Journal Website

Article history:
Received 17 March 2024
Revised 01 June 2024
Accepted 18 June 2023
Published online 30 June 2024

International Journal of Education and Cognitive Sciences

Volume 5, Issue 3, pp 1-8

E-ISSN: 3041-8828

Innovating Assessment Through the use of Tailored Testing on Student Achievement of Senior Secondary Mathematics Students in Kogi State

Onuh Omale [1*]

[1] University of Agriculture Makurdi, Benue State Nigeria

*** Corresponding author email address**: omaleonuh@gmail.com

A r t i c l e I n f o

Article type:
Original Research

How to cite this article:
Omale, O. (2024). Innovating Assessment Through the use of Tailored Testing on Student Achievement of Senior Secondary Mathematics Students in Kogi State. *International Journal of Education and Cognitive Sciences, 5*(3), 1-8.
https://doi.org/10.61838/kman.ijeas.5.3.1

A B S T R A C T

Purpose: The research aimed to assess the effectiveness of tailored testing on the academic performance of senior secondary three (SS3) Mathematics students in Kogi State.

Methodology: The study employed a quasi-comparative experimental research design, involving a total of 492 students from the state. The sample specifically consisted of SS3 Mathematics students, and a combination of disproportionate and purposive sampling techniques was utilized for participant selection. Data collection involved the use of three instruments, including a 50-item National Examination Mathematics past Questions paper (NECOMPQ). The study was guided by two research questions and two hypotheses. Descriptive statistics, such as mean and standard deviation, were employed to address the research questions, while the t-test statistic was utilized to assess the hypotheses.

Findings: The findings indicated a notable difference in the mean achievement scores between high and low-ability students when using standardized test items. In contrast, there was no significant difference in mean achievement scores between high and low-ability students when tailored testing was employed.

Conclusion: It can be concluded that when appropriate assessment tools are employed, there is a higher likelihood of revealing a student's true academic ability. Consequently, the study offers the following recommendations: Emphasize the importance of using tailored testing methods in educational assessment practices to better accommodate students with varying abilities. Encourage educators and institutions to consider the individualized needs and capabilities of students when designing assessment instruments. These recommendations are grounded in the research's findings and underscore the potential benefits of tailored testing in enhancing the accuracy of academic assessments.

Keywords: *Innovation, Tailored testing, Achievement, Secondary school, Mathematics*.

1. Introduction

The knowledge of capability of a person on a job or the ability of a student to cope with a specified task at any given level of study is of great importance to any institution that is progress driven. Therefore, there is need for the right instrument to be used in measuring the ability of persons before they are assigned a task or promoted to the next level. When a wrong instrument is used in measuring the ability of someone, it is obvious that the result will be invalid and unreliable. There are several instruments that can be used in measuring one's achievement and this includes among others, tailored test, aptitude test, intelligence test parallel test. Test could mean different things to different people depending on the area of interest. For instance, in Medicine, test means medical examination conducted in order to check or discover somebody's health condition. It could also be seen as an activity or instrument used or intended to find out whether somebody or something has the required qualities (Agbir, 2021).

But Onuka (2013) sees test as a systematic procedure for observing a person's behaviour and describing it with the aid of numerical scale or category (Onuka, 2013). In a different view, Odinko (2014) define test as an intended to show that something works or works well. A test therefore, is a measurement instrument or device administered to someone to determine the relative value of the trait or skills which include cognitive, affective and psychomotor skills (Odinko, 2014). For instance, a spelling test measures how well someone spells or the extent to which someone has learned how to spell a specific list of words (cognitive skills). A Mathematics interest test measures the extent to which someone has interest/likeness towards Mathematics (affective skills), while a test on filling the burette with acid measures someone's dexterity in filling the burette with acid (psychomotor skills). When information is elicited about the testees using test items that tally with their individual abilities, tailored testing is inferred. Tailored Testing is a process that systematically matches the difficulty of the test items with the abilities of the examinee being tested (Willis & Hyde, 2015). It is also a process in which different sets of test questions (items) are administered to different individuals depending on each individual's status on the trait being measured. Tailored testing is a system designed to administer tests or, more specifically, individual items, tailored to an individual's ability. The system utilizes the computational and storage capabilities of the computer to identify, select, and administer to each individual the items best able to measure the ability of the individual. Items that tend to be too easy or too difficult for an individual are avoided. In tailored testing, one or more items are administered to an examiner, and scored correct or incorrect. Based on the response of the examinee, additional items are selected from an item bank with items of known difficulties and discrimination. Therefore, a routine testing is done to establish the individual abilities of the testees.

Tailored testing has also been variously referred to as adaptive, programmed, response contingent, computerized, automated, individualized, branch and sequential testing. Tailored testing can also result in efficient and more accurate mastery classifications, and provides an efficient and practical approach to the measurement of individual change. Tailored testing solves the problem by selecting from an item bank for each individual a test designed of items which are appropriate in difficulty level for each examinee. This is in line with the idea of Item Response Theory (IRT). The idea behind IRT is that, tailoring of a testing to the ability of the testee should be advocated (Asiret & Sünbül, 2016).

Item Response Theory deals with the characteristics of an item with respect to item difficulty, item discrimination as well as item response pattern of examinee. It is interested in determining what a particular examinee might do when confronted with test item. Such information is necessary if the test designer desires to predict test scores characteristics in one or more population of examinee. An appealing feature of item response theory is that, with its application, once an examinee ability level has been established, it is possible to determine the probability of a correct response to an item the examinee has never taken assuming that certain item parameters have already been determined.

As highlighted by Orluwene (2012), test plays very many crucial roles not just in educational development but also in the facets nation (Orluwene, 2012). Hansen (1969), using a computer-based branched testing model, found a significant improvement in internal consistency reliability for computer presentation (r = .80) when compared to a conventional classroom achievement (r = .43). Hansen further reported a significant relationship (r = .76) between computer-based test results and college aptitude scores (Orluwene, 2012).

Mathematics plays a foundational role in the study of science subjects like, Physics, Chemistry, Biology and Computer Science. Mathematics stands out as the "queen of Sciences". The role this subject plays in Science and Technology is enormous and far-reaching. In recognition of this, the Federal Government of Nigeria in 1979 made

E-ISSN: 3041-8828

Mathematics a compulsory subject to be offered in both primary and secondary school levels (FRN, 2014). Mathematics is a pre-requisite for admission into technologically based and other science-oriented courses in higher institutions of learning (Kiss & Kónya, 2020; Nelson et al., 2023; Peters, 1998; Raghubar et al., 2010). In spite of the crucial role Mathematics plays in everyday life, it has remained one of the subjects with the lowest performance in Nigerian schools. Studies between 2016–2020, showed that students' performance has not yet significantly improved. The percentage of students that passed Mathematics at credit level from 2016-2020 still fell between 30% and 47% . Many research findings in Nigeria have shown that there are always differences in performance between examinees (Diahyleva et al., 2021; Yang et al., 2022). However, some authors and researchers have identified various factors that affect students' performances in Mathematics especially at the secondary school level. Prominent among these factors are the nature of the test items and the learners' characteristics. However, despite solutions being proffered by researchers, the low level of performance of students at the SSCE Mathematics persists. Chief examiner report of the three examinations in 2016 indicated that NECO had high enrollment with 69.74% pass in Mathematics while WAEC had 26% pass in Mathematics in 2016 and NABTEB with the least enrollment had 55.93 % pass in Mathematics which is higher than that of WAEC. Also in 2017 WAEC recorded an increase in the percentage of pass in Mathematics with 38.30% while NECO still had 70.85% in Mathematics and NABTEB had 83.31% indicating high performance (Diahyleva et al., 2021; Yang et al., 2022).

Several studied have been carried out on tailored testing. Previous results indicated that the students who used the proposed assessment system outperformed the students who used the other two systems in terms of learning performance and engagement in practice tests and reading materials (Yang et al., 2022). The present study provides insights for researchers who wish to develop formative assessment systems that can adaptively generate practice tests. The results of an experiment proved positive impact of computerized adaptive testing on communicative competence of future ship engineers. Further investigation of adaptive testing can also be done for learning system of maritime education establishments using simulation technologies of virtual, augmented and mixed realities (Diahyleva et al., 2021).

While Ozyurt et al. (2012) in their study described development and architecture of adaptive assessment module integrated into UZWEBMAT. Thanks to incorporation of adaptive assessment modules into e-learning environments, it becomes possible to measure each student with questions appropriate for his/his own level. Therefore, results of this study may contribute to development and implementation of these kinds of adaptive assessment modules in e-learning environments aimed at teaching at different levels. By this means, individualized assessment will become more effective through integration of these modules into distance education software. Adaptive assessment is not advantageous only for students. It also provides certain benefits for teachers (Ozyurt et al., 2012).

The results of this study show that the use of CATs in language testing may have its own advantages and disadvantages. Therefore, authorities should be aware of the positive and negative aspects of a computer adaptive test when they want to administer a test in adaptive format (Rezaie & Golshan, 2015).

It should be noted that examination is pivotal to promoting accomplishment of objective of the curricular in the education industry and this helps in selection, placement, promotion and certificate of students in a particular program according to Onuka (2013) to do these, a right instrument that matches the student's ability should be used. It is in the light of this that the study investigated the efficiency of tailored testing on achievements of SS3 Mathematics students in Ebonyi State (Onuka, 2013).

Two research questions and hypotheses guided the study.

1. What is the difference in mean achievement scores of high and low ability students using parallel test items?
2. What is the mean achievement scores of high and low ability students using tailored testing.

2. Methods and Materials

2.1. Study Design and Participants

The study adopted a pre-post-test randomised parallel control design with three groups. Two groups were the experimental group (Group A and B control design with three groups. Two groups were the experimental group (Group A and B), while one group (Group C) was the control group. The total number of respondents that were used in this study is 1456 senior secondary school students offering physics and are distributed among 21 schools in Calabar Metropolis, Cross River State, Nigeria. The selection involved a rigorous process in that certain criteria were used for the selection of the schools and students that were used

E-ISSN: 3041-8828

in the study. First, the school must have existed for 10 years. Secondly, the school must have qualified physics teachers with a minimum of 5 years' experience. Thirdly, the school must have written a national examination at least five times. Finally, the school must have a standard laboratory for experiments. These criteria reduced the study school population to eleven (11) with a total of 465 science students. A simple random method was used in determining the schools that would be selected for the study. This was done through the balloting method, and 5 schools, which represent 50% of the total number of schools, were selected from a total of 245 respondents (132 males for 54% and 113 females for 56%). The researcher sets the basis for grouping using the pre-test score, which contains items from a test that were randomly arranged. The researchers ensured that all the test papers were arranged to be useful in the assignment process. To obtain a random assignment of subjects in equivalent groups, aromatic progression with a common difference method was used. Here, those who have serial numbers of 1, 4, 7, 10, were in Group A (Experimental Group A), those with serial numbers 3, 6, 9, and 12 were in Group B (Experimental Group Bas the control group. The total number of respondents that were used in this study is 1456 senior secondary school students offering physics and are distributed among 21 schools in Calabar Metropolis, Cross River State, Nigeria. The selection involved a rigorous process in that certain criteria were used for the selection of the schools and students that were used in the study. First, the school must have existed for 10 years. Secondly, the school must have qualified physics teachers with a minimum of 5 years' experience. Thirdly, the school must have written a national examination at least five times. Finally, the school must have a standard laboratory for experiments. These criteria reduced the study school population to eleven (11) with a total of 465 science students. A simple random method was used in determining the schools that would be selected for the study. This was done through the balloting method, and 5 schools, which represent 50% of the total number of schools, were selected from a total of 245 respondents (132 males for 54% and 113 females for 56%). The researcher sets the basis for grouping using the pre-test score, which contains items from a test that were randomly arranged. The researchers ensured that all the test papers were arranged to be useful in the assignment process. To obtain a random assignment of subjects in equivalent groups, aromatic progression with a common difference method was used. Here, those who have serial numbers of 1, 4, 7, 10, were in Group A (Experimental Group A), those with serial numbers 3, 6, 9, and 12 were in Group B (Experimental Group B), and those with serial numbers 2, 5, 8, and 11 were in Experimental Group C (Control Group). Thus, applying the arithmetic progression formula for determining the nth term, (245 = a+(n-1)d), given the first terms as 1, 2, and 3, and the common difference as 3, a total of 81 respondents were assigned to EG1, 82 respondents were assigned to EG2, and 82 respondents were assigned to EG3.

2.2. Measures

Two instruments were used for data collection. These were the Test Anxiety Inventory (TAI) and the Physics Achievement Test (PAT). The TAI was developed by Spielberger in 1980 and has verifiable psychometric properties. The instrument has a total of twenty (20) items and is divided into three groups: Test Anxiety Inventory Total (TAI-T), Test Anxiety Worry (TAI-W), and Test Anxiety Emotional (TAI-E). The test was developed on a four-point modified Likert scale of Likert response options. These options range from nerve to almost away, with a strong pattern of never (1-) and almost never (4-). The degree of consistency of the scale as originally established was 0.96 for TAI-T, 0.91 for TAI-W, and 0.91 for TAI-E. These coefficients, as established by Spielberger (1980), were not established in Nigeria; thus, the researchers had to verify these psychometric attributes in the Nigerian context. A reliability estimate was carried out using 50 students who were not part of the study. The coefficients of the subscales (.87 for TAI-T, 0.83 for TAI-W, and 0.88 for TAI-E), even though not as high as originally established, showed that the instrument can function well in this area and can be used for the study.

The Physics Achievement Test (PAT) was developed and validated by Ibout (2022). The PAT was made up of 50 items with four options. Response of A-D Face, content, and construct validity were already ascertained by the researcher (see Ibout, 2022). The content validity was carried out using a Table of Specifications with five content areas such as properties of waves (15%), radioactivity (10%), energy and society (20%), electric fields (25%), basic electronics (20%), and electromagnetism (10%), with their respective percentages or weights in brackets. The items index was accessed, and the coefficients for p-values ranged from 0.25 to 0.70. Items whose p-values were below or above these values were seen as too easy or too difficult to be considered for the test. With the knowledge of the p-values, the researchers then arranged the items into three formats.

E-ISSN: 3041-8828

Format A was easy-to-difficult (EAD) items; format B was items arranged from difficult-to-easy (DEA) items; and format C was items that were randomly arranged (RAN).

The reliability of each of the formats was ascertained using the equivalent form reliability method. The coefficient values for each format are: format A (easy-to-difficult = 0.77, difficult-to-easy = 0.89, and random format = 0.79). The instrument for this study was adopted because of its sound psychometric properties, and the respondents that were used in the study were recently used in another study.

2.3. Data Analysis

The data collection was carried out in phases. First, the researchers administered the pre-test test of the PAT to all the respondents together with the TAI instrument in the three groups. The RAN instrument was used as the control measure. The data coaction was done after the researcher have collected ethical clearance in line with best practices of research from the institute of research and quality assurance, University of Calabar, Calabar (See code No: UIE/QAD/02/0678). This researcher appropriately obtained the informed consent of the respondents and explained the rationale of the study to the students. The students were also informed that the data that they are to provide will be used to publish a paper whose recommendation may be valuable for their future academics As a form of compensation, each of the respondents was given N300 to reinforce their participation in the study. Those who did not consent to the study were dropped. Out of the 245 students that were selected for the study, 2 were dropped, and these were those in the random and difficult-to-teach format. More so, the students were promised that the data would be protected from public invasion, and this was done by the lead researcher by protecting the data with a password that only the analyst has access to. After the pre-test was administered, the researchers allowed the school teachers to teach under supervision according to the weekly schedules that were drawn for the exercise. The researchers waited for four weeks before administering the post-test. The use of four weeks was to partially eliminate test-wiseness that may constitute a treat if administered within a very short interval. Data collection was done with the initial serial numbers that were given to each script to facilitate easy coding. Each respondent's data was entered accordingly, and the data collected were analysed using descriptive statistics (means and standard deviation) and an analysis of covariance (ANCOVA) using SPSS version 20.

3. Findings and Results

Research question 1: what is the difference in mean achievement scores of high and low ability students using parallel test items?

Table 1

Difference in Mean Achievement Scores of High and Low Ability Students

Group	N	Mean ($\bar{X}$)	S D
High ability	133	66.41	6.57
Low ability	133	25.58	7.29
Mean difference		40.83	

The result shows that high ability group have a $\bar{X}$ of 66.41 and S.D of 6.57 while low ability group have a $\bar{X}$ of 25.25 and S.D 7.29. The $\bar{X}$ difference of the two groups is 40.83.

Research question 2: what is the mean achievement scores of high and low ability students using tailored testing.

Table 2

Difference in Mean Achievement Scores of High and Low Ability Students using Tailored Testing

Group	N	Mean Achievement Scores ($\bar{X}$)	Standard Deviation
High ability	133	72.87	8.94
Low ability	133	68.59	8.74
Mean difference		4.28	

E-ISSN: 3041-8828

Table 2 indicates that the difference in mean achievement scores of both ability groups is 4.28

Hypothesis 1: there is no significant difference in the Mean achievement scores of high ability group and low ability student

Table 3

t-test result of mean achievement scores of high and low ability students

Group	N	Mean	Std.dev	Df	t	Sig	Alpha	Remarks
High ability	133	66.41	6.57	264	47.98	0.000	0.05	Rejected
Low ability	133	25.58	7.29					

Table 3 reveals that P < .05 at t= 47.98, df = 264 and therefore the test statistic is significant therefore the null hypothesis that state that there is no significant difference in the Mean achievement scores of high ability group and low ability student is rejected.

4. Discussion and Conclusion

The research outcomes indicate substantial disparities in mean difference scores, with a difference of 40.83 observed between ability groups when employing a parallel test and a much smaller difference of 4.28 when using a tailored test. This underscores the inherent diversity among individuals, emphasizing the importance of assessing individuals with instruments aligned with their unique abilities. Such an approach also contributes to a reduction in the prevalence of examination malpractice.

These results align with a study conducted by Yanget al. in 2022, which demonstrated that students utilizing the proposed assessment system exhibited superior performance and greater engagement compared to those using alternative systems for practice tests and reading materials (Yang et al., 2022). These findings also resonate with prior research (Ozyurt et al., 2012), which suggested that individualized assessment becomes more effective through the integration of such modules into distance education software.

Notably, adaptive assessment benefits not only students but also teachers. The study's results indicated that mean achievement levels did not significantly differ between high and low-ability students when a tailored test was utilized. This is attributed to the careful selection of items tailored to measure individual abilities effectively, thereby improving measurement precision by aligning item difficulty with individual ability levels.

The implications of these findings are evident: when an assessment instrument that does not align with an individual's ability is used, the resulting information is inherently flawed. Therefore, it is advisable to consider individual variation during item construction.

These findings shed light on the prevailing situation in Nigerian secondary schools, where conventional testing methods are employed to gauge students' abilities. However, the use of appropriate assessment tools holds the potential to unveil students' true capabilities, possibly enabling the prediction of their performance. This aligns with the concept of Item Response Theory (IRT), where test performance can be predicted by defining examinee characteristics (traits), estimating scores based on these traits, and using the scores to forecast or elucidate test performance. Consequently, tailored testing emerges as an efficient means of measuring students' achievements.

Hence, it can be inferred that when students are assessed using questions aligned with their unique abilities, their full potential in cognitive, emotional, and physical aspects can be unveiled. This not only benefits the field of education but also has broader societal advantages. Consequently, it is suggested that teachers receive training in crafting questions tailored to individual capabilities. Additionally, workshops and seminars on education should be conducted for various stakeholders, emphasizing the significance of tailored testing for educational progress in the nation. Given that high-quality education fosters meaningful development, it is recommended that examination bodies like NECO, WAEC, NABTEC, etc., incorporate tailored testing into the assessment of secondary school students.

Authors' Contributions

All authors significantly contributed to this study.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

E-ISSN: 3041-8828

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

We hereby thank all individuals for participating and cooperating us in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

In this study, to observe ethical considerations, participants were informed about the goals and importance of the research before the start of the interview and participated in the research with informed consent.

E-ISSN: 3041-8828

References

Agbir, J. D. (2021). *Relative Efficiency of Classical Test and Item Response Theories in Test Scores Equating Methods using Standardized Chemistry Achievement Tests Unpublished* Federal University of Agriculture Makurdi Benue State.

Asiret, S., & Sünbül, S. Ö. (2016). Investigating Test Equating Methods in Small Samples Through Various Factors. *Educational Sciences: Theory and Practice*, *16*(2), 647-668. https://eric.ed.gov/?id=EJ1101189

Diahyleva, O. S., Gritsuk, I. V., Kononova, O. Y., & Yurzhenko, A. Y. (2021). Computerized adaptive testing in educational electronic environment of maritime higher education institutions. *CTE Workshop Proceedings*, *8*, 411-422. https://doi.org/10.55056/cte.297

Kiss, M., & Kónya, E. (2020). Is it possible to develop some elements of metacognition in a Mathematics classroom environment? *Teaching Mathematics and Computer Science*, *18*(3), 123-132. https://doi.org/10.5485/TMCS.2020.0485

Nelson, G., Kiss, A. J., Codding, R. S., McKevett, N. M., Schmitt, J. F., Park, S., Romero, M. E., & Hwang, J. (2023). Review of curriculum-based measurement in mathematics: An update and extension of the literature. *Journal of School Psychology*, *97*, 1-42. https://doi.org/10.1016/j.jsp.2022.12.001

Odinko, M. (2014). Evaluation research, theory and practice. In: Giraffe books.

Onuka, A. O. U. (2013). *Issues in Contemporary Evaluation*. Nigeria PG Press.

Orluwene, G. (2012). Fundamentals of testing and non-testing tools in educational psychology. *Port Harcourt Harley publications coy*.

Ozyurt, H., Ozyurt, O., Baki, A., & Guven, B. (2012). An Application of Individualized Assessment in Educational Hypermedia: Design of Computerized Adaptive Testing System and its Integration Into UZWEBMAT. *Procedia - Social and Behavioral Sciences*, *46*, 3191-3196. https://doi.org/https://doi.org/10.1016/j.sbspro.2012.06.035

Peters, S. (1998). Playing games and learning mathematics: The results of two intervention studies. *International Journal of Early Years Education*, *6*(1), 49-58. https://www.tandfonline.com/doi/abs/10.1080/0966976980060105

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*(2), 110-122. https://doi.org/10.1016/j.lindif.2009.10.005

Rezaie, M., & Golshan, M. (2015). Computer adaptive test (CAT): Advantages and limitations. *International Journal of Educational Investigations*, *2*(5), 128-137.

Willis, B. H., & Hyde, C. J. (2015). What is the test's accuracy in my practice population? Tailored meta-analysis provides a plausible estimate. *Journal of Clinical Epidemiology*, *68*(8), 847-854. https://doi.org/10.1016/j.jclinepi.2014.10.002

Yang, A. C. M., Flanagan, B., & Ogata, H. (2022). Adaptive formative assessment system based on computerized adaptive testing and the learning memory cycle for personalized learning. *Computers and Education: Artificial Intelligence*, *3*, 100104. https://doi.org/https://doi.org/10.1016/j.caeai.2022.100104

E-ISSN: 3041-8828