DEVELOPMENT AND VALIDATION OF MATHEMATICS ACHIEVEMENT TEST FOR SENIOR SECONDARY SCHOOLS IN KATSINA ZONAL EDUCATIONAL QUALITY ASSURANCE

ATUMA OCHENU LAWRENCE[1]; DR. ADEKUNLE THOMAS OLUTOLA[2] & PROF. RAFIU ADEMOLA OLATOYE[3]

[1]Department of Science Education, Faculty of Education,
Federal University Dutsin-Ma, Katsina State.
[2]Department of Educational Psychology and Counselling, Faculty of Education,
Federal University Dutsin-Ma, Katsina State
[3]Department of Educational Foundations, Faculty of Education,
Federal University Lokoja, Kogi State.
generalatuma@gmail.com; aolutola@fudutsinma.edu.ng; olatoyedemola@yahoo.com

Abstract

The study examined the development of mathematics achievement test (MAT) for senior secondary schools in Katsina State, Nigeria. Three research questions were asked and tested at 0.05 level of significance. The research design adopted for the study was instrumentation research design. The population of the study comprised seventeen thousand five hundred and seventy two (17572) students in public secondary schools in Katsina zonal educational quality assurance. Multi stage sampling technique was used to select three hundred and seventy eight (378) students in public secondary schools in the State. The instrument for data collection for this study was mathematics achievement test (MAT) developed by the researcher. The instrument was validated by two experts and the reliability of (MAT) was carried out using Kuder Richardson Formula-20 with reliability co-efficient of 0.75. Data obtained were subjected to statistical analysis by calculating difficulty, discrimination and distracters indices using partial credit model. The findings of the study showed that the developed multiple choice test items had the infit and outfit statistics ranging from 0.57-1.41 and 0.81-3.04 respectively. It is only items 13, 17 and 29 that have outfit of 2.09, 2.22 and 3.04 respectively beyond the accepted range of 0.7-1.5. Furthermore, the findings revealed that the Mathematics achievement test (MAT) was estimated to be 0.753. This shows that the Mathematics achievement test (MAT) developed is highly reliable. Tthe findings also indicated that items for the Mathematics achievement test (MAT) developed had difficulty estimate that ranges from -1.28 of item (1) (the easiest item) to 1.52 of items 20 and 44 (the most difficult item). Based on the above findings, it was recommended among others that psychometricians, test developers, teachers and other persons involved in any kind of measurement should be trained in item response theory framework.

Keywords: Development, Validation, Mathematics, Achievement Test

Introduction

Mathematics is the foundation of science and technology without which a nation can never become prosperous and economically independent. Arsevens (2015) and Mefor (2014) stated that the world today is extremely mathematical; hence modern citizens cannot afford to be ignorant of it. Arsevens also mentioned that in the 21st century, teachers are expected to teach mathematics with the aim of having students create effective solutions. In a similar vein, Acharya (2017) posited that mathematics is one of the most important subjects in our modern lives. Without the knowledge of mathematics, it may be

difficult to progress in today's global society.

Mathematics is a unique discipline that promotes and facilitates critical thinking among its learners. Mathematics has been at the core of ancient civilizations, dating as far back as the ancient Greeks and Roman empires. Today, mathematics is equally value and continues to form the foundation for many other disciplines. Ernest (2015) believed that mathematics is useful in developing intellect and consistency of thoughts and ideas. In referencing these skills. Mathematics has been characterized as a tedious subject to learn as well as to teach. Sa'ad, Adamu, and Sadiq (2014) found that poor performance in mathematics can be attributed to students being fearful of the subject. Gafoor and Kurukkan (2015) agreed that many students are of the view that the subject is boring and expressed that they find it difficult to remember and understand formulae.

However, Mathematics is usually taught and assessed through tests. A test is an assessment intended to measure a test-taker's knowledge, skill, aptitude, physical fitness, or classification in many other topics (Fajobi, 2019). A test may be administered verbally, on paper, on a computer, or in a predetermined area that requires a test taker to demonstrate or perform a set of skills. Similarly, tests are used to determine whether students have learned what they were expected to learn or the degree to which students have learned the material. Tests make it possible for both teachers and other stakeholders in education to discover how class, school or authority compares in the national scene.

Testing has been seen worldwide as a best way to determine the success or failure of the teaching and learning process in the school setting (Olutola, Ogunjimi, Daramola & Sheu, 2017). A test is used to find out the entry behaviour of the pupils to whom the subject would be taught. That is it is used to measure how much the pupils have mastered on the course already. The objectives for administering tests are: to compare students' performance, to determine change in behavior, for feedback purposes, for placement purposes, for promotion purposes, to determine teacher teaching efficiency and to promote healthy rivalry among students. Ajeigbe and Afolabi (2023) in their view stressed that a typical multiple-choice objective test consists of a widely sample of subject syllabus or course outline within a well-defined ranged of cognitive levels.

However, before a test is developed, a table of specification must be constructed to guide the developer. According to Obidiegwu, (2018) a table of specification is a plan or guide for test preparation.

Test validity and reliability is a vital factor to consider in testing and test construction. A content valid test measures what it is projected to measure — the content as depicted by the test blueprint (Fajobi, 2018). The tests in the field of education are generally constructed to measure some specific qualities or abilities (variables) of the students. If a test measures the traits or abilities for which it was constructed, then such a test is called valid, and these traits of the test is called validity. According Olatoye and Olutola, (2020) validity refers to the degree to which attest measures what it is supposd to measure. Also Bandele (2014) defines validity as the extent to which an instrument measures what it purports to measure. Another psychometric property of a test instrument is its reliability. According to Drost (2011), reliability is the extent to which test items used for measurements are repeatable when different examinees perform equally on different occasion, under different condition, supposedly with alternative instruments which measure the construct or skill. Reliability is the measure of the consistency, stability, dependability, predictability, precision and accuracy (Olatoye & Olutola, 2020).

There are three common types of item analysis which provide teachers with three different types of information:

Difficulty Index - Teachers produce a difficulty index for a test item by calculating the proportion of students in class who got an item correct (Fajbi, 2018). is, not the difficulty of the item.) The larger the proportion, the more students who have learned the content measured by the item.

Discrimination Index - The discrimination index is a basic measure of the validity of an item. It is a measure of an item's ability to discriminate between those who scored high on the total test and those who scored low. Though there are several steps in its calculation, once computed, this index can be interpreted

as an indication of the extent to which overall knowledge of the content area or mastery of the skills is related to the response on an item. Perhaps the most crucial validity standard for a test item is that whether a student got an item correct or not is due to their level of knowledge or ability and not due to something else such as chance or test bias (Fajbi, 2018).

Analysis of Response Options - In addition to examining the performance of an entire test item, teachers are often interested in examining the performance of individual distractors (incorrect answer options) on multiple-choice items. By calculating the proportion of students who chose each answer option, teachers can identify which distractors are "working" and appear attractive to students who do not know the correct answer, and which distractors are simply taking up space and not being chosen by many students. To eliminate blind guessing which results in a correct answer purely by chance (which hurts the validity of a test item), teachers want as many plausible distractors as is feasible (Fajobi, 2018). Analyses of response options allow teachers to fine tune and improve items they may wish to use again with future classes.

Statement of the Problem

In the secondary school system, the curriculum demands that different assessments must be carried out in the course of the instruction to guide effective teaching, learning and to evaluate the level of mastery of the learning objectives by the students. To achieve this, many teachers rely on the teacher made test which often lack the required psychometric properties of a good test.

Unfortunately, most teachers do not know how to construct a valid and reliable tests. According to (Osadebe, 2014) these test usually lack psychometric properties such as validity, reliability, difficulty, discrimination and functionality. Also, researchers like Chime, (2012); Ugwu, (2012) and Obilor, (2019) are of the opinion that the teacher-made tests lack basic psychometric properties and as a result they are not very appropriate for the assessment of students. It has been observed that most teachers, always pick up published past questions and some simply write down their own test items without considering the psychometric properties of the test items.

It is believed that teacher-made tests have some flaws which affect the performance of the students. Therefore, in a bid to overcome these problems and to improve the performance of students in Mathematics, there is need to develop and validate mathematics achievement test that will have all the required psychometric properties of a standardized test. Thus, this research seeks to construct and validate an achievement test on Mathematics for Senior Secondary Schools in Katsina Zonal Educational Quality Assurance.

Objectives of the Study

The objective of this study is to develop Mathematics Achievement Test (MAT) for Senior secondary schools (SSI) in Katsina Zonal Educational Quality Assurance. Specifically, the study sought to;

i. determine the validity of Mathematics achievement test (MAT) in Katsinaa Zonal Educational Quality Assurance
ii. find out the reliability of Mathematics achievement test (MAT) in Katsina Zonal Educational Quality Assurance.
iii. determine the difficulty index of Mathematics achievement test (MAT) in Katsina Zonal Educational Quality Assurance.

Research Questions

The following research questions were asked in the study:

i. What is the validity of Mathematics achievement test (MAT) in Katsina Zonal Educational Quality Assurance?
ii. What is the reliability of Mathematics achievement test (MAT) in Katsina Zonal Educational Quality Assurance?
iii. What is the Difficulty index of Mathematics achievement test (MAT) in Katsina Zonal Educational

Quality Assurance?

Methodology

The study adopted an instrumentation research design. The population for this study consists of all SSI students in public secondary schools in Katsina educational quality assurance zone, Katsina State. Katsina State is one of the states in the North Western part of Nigeria. The state has a largely agrarian economy and most communities engage in growing crops and rearing of animals. Multistage sampling techniques was adopted for the study. The researcher clustered the schools into zones. Out of the notable three zones in Katsina quality assurance, two schools were randomly selected per each zone. The schools selected were Kaita zone; government secondary school Girka, Sabitu Mohammad Yahaya secondary school Jifatu, Jibia zone; government senior secondary school Danddara, government senior secondary school Gangara, Katisina zone; Family support senior secondary school Katsina and Government college Pilot Katsina Also, the selected schools were clustered into classes and SSI was randomly selected. Thus, a total of three hundred and seventy eighty students were sampled for the study.

Mathematics Achievement Test (MAT) was developed by the researcher for data collection. MAT is consists of sixty (60) multiple choice test items with four (4) alternative options lettered A – D from which students choose the correct answer, each question carries one (1) mark given a total of sixty (60) marks. To determine the number of text items to be generated from each topic, a test blue print or table of specifications was constructed and used based on Bloom's taxonomy of Educational Objectives of learning in the cognitive domain (Knowledge, comprehension, Application, Analysis, Synthesis and Evaluation).

The data collected for this study were analyzed using using Kuder-Richardson (KR-20) and partial credit model.

Results

Research Question One: What is the validity of Mathematics achievement test (MAT) in Katsina Zonal Educational Quality Assurance?

Table 1: Validity of Multiple Choice test items of Mathematics achievement test (MAT) in Katsina Zonal Educational Quality Assurance based on IRT

Items	Infit 2023	Outfit 2023	Items	Infit 2023	Outfit 2023
1	1.11	1.13	**31**	0.82	0.89
2	1.04	0.98	**32**	1.13	0.99
3	0.82	0.81	**33**	1.24	2.28
4	0.91	0.86	**34**	0.57	0.78
5	1.32	1.51	**35**	1.03	1.09
6	1.01	1.06	**36**	1.06	0.95
7	1.13	1.14	**37**	0.81	0.85
8	1.11	1.07	**38**	0.81	0.93
9	1.22	1.11	**39**	1.28	1.53
10	0.93	1.11	**40**	1.01	1.04
11	1.11	1.13	**41**	1.11	1.14
12	0.88	0.84	**42**	1.04	1.14
13	1.53	2.09	**43**	1.12	1. 26
14	0.83	0.84	**44**	0.94	1.12
15	0.89	0.96	**45**	1.12	1.17
16	0.92	0.96	**46**	0.82	0.87
17	1.51	2.22	**47**	0.73	0.93
18	0.72	0.85	**48**	0.85	0. 91
19	1.14	1.08	**49**	1.09	1. 12

20	0.84	0.93	**50**	0.96	1.12
21	0.94	0.98	**51**	1.14	1.17
22	0.72	0.86	**52**	0.87	0.87
23	0.86	0.94	**53**	1.08	1.17
24	0.91	0.86	**54**	0.81	0.83
25	1.11	1.13	**55**	0.89	0.97
26	0.92	1.15	**56**	0.99	1.15
27	1.09	1.18	**57**	0.78	0.85
28	0.82	0.82	**58**	1.03	1.10
29	1.41	3.04	**59**	0.95	1.04
30	0.82	0.89	**60**	0.83	0.87
Mean	**0.99**	**1.08**			
SD	**0.49**	**0.52**			

In Table 1 the results showed that the developed multiple choice test items had the infit and outfit statistics ranging from 0.57-1.41 and 0.81-3.04 respectively. It is only items 13, 17 and 29 that have outfit of 2.09, 2.22 and 3.04 respectively beyond the accepted range of 0.7-1.5. The fit statistic of the developed items apart from items 13, 17 and 29 indicate that all the items are perfect and valid since they all fall within the range of fit regarded valid. Since the mean of the infit and outfit are 0.99 and 1.08 with generally low S.D of 0.49 and 0.52; it implies that 97% of the items are valid and unidimensional. It could therefore be said that apart from items 13, 17 and 29 in the developed multiple choice test items, all the items are within the accepted range and therefore valid and sufficiently demonstrates unidimensionality.

Research Question Two: What is the reliability of Mathematics achievement test (MAT) in Katsina Zonal Educational Quality Assurance?

Table 2: Reliability of Mathematics achievement test (MAT) in Katsina Zonal Educational Quality Assurance

Method of Establishing Reliability	Kuder-Richardson formular 20 of MAT Based on Standardized Items	N of Items
Kuder-Richardson Formular 20 (KR-20)	0.753	60

The reliability index of the Mathematics achievement test (MAT) was estimated to be 0.753 using Kuder-Richardson formular 20 (KR-20).This shows that the Mathematics achievement test (MAT) developed is highly reliable.

Research Question Three: What is the Difficulty index of Mathematics achievement test (MAT) in Katsina Zonal Educational Quality Assurance?

Table 3: Difficulty index of Mathematics achievement test (MAT) in Katsina Zonal Educational Quality Assurance based on Model of IRT

Items	Difficulty Estimates	Items	Difficulty Estimates	Items	Difficulty Estimates
1	-1.28	**21**	0.47	**41**	-0.22
2	-1.06	**22**	0.14	**42**	0.61
3	-0.91	**23**	0.28	**43**	-0.07
4	1.15	**24**	0.43	**44**	1.52
5	0.90	**25**	1.38	**45**	0.45
6	0.01	**26**	-1.18	**46**	0.18
7	-0.39	**27**	0.94	**47**	0.28
8	-0.23	**28**	1.16	**48**	0.42

9	-0.98	**29**	0.91	**49**	1.26
10	-0.89	**30**	0.03	**50**	1.06
11	-0.55	**31**	-0.37	**51**	-0.84
12	-0.45	**32**	-0.28	**52**	1.12
13	-0.99	**33**	-0.96	**53**	0.94
14	-0.39	**34**	-0.87	**54**	0.04
15	-0.58	**35**	-0.52	**55**	-0.49
16	1.19	**36**	-0.45	**56**	-0.34
17	0.27	**37**	-0.99	**57**	-0.94
18	-0.73	**38**	-0.36	**58**	-0.99
19	-0.08	**39**	-0.55	**59**	-0.53
20	1.52	**40**	1.19	**60**	0.41
	Mean 0.68	**SD**	**0.41**		

The results show that items for the Mathematics achievement test (MAT) developed had difficulty estimate that ranges from -1.28 of item (1) (the easiest item) to 1.52 of items 20 and 44 (the most difficult item). Within this range, thirty two items had negative difficulty estimate which means thirty two fairly easy items. Also, twenty eight items had positive difficulty estimate which implies fairly difficult questions.

Data also showed that 45 items which constitutes 75% of total test items with difficulty estimates below 0.5 as easy items; and also, 15 items constituting 25% of the total items with difficulty estimates above 0.51 were presented as difficult items. Based on this information, it can be deduced that 75% of the items were made too easy to even the lower ability examinees to have performed well. The mean of the estimate distribution is 0.68 which suggest that fairly easy items balance fairly difficult items. The difficulty indices are desirable since both the positive and negative range, are close to 0.68 and the standard deviation is low 0.41.

Discussion of Findings

The study showed that the developed multiple choice test items had the infit and outfit statistics ranging from 0.57-1.41 and 0.81-3.04 respectively. It is only items 13, 17 and 29 that have outfit of 2.09, 2.22 and 3.04 respectively beyond the accepted range of 0.7-1.5. The fit statistic of the developed items apart from items 13, 17 and 29 indicate that all the items are perfect and valid since they all fall within the range of fit regarded valid. Since the mean of the infit and outfit are 0.99 and 1.08 with generally low S.D of 0.49 and 0.52; it implies that 97% of the items are valid and unidimensional. It could therefore be said that apart from items 13, 17 and 29 in the developed multiple choice test items, all the items are within the accepted range and therefore valid and sufficiently demonstrates unidimensionality. The finding is in agreement with the finding of Bryce (2018) who revealed an item will be valid or of good fit to the model if it had fit statistics of 1.5 and below. A large positive fit indicate a poor fit, but fit statistic nearer one (1) indicate a better fit. The findings are also in line with the finding of Boman (2017), Lian and Idris (2006), noted that infit and outfit statistics should be between 0.7 and 1.5 for items validity or fitness to be acceptable for moderately rigorous assessment purposes and for such items to be considered unidimensional.

The study showed that the Mathematics achievement test (MAT) was estimated to be 0.753. This shows that the Mathematics achievement test (MAT) developed is highly reliable. The finding is in agreement with the finding of Aliyu (2016) who showed that items and person separations indices were 13.17 and 2.93 while item and person reliability were 0.99 and 0.78 respectively. The MNSQ for both infit and outfit were 0.94 and 1.08 respectively while the ZSTD for both in fit and outfit are -1.7 and +2.0 respectively which were within the acceptable range of 0.7-1.1 for MNSQ for sample > 1000 while -2.0 to +2.0 for ZSTD.

The study revealed that items for the Mathematics achievement test (MAT) developed had difficulty estimate that ranges from -1.28 of item (1) (the easiest item) to 1.52 of items 20 and 44 (the most difficult item). Within this range, thirty two items had negative difficulty estimate which means thirty two fairly easy items. Also, twenty eight items had positive difficulty estimate which implies fairly difficult questions.

Data also showed that 45 items which constitutes 75% of total test items with difficulty estimates below 0.5 as easy items; and also, 15 items constituting 25% of the total items with difficulty estimates above 0.51 were presented as difficult items. Based on this information, it can be deduced that 75% of the items were made too easy to even the lower ability examinees to have performed well. The mean of the estimate distribution is 0.68 which suggest that fairly easy items balance fairly difficult items. The difficulty indices are desirable since both the positive and negative range, are close to 0.68 and the standard deviation is low 0.41. The findings are similar to the finding of Nkpone (2001) who revealed that in item response theory, validity connotes a fit to the model; that item discrimination are uniform and substantial and that there is no error in scoring.

Conclusion

It was concluded that thirty two items had negative difficulty estimate which means thirty two fairly easy items. Also, twenty eight items had positive difficulty estimate which implies fairly difficult questions. Data also showed that 45 items which constitutes 75% of total test items with difficulty estimates below 0.5 as easy items; and also, 15 items constituting 25% of the total items with difficulty estimates above 0.51 were presented as difficult items. Based on this information, it can be deduced that 75% of the items were made too easy to even the lower ability examinees to have performed well.

Recommendations

Based on the findings of this study, the following recommendations were made;

1. Psychometricians, test developers, teachers and other persons involved in any kind of measurement should be trained in item response theory framework. This will enable the advantage of the framework and its overall essence to be appreciated and popularized in our local situation.
2. The government, ministries of education and high profile stakeholders in education should procure the various IRT analytical software and sponsor the training of individuals to learn the analysis using IRT framework. This way, the framework would have been popularized and the interpretation of the results and usage of the framework will thus be demystified.
3. Given the obvious advantages of IRT over other popular measurement framework, the government should encourage our examination bodies such as WAEC, NECO, NABTEB, NTI etc to adopt this measurement framework. This will ultimately surmount the measurement problems we frequently encounter in Nigeria. Such measurement problem as test score equating has nearly gone into extinction in the foreign countries that have adopted the IRT measurement framework. IRT framework can also do the magic for us in Nigeria.

References

Acharya, B. R. (2017). Factors affecting difficulties in learning mathematics by mathematics learners. *International Journal of Elementary Education, 6*(2), 8–15.

Ajeigbe, T. O. & Afolabi, E. R. I (2023). A study of Item Response and Anchor Bias in Economics Objective Tests among Senior Secondary School Students in Osun State. Academic Journal of Interdisciplinary Studies, Vol. 2 No 3 p. 173-179.

Aliyu, R.T & Akinoso, S.O (2016). Development and validation of mathematics aptitude test (MAT) using

Rasch and 2-Pi Model of IRT. IJES, 16(2), 1-1

Arsevens, A. (2015). Mathematical modelling approach in mathematics education. *Universal Journal of Educational Research, 3*(12), 973–980.doi:10.13189/ujer.2015.031204

Asuru, V.A. (2015). Measurement and Evaluation in Education and Psychology (2nd Ed). Port Harcourt: Pearl, Publishers International Ltd

Chime, U. M. (2022).Development and validation of economics achievement test for senior secondary school students. Unpublished M.Ed Thesis, Faculty of Education, University of Nigeria Nsukka

Drost, E., A. (2021). Validity and reliability in social science research. *Education Research and Perspectives, 38*(1), 105-124.

Ernest, A. (2015). *Designing gamification for the recreational shopper* (Master's thesis). Chalmers University of Technology, Gothenburg, Sweden.

Fajobi, O.O (2019). The Development and Validation of Mathematics Multiple Choice Test Items for Junior Secondary School Three (JSS 3) Students in Oyo State, Nigeria. *Global Scientific Journals (GSJ)*, 7(11), 23-43.

Gafoor, K. A., & Kurukkan, A. (2015). *Why high school students feel mathematics difficult? An exploration of affective beliefs*. Paper presented at the UGC Sponsored National Seminar on Pedagogy of Teacher Education, Trends and Challenges, Kozhikode, Kerala, India.

Mahajan, G., 2015. Construction and validation of achievement test in economics. International Journal of Humanities and Social Sciences Studies, 1(6): 54-60. View at Google Scholar

Mefor, C. (2014). Nigeria: Identifying problems of poor performance in mathematics and the way out. Retrieved from http://allafrica.com/stories/201101200591.html

Obidiegwu, U. J. (2018).Development and Validation of Physical Education Achievement test (PEAT) for adult learners in Anambra state.Selected work.Retrieved April, 22, 2013, from http://works.bepress.com druche_Obidiegwu/6.

Obilor, I. E. (2019). Essentials of Measurement and Evaluation. Port Harcourt: Sabcos Printers & Publishers.

Olutola, A. T., Ogunjimi, M. O., Daramola, S.O. and Sheu, A.L. (2017). Effect of Convergent Learning and Divergent Learning Strategies on Students' Academic Performance in Biology in Ilorin, Nigeria. *African Journal of Interdisciplinary Studies,* 10, 33-39.

Olutola, A.T. & Olatoye, R.A. (2020).Introduction to Educational Measurement and Evaluation. Department of Educational Foundations, Faculty of Education, Federal University, Dutsin-Ma, Katsina State, Nigeria.

Osadebe, P.U (2014).Construction of Economics Achievement Test for Assessment of Students.*World Journal of Education*, 4(2), 58-64

Sa'ad, T. U., Adamu, A., & Sadiq, A. M. (2014). The causes of poor performance in mathematics among public senior secondary school student in Azare Metropolis of Bauchi State, Nigeria. *IOSR Journal of Research & Method in Education, 4*(6), 32–40.

Ugwu, S. N. (2020).Development and validation of criterion referenced achievement test in Biology.Unpublished M.Ed Thesis, Faculty of Education, University of Nigeria, Nsukka.