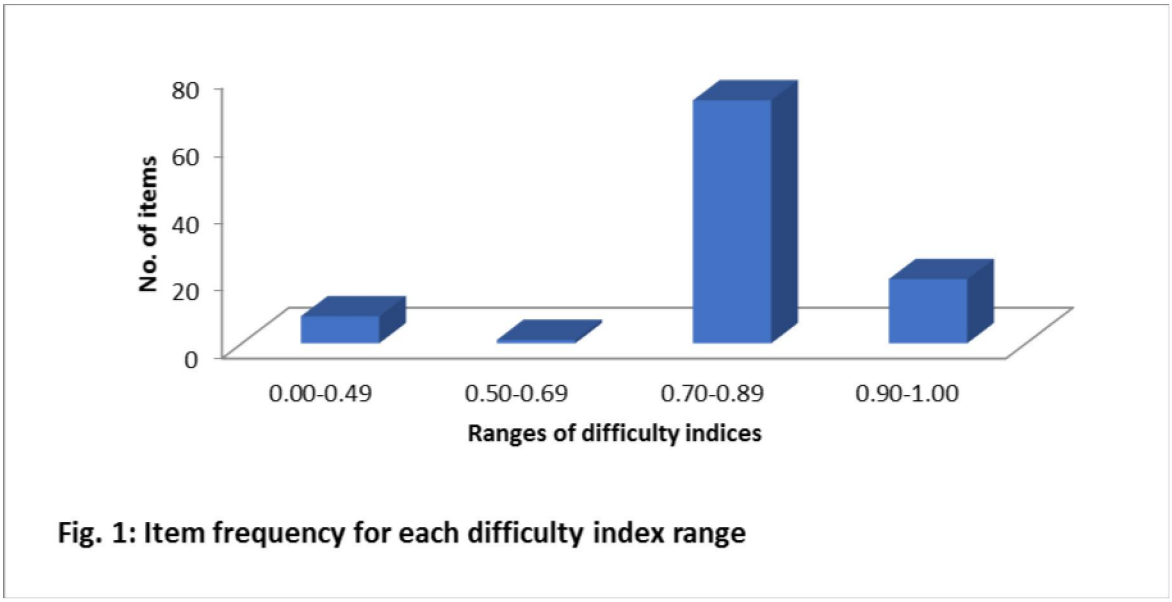

Fig. 1: Item frequency for each difficulty index range

The data on table 3 contains the range of discrimination indices of the items of AMCQAT and its proportion. The result has shown a higher proportion in the discrimination index level with negative value for the items of AMCQAT reporting that the quality of the item is worst of which suggests that they should be definitely discarded. The discrimination indices with the range above 0.40 accounts 34% of the result as the result considers these qualities of the items as excellent reporting that the items should be retained. The discrimination indices with the range 0.30 – 0.39 accounts 12% of the study rating the quality of these items as good, therefore reporting that there is room for improvement for these items. The range 0.20 – 0.29 being about 3% of the study is seen as mediocre in terms of quality of the item and should be reviewed. The range 0.01 - 0.20 being about 12% of the study is seen as poor in terms of quality of the items and should be discarded. The result is further graphically represented in figure 2 below.

Table 3. Range of discrimination indices of AMCQAT

	Discrimination Index Level			
Ranges	F	(%)	Quality of Item	Report
Negative	39	(0.39)	Worst	Definitely discard
0.01 - 0.20	12	(0.12)	Poor	Discard
0.20 - 0.29	3	(0.03)	Mediocre	Need to review
0.30 - 0.39	12	(0.12)	Good	Scope of improvement
Above 0.40	34	(0.34)	Excellent	Retain
Total	100			

DOI: 10.9756/INT-JECSE/V14I3.340

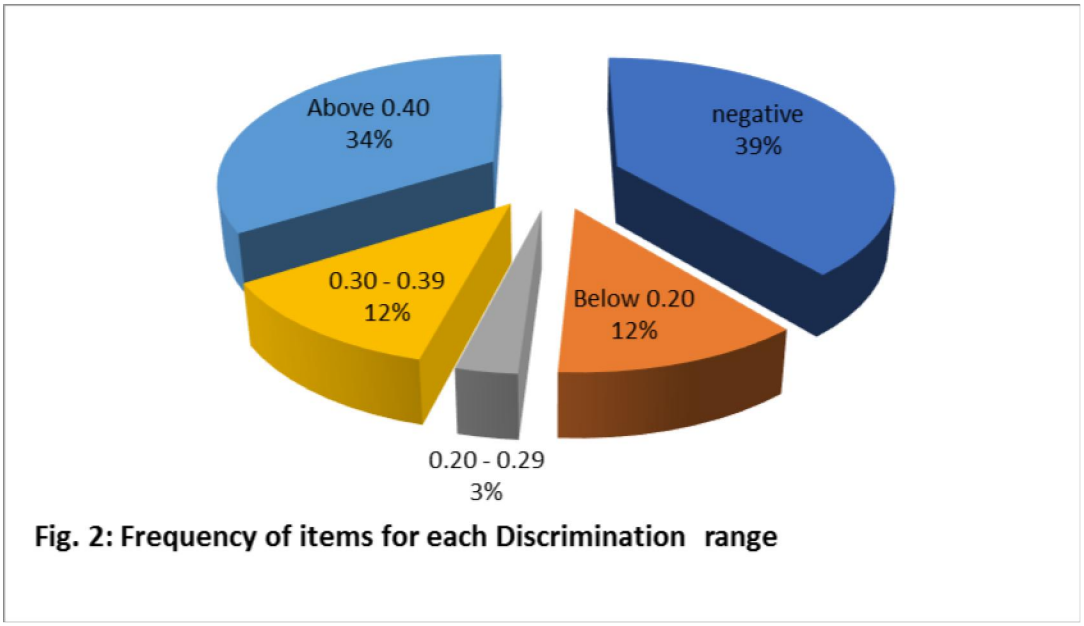

Fig. 2: Frequency of items for each Discrimination range

Table 4 shows the respective distracter indices of all the options in all the items. The result shows a higher proportion in the distracter index level with negative range for AMCQAT. The result accounts about 61% are in negative range, showing a larger proportion while 39% of the distracter indices were reported positive. See figure 3 for a chart representing the result.

Table 4; Distracter indices of all the options in all the items

Distracter Index Level

Ranges	F	(%)	Report
Negative	61	(0.61)	Bad
Positive	39	(0.39)	Good
Total	100		

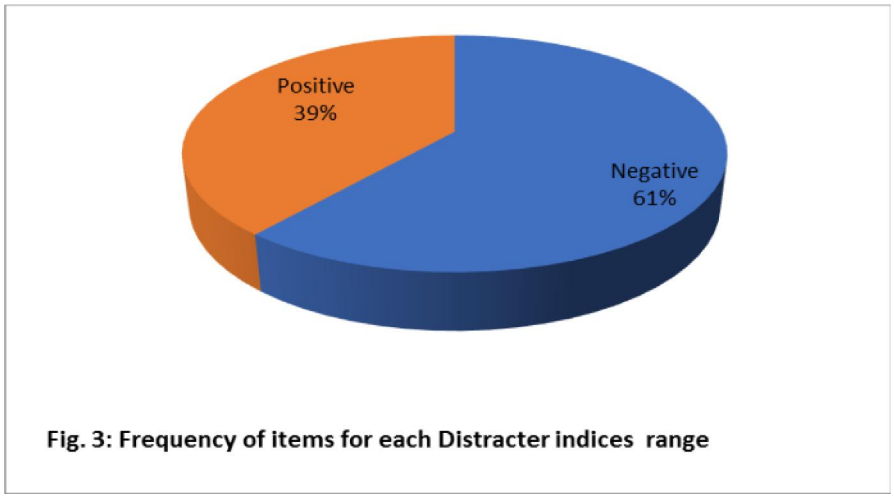

Fig. 3: Frequency of items for each Distracter indices range

DOI: 10.9756/INT-JECSE/V14I3.340

Discussion

The findings revealed that out of the seven experts who rated the content validity of the instrument, four (4) of the experts agreed that the instrument passed content validity while the rest strongly agreed. The result also shows that only two (2) experts being about 28.6% of the respondents rated that the psychomotor level of the instrument was weak but in overall reported that the instrument passed the test of validity. The finding indicated that the distribution of items to the objectives and content areas are enough evidence that the test items has a high content validity. This finding imply that the validity co-efficient of the items on AMCQAT as judged by the experts on its content is significant. The result revealed a high grand mean indicating a total agreement by the respondents. This finding is in line with Kpolovie (2010) who stated that test must be constructed in such a vigorous and rigorous manner that by all theoretical, abstract and empirical indications, it possesses unquestionable validity. The finding agrees with Nworgu (2015) who reported that for a test to be valid questions must be set from all parts of the syllabus and that this emphasizes the need to ensure adequate coverage of both subject matter area and the instructional objectives which the students' learning centred on.

The result further revealed a higher proportion in the difficulty level of the test items and majority of the items are moderately difficult. Only a proportion out of the entire items was reported very easy while eight percent was reported very difficult. One percent of the item was reported fairly difficult. This finding agreed with the findings of Agbo (2007) who found out that some test items are difficult, that is, relatively low proportion of testees answer them correctly, while some test items are easy, that is, a relatively high proportion of the testees answer them correctly. These are within the acceptable range of item difficulty as stated in Ezechukwu, at el., (2020). Item difficulty is relevant for determining whether students have learned the concept being tested.

The result has shown a higher proportion in the discrimination index level with negative value for the items of AMCQAT reporting that the quality of the item is worst of which suggests that they be definitely discarded. The discrimination indices with the range above 0.40 accounts 34% of the result as the result considers these qualities of the items as excellent reporting that the items should be retained. The discrimination indices with the range 0.30 – 0.39 accounts for 12% of the study rating the quality of these items as good, therefore reporting that there is opportunity for improvement for these items. The range 0.20 – 0.29 being about 3% of the study is seen as mediocre in terms of quality of the item and should be reviewed. The range 0.01 - 0.20 being about 12% of the study is seen as poor in terms of quality of the item and should be discarded. There is a significance difference among the range of discrimination indices of the items constructed ($P<0.05$). This finding agreed with the findings of Onunkwo (2002) which provides a guideline for interpreting different values of discrimination indices. This finding was similar to a study conducted by Ogomaka, et al (2017) and El-Uri and Malas (2013) who reported that 38% of the test items had the discrimination coefficient less than 0.2 with 23 questions obtained negative discrimination.

The result shows a higher proportion in the distracter index level with negative range for AMCQAT. The result shows that about 61% are in negative range, showing a larger proportion while 39% of the distracter indices were reported positive. The finding is in line with Onunkwo (2002) who identified that the value of distracter index ranges from -1.00 to +1.00. The authors further stated that a positive value indicates that the distracter is effective (good), since it is chosen by more of the students in the low ability group than those in the high ability group. A negative value of the index indicates that the distracter is not effective (it is poor or bad) since it is chosen by more of the students in the high ability group than those in the low ability group. A zero-distracter index indicates that the distracter does not distract or confuse any students. In other words, it makes no differentiation in the amount of confusion it poses to students in the two ability groups.

Conclusions

The overall report showed that the instrument passed the test of validity. The difficulty level of the items showed that the majority of the items were moderately difficult, very easy, and very difficult while only one item was reported as fairly difficult. The result showed a higher proportion in the discrimination index level with negative values, meaning that the quality of items is worst and should be definitely discarded. The result also showed a higher proportion in the distracter index level with negative while few of the distracter indices were reported as positive. The number of items that constitute the AMCQAT was a limitation to the study. This could pose a problem to the students due to their short attention span and endurance rate. This might contribute to the general poor performance on the AMCQAT exhibited by the students. This study is an improvement on similar instruments that are currently presenting challenges to the users. The instrument developed from this study may be used to collect information about how students are performing and thus contribute to the description of educational phenomenon. To the educators, the instrument could be used for diagnostic, formative and summative evaluation of students' achievement in accounting.

DOI: 10.9756/INT-JECSE/V14I3.340

Recommendations

To address the problem of subjectivity in teacher made test, it wass recommended that:

1. Students with special needs and their teachers should use the developed instrument to get prompt feedback on their competence in Accounting, so that they are able to identify areas where they may need remediation.
2. It should be used for both formative and summative evaluation of students with special needs in senior secondary schools.

Acknowledgements

The researchers are very grateful to our colleague's and all the authors whose works were consulted during the process of this study. This research did not receive any specific grant from any funding agency in the public, commercial, or non-profit making organization.

References

Agbo, O.S (2007). Assessment of science process skill possessed by senior secondary school students. Unpublished B.sc (ED) Project University of Nigeria Nsukka.

Asuru, V.A (2012). *Measurement and evaluation in education and psychology*. Port-Harcourt Minson Publishers.

Black, P. J., & William, D. (2009). Assessment and classroom learning. *Assessment in education*. 5, 7-74. https://doi.org/10.1080/0969595980050102

Baker, F. B. (2001). *The basics of item response theory*. New York, NY: ERIC Clearinghouse.
https://files.eric.ed.gov/fulltext/ED458219.pdf
https://doi.org/10.1007/978-3-319-54205-8

Chong, H. Y. (2013). A Simple guide to the Item Response Theory (IRT) and Rasch modeling. Retrieved from March, 2019 from http:// www.creativewisdom.com.

El-Uri, F.I & Malas, N. (2013). Analysis of use of a single best answer format in an undergraduate medical examination. *Qatar medical journal*. 2013(1), 3-6 doi: 10.5339/qmj.2013.1

Ezechukwu, R.I., Oguguo, B.C.E.; Ene, C.U & Ugorji, C.I. (2020). Psychometric Analysis of Economics Achievement Test Using Item Response Theory. *World Journal of Education*, 10(2) pp. 59-68. doi:10.5430/wje.v10n2p59. URL: https://doi.org/10.5430/wje.v10n2p59

Federal Republic of Nigeria (2014). *National policy on education*. Lagos: Federal Government Press

Flateby, B.A (2014).Outstanding issues in assessment for learning. A Paper Presented At the 3rd Annual Conference of *International Association for Education Assessment*, Abuja.

Grounlund, N.E (1981). *Measurement and evaluation in teaching* (4th Edition). New York: Macmillan Publishers Inc.

Ibe, A. E. (2006). *Psychology of education, child-hood, adolescence and adult-hood in teaching- learning transaction*. Lagos, Kola Okanlawon Publishers.

Lovely Professional University (2012). *Educational measurement and evaluation*. New Delhi: USI Publications

Jonah-Eleli, P (2000). Assessment misunderstood grail educational leadership. *Educational Leadership*, *66*(1), 82 - 83. http://www.ascd.org/publications/educational-leadership/sept08/vol66/num01/A-Misunderstood-Grail.aspx

Kpolovie, P. J. (2010). *Advanced Research Methods*. Owerri: Springfield Publishers Ltd.

Mislevy, R.J. (2018). On measurement in educational assessment. In C. Secolsky & D.B Denison (Eds), *Handbook on Measurement, Assessment, and Evaluation in Higher Education* (pp. 11). New York, NY: Routledge.

Nworgu, B. G (2015). *Educational measurement and evaluation*: Theory and Practice (2nd Edition).

Obinne, A. D. E. (2008). Psychometric properties of senior certificate biology examinations conducted by West African Examinations council: Application of item response theory. *Unpublished doctoral dissertation*, University of Nigeria, Nsukka.

Ogomaka, P.M.C; Ukozor, F.I; & Oguguo, B.C.E (2017). Development of financial accounting achievement test for assessment of senior secondary school students in Imo State. *Nigerian Journal of Educational Research and Evaluation (ASSEREN)*, *16*(1), 25 – 33.

Onwuka, A.O.U. & Duruwoju, E.O. (2011), Continuous assessment for improved higher education learning achievements in business management. A Paper Presented at Herpnet In August 8-12, 2011.

Onunkwo, G (2002). *Fundamentals of education measurement and evaluation*. Owerri Cape Publishers International Ltd.

DOI: 10.9756/INT-JECSE/V14I3.340

Secolsky, C. & Denison, D.B. (2018). Improving Institutional Decision-making through Educational Measurement, Assessment, and Evaluation. In C. Secolsky & D.B Denison (Eds), *Handbook on Measurement, Assessment, and Evaluation in Higher Education* (pp. 11). New York, NY: Routledge.

Thorndike, M.R. & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and Education* (8th ed). New Delhi: Asoke K. Ghosh.

DOI: 10.9756/INT-JECSE/V14I3.340