BSU Journal of Science, Mathematics and Computer Education (BSU-JSMCE) Volume 5, Issue 1, June 2025

DEVELOPMENT AND VALIDATION OF UPPER BASIC II TWO-TIER MATHEMATICS DIAGNOSTIC TEST USING RASCH MODEL IN KOGI STATE

[1]Olorunnishola Olufikayo Agnes, [2]John Iorhemen Kyeleve, [3]Jerry Ebere Omenka and [4]Arivi Sediq Saliu

[1&4] Department of Science Education, Prince Abubakar Audu University Anyigba
and
[2&3] Department of Mathematics and Science Education, Rev. Fr. Moses Oshio Adasu University, Makurdi
olufikayoolorunnishola@gmail.com, jkyeleve@bsum.nigeria.ng.edu

Abstract

This research focused on development and validation of Mathematics two-tier diagnostic test using the Rasch model. Instrumentation research design was employed. The scales used to measure achievement test in Nigeria rely on classical test theory (CTT) approach. As a result, they have limitations like poor precision, sample dependency and undue focus on aggregate scores that deny test developers the opportunity of determining how the examinees performed on a test item. Multistage sampling procedure was used to sample 266 students from the population of 2150 students. The instrument was a self-developed 20 multiple choice questions. The instrument was face validated by three experts from Mathematics Education and two from measurement and Evaluation while the content and parametric properties were ascertained using the table of specification and Rasch model. Internal consistency reliability value of the instrument using KR20 was 0.85 while the inter-rater reliability was 0.74. The result showed that items difficulty ranged from -0.02 to 1.00 while the difficulty level indices were 1.2 to 2.5. The infit and outfit ranged from 0.6 to 1.2. Findings also showed that gender difference in upper basic II students' conceptual understanding was not statistically significant. It was therefore concluded that two-tier mathematics diagnostics test (TTMDT) developed is reliable and valid to be used for both summative and formative assessments in schools and by examination bodies. It was recommended among others that seminars should be organized for mathematics teachers on how to structure and use two-Tier diagnostic mathematics test in the classroom.

Key words: Development, validation, diagnostic test, Mathematics, Rasch model

Introduction

The role of Mathematics towards realizing the nation's scientific technological aspiration is indisputable. Thus, mathematics plays a significant role in human kind's ability to survive on Earth. Its significance transcends mere academic pursits as it functions as a crucial tool for understanding

and solving real-world problems across various disciplines (Yadav, 2020). This likely explains why arithmetic is taught as a required subject in elementary schools everywhere, including Nigeria. As such, no nation can thrive well without mathematics. This implies that Mathematics has much to offer in solving the problems of mankind by being the gateway to scientific and technological development of any nation. Furthermore, no academic performances of students can be measured without assessment. Assessment in simple terms is the "what and how" a student progress is measured academically. Oduro (2015) explains that assessment in the classroom is viewed as being essential to achieving high quality education. Tefera (2014) also assert that a succession of important choices about the content, the instruction and the assessment of a mathematics session lead s to an effective lesson. This means that assessment is very important if we must know the progress in the mathematics knowledge of a student. Some examples of assessment tools are multiple-choice questions, short-answer questions, essay questions, standardized, quizzes, two-tier diagnostic test among others.

A two-tier diagnostic test is a multilevel assessment model that helps in knowing the weakness and mistakes of students in understanding a concept. Two-tier diagnostic tests combine multiple test and short answer test which consist of two sections at least. The first section aims to diagnose how the individual interprets discipline knowledge. In the second section, students are asked to state the reason (s) for their answer (s) in the first section. The use of two-tier test is aimed at reducing the guessing factor of students because students are required to provide reasons to the answers they have chosen. (Lengkong, Istiyono., Rampean, Rejeki, Tumanggor & Nirmala, 2020). Consequently, the need to develop the necessary mathematical, scientific, and technological knowhow of the next generation demands assessment using valid and reliable diagnostic test. That is, not just answering questions in mathematics but knowing the underlying reasoning for the answers given. Some researchers like; (Vee Diagrams (Novak, 1198), Bloom's Taxonomy (1956) have also tried to develop instruments to cater for the proper assessment of the conceptual understanding of students but some of the instruments are challenged haven been noticed to be over simplifying the learning process, the package been confusing for students to use and so on. However, Ermie, (2017) gave general guidelines for ExamSoft statistical interpretation of a developed two-tier diagnostics test which are: item difficulty – acceptable item difficulty is how many exams takers answered the item correctly, discrimination index – this ranges from -1.0 to +1.0; the closer to 0.0 indicates no discrimination among high and low -performing students. Achieving closer to 1.0 discrimination index is optimal and Point Biserial – the point -biserial correlation is the Pearson correlation between responses to a particular item and scores on the total test. It also ranges from -1.0 to 1.0. Aliyu (2015) developed and validated Mathematics Achievement Test using the Rasch model. The result 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 infit 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 difficulty level of the items ranges between -1.95logit to 7.45logit. this shows that the instrument is valid. Also, Mutlu and

Sesen (2015) investigated the development of a two-tier multiple-choice diagnostic test to assess undergraduates' understanding of some chemistry concepts. The findings indicated that difficult indices ranged from 0.30 to 0.77 with an average of 0.50 and the discrimination indices ranged from 0.20 to 0.64 with an average of 0.39. This result indicates that the two-tier diagnostic test was valid and reliable in assessing undergraduates' misconceptions and in identifying their conceptual understanding of the chemistry concepts under investigation. Furthermore, Idenhen and Omoifo (2015) examined the development and validation of a two-tier multiple-choice diagnostic instrument for assessing senior secondary school students' conceptions of selected mathematical concepts in Edo state, Nigeria. And finally, according to kusumaningsih, Saputra and Aini (2019) in their study cognitive style and gender difference in conceptual understanding of mathematics students in the United kingdom, male and female students have similar understanding of mathematical concepts in some aspects.

Furthermore, worthy of note is the fact that several studies (Fadilah, Mahdzir. Mazlina, Khuzzan, Sunarti, Mastan, Ahmad & Yahya, 2023; Onah, 2019; Ode, Odoh, & Amana, 2016) have concentrated on external factors influencing the learner that can improve learners' performance, such as teaching methods and strategies, teacher variables, instructional materials, and classroom environments without considering learners' internal factors such as intelligence, lack of understanding among others in the learning process. It is against this background that the researcher seeks to develop a reliable and validate instrument to assess upper basic II students conceptual understanding in mathematics knowledge using two-tier diagnostics test in Kogi State.

Purpose of the Study

The purpose of this study was to develop and validate Mathematics two-tier diagnostic test using the Rasch model. Specifically, the study:

(1) estimated the item parameters of Two-tier Mathematics Diagnostic Test (TTMDT) for upper basic II using the Rasch model
(2) ascertained the conceptual understanding of Male and Female upper basic 2 students' Mathematics knowledge assessed using TTMDT?

Research Questions

The following research questions guided the study

1. What is the estimate of item parameter of Two-tier Mathematics Diagnostic Test (TTMDT) for upper basic II using the Rasch model?
2. What is the mean conceptual understanding of Male and Female upper basic 2 students' Mathematics knowledge assessed using TTMDT?

Hypothesis

1. There is no significance difference in the mean conceptual understanding of Male and Female upper basic II students' Mathematics knowledge assessed using TTMDT?

Methodology

The study adopted instrumentation design. It is a design which typically involves creating tools to assess and analyze data in various fields. Two-Tier Mathematics Diagnostic Test (TTMDT) was developed by the

researchers in three stages adopting procedures by Treagust in (widiyatmoko & shimizu, 2018). The instrument TTMDT cycles I and II for the study were subjected to face, content and construct validation by five experts. The final version of the instrument TTMDT was trial tested on 10 students outside the sampled schools to establish the reliability of the instrument. Kuder-Richardson Formula 20 was used to measure the internal consistency reliability of the test which is 0.85. The study adopted partial credit model in scoring. In this model, student received a score of 2 marks if he/she responded correctly to the first tier (content choice) and correctly to the second tier (reasoning part), signifying full conceptual understanding of mathematical concepts. Also correct answer to the content part and wrong reason attracted a score of 1 mark, signifying partial understanding (misconception) while wrong answer and correct reason or wrong answer and wrong reason earned 0 marks, signifying no conceptual understanding or lack of knowledge. This is because if students' ability is assumed to progress from knowing to explanation, students whose answer to first tier is incorrect but correct explanation should not exist on the progression path and this pattern can be attributed to guessing. The data were analyzed using Rasch model which relies on Item Response Theory (IRT) to answer the research question one. Research question two was answered using mean and standard deviation while t-test was used to test the hypothesis at 0.05 level of significance

Results

Research Question One

What is the estimates of item parameter of Two-tier Mathematics Diagnostic Test (TTMDT) Two-tier Mathematics Diagnostic Test (TTMDT) for upper basic II using the Rasch model?

Table 1: *Shows Estimates of Item Parameter of Two-tier Mathematics Diagnostic Test (TTMDT) Two-tier Mathematics Diagnostic Test (TTMDT) for Upper Basic II*

	Proportion	Measure	Standard Error Measure	Infit	Outfit
q1	0.461	0.17301	0.127	0.999	1.002
q1b	0.331	0.76324	0.134	0.961	0.949
q2	0.677	-0.79365	0.135	1.065	1.098
q2b	0.446	0.23787	0.128	1.043	1.048
q3	0.494	0.02789	0.127	1.059	1.064
q3b	0.361	0.62204	0.132	1.074	1.091
q4	0.632	-0.58097	0.131	1.002	1.026
q4b	0.539	-0.16560	0.127	1.001	1.011
q5	0.465	0.15684	0.127	1.091	1.104
q5b	0.316	0.83604	0.136	1.013	1.018
q6	0.520	-0.08482	0.127	1.015	1.024

	Proportion	Measure	Standard Error Measure	Infit	Outfit
q6b	0.372	0.57036	0.131	0.955	0.931
q7	0.483	0.07618	0.127	1.063	1.066
q7b	0.238	1.25333	0.147	1.012	0.991
q8	0.710	-0.96272	0.139	1.059	1.141
q8b	0.353	0.65685	0.132	1.025	1.024
q9	0.361	0.62204	0.132	1.094	1.103
q9b	0.130	2.02485	0.185	1.019	1.059
q10	0.323	0.79943	0.135	1.009	1.037
q10b	0.190	1.55775	0.159	0.969	0.943
q11	0.468	0.14068	0.127	1.011	1.012
q11b	0.230	1.29722	0.149	1.005	1.050
q12	0.416	0.36886	0.128	1.003	1.008
q12b	0.257	1.14733	0.144	0.964	0.926
q13	0.394	0.46869	0.130	0.971	0.973
q13b	0.257	1.14733	0.144	0.924	0.880
q14	0.487	0.06008	0.127	0.919	0.908
q14b	0.268	1.08602	0.142	0.917	0.871
q15	0.468	0.14068	0.127	0.908	0.900
q15b	0.249	1.18913	0.145	0.937	0.898
q16	0.658	-0.70352	0.133	1.000	1.014
q16b	0.491	0.04398	0.127	1.076	1.081
q17	0.502	-0.00430	0.127	1.016	1.020
q17b	0.442	0.25414	0.128	1.038	1.042
q18	0.561	-0.26315	0.128	0.994	0.988
q18b	0.390	0.48549	0.130	0.934	0.917
q19	0.506	-0.02040	0.127	0.963	0.957
q19b	0.331	0.76324	0.134	0.945	0.911
q20	0.509	-0.03649	0.127	0.957	0.958
q20b	0.335	0.74529	0.134	0.963	0.934

Note. Infit= Information-weighted mean square statistic; Outfit= Outlier-sensitive means square statistic.

Data in Table 1 shows the Rasch item statistics where the first column identifies the items (q1, q2, q3…. q20) and reasons to each corresponding item (q1b, q2b, q3b……qb20). Column 2 is the proportion, indicating the percentage of respondents that got the items (option and reason) correctly and this range from -0.11 to 2.63. The third column is the measure of item difficulty level, followed by the Standard Error (SE). The last two columns are the infit and outfit measures. From the table, item 8 is the most correctly scored item with 71% while item 10 was least correctly scored with 32.3% for the part a. The analysis showed that respondents find it difficult to provide correct reasons despite that the first items were correct. The range for correct reasons is from 0.13 to 0.24. Item 9 was the most difficult item for respondents to provide reason for correct option with only 13% correctly and the highest percentage (35.3%) of respondents got the correct reason on item 8.The third column (measure), further confirmed the easiest items (q8, q2, q4) with (-0.96, -0.80, -0.58) and most difficult items (q9, q10, q7) with (0.62,.0.80, 0.08). The acceptable threshold of Infit and Outfit (stable performance) ranges from 0.6 and 1.2, the spread of the values under Infit and Outfit are all within this range: infit (1.0 – 1.1) and outfit (1.0 -1.1) which shows that all the items are pretty good. Fig1a shows the wright map which is also known as item map and shows the spread of item difficulty and person's ability. The item difficulty ranges from -0.02 to 1.00 for the options while the difficulty level of reasons for the correct options are 1.2 to 2.5. This is estimates of item parameter of Two-tier Mathematics Diagnostic Test (TTMDT) for upper basic II. The model fit which indicate the reliability is given as 0.735, which is approximately 0.74 and shows that the items are good and very reliable. However, the reliability coefficient from Kuder Richardson formular (KR20) is 0.85.

Research Question Two

What is the mean conceptual understanding of Male and Female upper basic II students' Mathematics knowledge assessed using TTMDT?

Table 2: Mean and Standard Deviation of Upper Basic II Male and Female Students' Conceptual Understanding of Mathematics Knowledge Assessed Using TTMDT

Gender	**N**	**Mean ($\bar{x}$)**	**SD**
Male	133	17.44	5.83
Female	136	16.90	11.59
Mean Difference		0.54	

The summary result of the analysis in Table 2 revealed that the mean scores of male Upper Basic 2 students' conceptual understanding of mathematics knowledge is 17.44 with SD 5. 83, while that of female Upper Basic 2 students is 16.90 with SD of 11.59. The mean score difference between the male and female students' conceptual understanding of mathematics knowledge assessed using two-tier diagnostic test is 0.54 in favour of the male students. This suggests that male Upper Basic students may have more understanding of the mathematical knowledge under study.

Hypothesis

There is no significance difference in the mean conceptual understanding of Male and Female upper basic II students' Mathematics knowledge assessed using TTMDT.

Table 3: *t-test Result of Male and Female Upper Basic II Students' Conceptual Understanding Scores of Mathematics Knowledge as Assessed Using TTMDT*

Gender	N	$(\bar{x})$	SD	DF	t	Sig.	Decision at P < 0.05
Male	133	17.44	5.83				
				267	-0.493	0.622	Not Sig.
Female	136	16.88	11.63				

The summary result of the analysis on Table 3 shows that the significant value (p) was greater than the set value for the study (p > 0.05) therefore, the hypothesis that there is no significant difference in the mean conceptual understanding of male and female upper basic 2 students in mathematics knowledge assessed using two-tier diagnostic test, was not rejected. This implies that there was no significant difference in the mean conceptual understanding of upper basic education 2 students as assessed using two-tier diagnostic test. The conclusion drawn was that there is no gender bias in students' understanding of the mathematics concepts as assessed using two tier diagnostic test.

Discussion of findings

Data in table 1 Column 2 is the proportion, indicating the percentage of respondents that got the items (option and reason) correctly and this range from -0.11 to 2.63 while the third column is the measure of item difficulty level while the range for correct reasons is from 0.13 to 0.24. The acceptable threshold of infit and outfit (stable performance) ranges from 0.6 and 1.2, the spread of the values under infit and outfit are all within this range, which shows that all the items are very good. Fig1 also shows the spread of item difficulty and person's ability. The item difficulty ranges from -0.02 to 1.00 for the options while the difficulty level of reasons for the correct options are 1.2 to 2.5. This shows that the instrument is reliable and valid. This finding corroborates that of Mutlu and Sesen (2015), Idenhen and Omoifo (2015), Aliyu (2015) and Ermie (2017) who found that when an instrument is having difficulty level ranging from – 1.0 to +1 is considered valid because the range of validity is similar. Furthermore, the reliability gotten from Rasch model package is given as 0.74 which shows that the items are good and very reliable. However, the reliability coefficient from Kuder Richardson formular (KR-20) was also given as 0.83 which is deemed reliable. This finding is also the consistent with the finding of Aliyu, (2015) who also

found reliability for diagnostic test as 0.75 which is within the range of very good reliability. In a nut shell two different reliability was generated but within the plausible range.

The result of the analysis in Table 2 showed that male and female students had a mean conceptual understanding score gain difference of 0.54. However, these figures are considered small and confirmed in Table 3 that the difference is not statistically significant with regard to mathematics conceptual understanding. This implies that male and female students hold equal conceptual understanding level as measured using two-tier diagnostic test. This finding can be related to the finding of kusumaningsih, Saputra, and Aini (2019) who in their study found that male and female students exhibit similar spatial ability in mathematics and hence performed equally. This similarity in male and female students' understanding of mathematics knowledge may be as a result of their ability to mentally represent and manipulate objects in space and these skills predict better understanding.

Conclusion

Based on the findings of this study, the researcher concluded that two-tier diagnostic test indicated that this test is valid and reliable to be used in assessing students' conceptual understanding of mathematics knowledge among upper basic II students in Kogi State. Furthermore, the use of Two-Tier Mathematics Diagnostic Test (TTMDT) at any giving time of instruction helps a competent teacher in the classroom to assess students understanding. Also, the conceptual understanding of upper basic 2 students using two-tier diagnostic test is not gender biased.

Recommendation

1. School administrators should encourage teachers in developing two-tier Mathematics diagnostics test so as to help them in their assessment of students' conceptual understanding their various classes.
2. Application of the Rasch model principles of test development and validation for performance test in Nigerian schools and Examination bodies is highly recommended.
3. Seminars should be organised for mathematics teachers on how to structure and use two-Tier diagnostic Mathematics test in the classroom

References

Aliyu, R.T. (2015) Development and validation of mathematics achievement test using the rasch model. An un published Thesis Submitted to the Postgraduate School in Partial Fulfillment of the Requirements for the Award of Doctor of Philosophy (Ph.D) Degree in Measurement and Evaluation of the Delta State University, Abraka.

Ermie, E. (2017) psychometric 101: know what your assessment data is telling you. Slide presented at ExamSoft Assessment Conference, Denver

Fadilah, N., Mahdzir, M., Mazlina,S., Khuzzan, M.S.,Sunarti,N., Mastan, Z. P., Ahmad, N., N. & Yahya, F.A.N (2023). Factors influencing learners' engagement in the classroom. *Asian Journal of Research in Education and Social Sciences, 2*(5), 53-62.

Idehen, F.O., & Omoifo, C.N. (2015). Development and validation of a two-tier multiple-choice diagnostic instrument for assessing senior secondary school students' conceptions of selected mathematics concepts. *International Journal of*

Research and Development (IJRD), *2* (1), 254-272.

Kusumaningsih, W., Saputra,H.A. & Aini, N (2019). *Journal of physics Conference series* 1280 (0): 042017

Lengkong, M., Istiyono, E., Rampean, B. A. O., Rejeki, A. M., Tumanggor., & Nirmala, M. F. T (2020). Development of two-tier test instruments to detect student's physics misconception. Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences Advances in Social Science, Education and Humanities Research: 528.

Mutlu, A., & Sesen, B.A. (2015). Development of a two-tier diagnostic test to assess undergraduates' understanding of some chemistry concepts. *Procedia-social and Behavioral science,* 174, 629-635. Available online @ www.sciencedirect.com

Ode, J.O., Odoh, C.O., & Amana, D. (2016). Influence of teachers' attitude on students' interest in basic science in Ankpa Local Government Area of Kogi State, *Benue Journal of Mathematics and Mathematics Education*, *1*(5), 42-48.

Oduro, E.O. (2015). Assessment in mathematics classrooms in Ghana: A study of Teachers' practices. An unpublished Doctorate Thesis, the University of Sussex.

Onah, J. (2019), Measures of students' conceptual understanding and reasoning of Junior Secondary School in Geology using two-tier diagnostic test in Benue State Nigeria. An unpublished doctorate thesis, submitted to the Postgraduate School, Benue State University, Makurdi, Partial Fulfillment of the Requirements for the Award of Doctor of Philosophy (Ph.D) in Mathematics Education.

Tefera, G. (2014). Teachers perceptions and practices of continuous assessment in mathematics class ': A masters thesis, Department of science and mathematics Education, Addis Ababa University.

Yadav, S. (2020). Role of mathematics in the development of society. *International journal of research and Analytical Reviews, 6*(4), 295-297.