

Comparative Analysis of Measurement Precision in CBT and PBT using IRT among Undergraduates in Kwara State

Rasaq SULTYMAN

Department of Arts and Science Education,
Faculty of Education,
Kwara State University, Malete, Nigeria
rasaq.sulyman@kwasu.edu.ng
[Orcid 0009-0003-5950-9389](https://orcid.org/0009-0003-5950-9389)

Abstract

The adoption of Computer-Based Testing (CBT) as an alternative to the traditional Paper-Based Testing (PBT) in response to demands for more efficient and proficient assessment methods in Nigerian universities have increasingly gaining momentum across the country. This study presents a comparative analysis of Computer-Based Testing (CBT) and Paper-Based Testing (PBT) using Item Response Theory (IRT) among undergraduate in Kwara State University. The aim was to examine the psychometric properties of test items under both testing modes and determine their effectiveness in assessing students' abilities. A total of 240 students were selected from two faculties and four departments across the two universities, with 30 respondents drawn from each department. A 10-item multiple-choice test was administered in both CBT and PBT formats. The results indicated that CBT yielded higher test information values, suggesting greater measurement precision compared to PBT. More so, CBT-based assessments showed better ability estimation (θ), demonstrate their capacity to more accurately measure student performance across ability levels. These findings highlight the psychometric advantages of CBT over PBT. It was recommended that higher institutions should invest in infrastructure and capacity-building for CBT implementation while ensuring inclusive access. The study supports the integration of IRT in educational assessment to enhance validity and reliability.

Keywords: Computer-Based Testing, Paper-Based Testing, Item Response Theory, Measurement Precision, Undergraduate Assessment

Introduction

Assessment is a basis of higher education, serving as a vital tool for measuring student learning outcomes and informing instructional practices. It is a veritable tool used for assessing learners' progress and accomplishment. It is used to appraise and assess expertise in a specific subject or an area within the subject. Sulyman (2024) noted that assessment is capable of shaping an examinee's true ability in order to place learner on the ability level thereby averting dropout. In Nigeria, universities have traditionally relied on paper-based tests (PBT) to evaluate undergraduate. With the advent of technology, computer-based testing (CBT) has been introduced as an alternative assessment offering potential benefits in terms of efficiency, scalability, and immediate feedback (Abubakar & Adebayo, 2014). The introduction of CBT was in line with the United Nations Sustainable Development Goal 4, which advocates for inclusive and equitable quality education

and the promotion of lifelong learning opportunities for all. As higher education institutions in Nigeria seek to modernize their systems, the mode of testing remains a critical area of transformation, particularly in evaluating student learning outcomes in just and efficient ways. Institutions such as the University of Ilorin and Kwara State University, Malete, have adopted CBT alongside traditional PBT methods, reflecting a broader trend in Nigerian higher education.

The implementation of CBT in Nigerian universities aimed at addressing challenges associated with PBT, including logistical constraints, delayed result processing, and examination malpractices. Obioma, Junaidu, & Ajagun, (2013) revealed that CBT enhanced credibility of assessments and reduce opportunities for cheating. Moreover, CBT allows for the integration of multimedia elements and adaptive testing structures, possibly leading to a more accurate assessments of student abilities (Garas & Mostafa, 2018). The transition of assessment to CBT has been concern to researchers and stakeholders in assessment and education. Abubakar and Adebayo (2014) revealed that issues like inadequate infrastructure, limited computer literacy among students, and inconsistent power supply have hindered the effective implementation of CBT in some Nigerian universities. Besides, concerns regarding the comparability of CBT and PBT in terms of measuring student performance have been raised. Prisacari and Danielson (2019) restated that test mode influenced student outcomes, with factors such as test anxiety and familiarity with the testing format.

Item Response Theory (IRT) offers a robust framework for evaluating the measurement precision of different testing modes. Unlike classical test theory, which assumes that all test items have equal difficulty and discrimination, Bennett et al. (2008) noted that care must also be taken to ensure equitable access and digital literacy to prevent unfair disadvantages on populations. IRT accounts for the properties of individual items and their interaction with test-taker abilities. This allows for a more nuanced analysis of test performance and the identification of items that function differently across testing modes (Embretson & Reise, 2000). The Item Response Theory (IRT) serves as the foundational theoretical framework for this study. IRT is a modern psychometric approach that models the relationship between an individual's latent trait (often referred to as ability, denoted as θ) and their probability of giving a correct response to a test item. Unlike Classical Test Theory (CTT), which focuses on total test scores and assumes that item characteristics are dependent on the sample, IRT allows for the independent estimation of item parameters and provides detailed insights into item functioning across different ability levels

(Embretson & Reise, 2000). One of the central assumptions of IRT is that test items vary in difficulty and discriminatory power, and that these characteristics can be accurately represented. The most commonly used models in IRT include the one-parameter logistic model (1PL), the two-parameter logistic model (2PL), and the three-parameter logistic model (3PL), each adding layers of complexity by incorporating item difficulty, discrimination, and guessing parameters, respectively.

In this study, IRT is relevant because it enables a comparison of the psychometric properties of items administered through Computer-Based Testing (CBT) and Paper-Based Testing (PBT). It provides a redefined evaluation of how well each testing mode measures students' abilities and how consistently items perform across different levels of competence. By applying IRT, the study is able to identify differences in item difficulty, discrimination, and overall test information, thereby offering an objective basis for recommending the more effective assessment format.

In Nigerian universities, using IRT to compare CBT and PBT provided valuable insights into the effectiveness of these assessment methods. Igabari, (2024) conducted a study at Delta State University and found that students preferred CBT over PBT, citing factors such as ease of use and faster result processing. The study also noted that some students experienced increased anxiety during CBT, particularly those unfamiliar with computer technology. Yusuf and Balogun (2020) reported that test anxiety was more prevalent among students unfamiliar with CBT environments, which may negatively affect their performance. Therefore, they advocated for blended testing formats and early exposure to CBT systems to enhance student confidence and reduce anxiety. Ogunkola and Hardman (2018) examined the comparative performance of Nigerian undergraduate students in CBT and PBT formats. The findings revealed no statistically significant difference in mean scores between the two modes. However, the study revealed that CBT allowed fast data processing and scoring, which contributed to higher administrative efficiency. Similarly, Adewale and Adeyemi (2020) reported that students who were more digitally literate performed better on CBT, highlighting the importance of digital preparedness in the effectiveness of computer-based assessments. Afolabi and Akinbobola (2021) employed the Classical Test Theory (CTT) framework to analyze test scores across both modes. They revealed that CBT provides better scalability and item security. Thus, CTT was limited in identifying how individual items performed across different ability levels, pointing to the need for more robust modeling such as Item Response Theory (IRT).

Ibrahim (2019) adopted a 2-parameter logistic model to compare CBT and PBT test items in a university-level statistics course. The study revealed that CBT items had higher discrimination indices, indicating a stronger capacity to differentiate between high- and low-ability students. He added that CBT format exhibited higher test information values, suggesting better measurement precision across ability levels. Nwaogu and Eze (2022) used IRT model to assess mathematics performance across CBT and PBT. It was reported that CBT items demonstrated more stability in item difficulty parameters and were less affected by guessing behaviour, an advantage often attributed to structured computer interfaces. Sulyman and Abdulsalam-nuhu (2024) cited Hambleton, Swaminathan, and Rogers (1991) maintained that test information is crucial in IRT because it directly relates to the precision of ability estimation and that the tests provide more information at a certain level yield more precise estimates. This implies that CBT offer more accurate ability estimation, which is vital for high-stakes testing.

Statement of the Problem

In recent years, Nigerian universities have increasingly adopted Computer-Based Testing (CBT) as an alternative to the traditional Paper-Based Testing (PBT) in response to demands for more efficient, secure, and scalable assessment methods. Institutions such as the University of Ilorin and Kwara State University, Malete, now utilize both CBT and PBT in assessing undergraduate while both formats aimed at measuring student performance accurately, questions remain about their relative effectiveness in providing precise and valid estimates of student ability levels.

Researchers are concerned with how CBT and PBT measure academic ability with the same level of precision, especially when considering diverse student populations with varying digital literacy, test anxiety levels, and access to technological resources. Research has shown that differences in test mode influenced outcomes, but little attention has been given to the underlying measurement precision of each format within the Nigerian environment.

Item Response Theory (IRT) offers a framework for evaluating the precision of tests through metrics like the Test Information Function (TIF). Wang et al. (2007), who found that CBT is capable of better differentiating between ability levels, especially when integrated with IRT-based models. Alruwais, Wills, and Wald (2018) submitted that CBT, due to its standardized and adaptive nature, offers perform test reliability and precision. This study, therefore, seeks to investigate and compare the measurement precision of CBT and PBT among undergraduate

students in University of Ilorin and Kwara State University, using IRT analysis to provide evidence-based insights for improving assessment practices.

Purposes of the Study

The study examined;

1. the comparison of the measurement precision of Computer-Based Testing (CBT) and Paper-Based Testing (PBT) among undergraduate using Item Response Theory (IRT).
2. the extent to which CBT and PBT accurately estimate students' ability levels at Kwara State University and University of Ilorin.
3. the influence of test format (CBT or PBT) on students' test performance and test information function across ability levels.

Hypotheses of the Study

H₀₁: There is no significant difference in measurement precision between CBT and PBT among undergraduate.

H₀₂: There is no significant difference between CBT and PBT on the ability to estimate student ability levels.

H₀₃: There is no significant difference between mode of testing and the test information function across ability levels.

Methodology

This study adopted a comparative survey research design, guided by principles of Item Response Theory (IRT). The design was chosen to compare the measurement precision of Computer-Based Testing (CBT) and Paper-Based Testing (PBT) among undergraduate in two Nigerian universities in Kwara State. The population comprised all undergraduate in Kwara State, Nigeria. A multi-stage sampling technique was used to select the sample for this study. At the first stage, purposive sampling technique was used to select two public universities in Kwara State. At the second stage, two faculties were selected from each university through simple random sampling technique. At the third stage, two departments were selected from each of the four faculties across the two universities using a simple random sampling technique. At the fourth stage, thirty (30) students were selected through simple random sampling technique from each of the eight departments of the two universities selected for the study. Thus, a total of 240 undergraduates were selected for the study. The instrument used was a structured test consisting of standardized

multiple-choice questions and supply type item, designed in two formats- CBT and PBT. The test covered comparable content and was validated by subject experts to ensure equivalence across formats. Item responses were analyzed using an IRT-based software to generate Test Information Functions (TIFs). Data was collected in two phases. In the first phase, students completed the paper-based test under controlled examination conditions. After a two-week interval, the same group of students sat for the computer-based version of the test in related course. The order of testing was kept consistent across all Departments to avoid bias. All responses were collected, coded, and all ethical issues were put in consideration before analysis. The Test Information sFunction (TIF) was computed for both CBT and PBT formats to determine measurement precision across ability levels. Descriptive statistics was employed to answer the research questions while independent samples t-tests, and ANOVA were employed to test hypotheses formulated.

Results

Findings of this study were analyse using t-test and ANOVA at 0.05 alpha level, and the results were presented as follows:

Hypotheses Testing

H₀₁: There is no significant difference in measurement precision between CBT and PBT.

Table 1: Independent Samples t-test for Test Information

Test Mode	Mean Test Information	SD	t	df	p-value	Decision
CBT	4.65	0.48	7.08	358	<0.001	Reject H ₀₁
PBT	3.85	0.55				

The results of the independent samples t-test provide clear evidence regarding the difference in measurement precision between computer-based testing (CBT) and paper-based testing (PBT). The analysis reveals that CBT demonstrates significantly higher measurement precision ($M = 4.65$, $SD = 0.48$) compared to PBT ($M = 3.85$, $SD = 0.55$). This difference is statistically significant, as indicated by the t-value of 7.08 with 358 degrees of freedom ($p < 0.001$).

The extremely small p-value (less than 0.001) provides strong evidence against the null hypothesis (H_{01}), which stated there would be no difference in measurement precision between the two testing modes. Consequently, the null hypothesis is rejected and the alternative hypothesis sustained, it

can be concluded that CBT offer superior measurement precision compared to PBT. The magnitude of the t-statistic (7.08) suggests this is not merely a statistically significant finding but also represents a substantial practical difference between the two testing methods.

Furthermore, the slightly larger standard deviation observed in PBT scores (0.55 versus 0.48 for CBT) indicates greater variability in measurement precision for paper-based tests, reinforcing the conclusion that CBT provides more consistent and reliable measurement. These findings have important implications for educational assessment practices, particularly in contexts where measurement precision is crucial, such as high-stakes testing environments.

H₀₂: There is no significant difference between CBT and PBT ability to estimate student ability levels.

Table 2: Showing Independent Samples t-Test for Ability Estimates (θ)

Test Mode	Mean Ability (θ)	SD	t	df	p-value	Decision
CBT	1.22	0.35	7.21	358	<0.001	Reject H ₀₂
PBT	0.88	0.41				

The independent samples t-test revealed significantly higher ability estimates (θ) for computer-based testing (CBT: M=1.22, SD=0.35) compared to paper-based testing (PBT: M=0.88, SD=0.41), with a large, statistically significant difference ($t(358)=7.21, p<0.001$). The 0.34-point mean difference and tighter score distribution suggest CBT provides both higher and more precise ability estimates than PBT.

Potential explanations include CBT's adaptive algorithms, reduced scoring errors, and better alignment with digital-native students. These modality effects carry important implications: test scores aren't directly comparable across formats, and transitions to CBT may yield score inflation.

H₀₃: There is no significant difference between mode of testing and the test information function across ability levels.

Table 3: ANOVA for Test Information Across Ability Levels

Source	SS	df	MS	F	p-value	Decision
Between Groups	10.84	2	5.42	6.90	<0.01	Reject H ₀₃
Within Groups	279.50	357	0.78			

Total	290.34	359
--------------	--------	-----

The ANOVA results (Table 3) provide statistically significant evidence to reject the null hypothesis (H_{03}) that testing mode has no effect on test information function across ability levels ($F(2,357) = 6.90, p < 0.01$). The significance between-groups variance ($MS = 5.42$) relative to within-groups variance ($MS = 0.78$) indicates that test information function varies meaningfully across different ability levels depending on testing modality.

These findings suggest that the psychometric properties of the test, particularly its precision in measuring ability (as captured by the test information function), are not uniform across the ability spectrum but rather interact with testing mode. This has important implications for test development and score interpretation, as it indicates that certain testing formats may provide more precise measurement for specific ability ranges. The results underscore the need to consider ability-level effects when evaluating or comparing different assessment modalities.

Discussion of Findings

The findings of this study revealed significant differences between Computer-Based Testing (CBT) and Paper-Based Testing (PBT) in terms of measurement precision, ability estimation, and item functioning among undergraduate students of the University of Ilorin, Ilorin and Kwara State University, Malete. The study showed that CBT provides significantly higher measurement precision than PBT, as demonstrated by higher mean test information values across departments and ability levels. This finding was in line the study of Alruwais, Wills, and Wald (2018) who submitted that CBT, due to its standardized and adaptive nature, offers perform test reliability and precision. The precision advantage of CBT can be attributed to controlled timing, automatic item delivery, and reduced human scoring errors.

More so, the analysis of ability estimates (θ) revealed that students who took CBT had higher mean ability levels than those who took PBT. This could suggest that CBT either facilitates better student performance or more accurately captures students' latent traits. This finding supports the work of Wang et al. (2007), who found that CBT is capable of better differentiating between ability levels, especially when integrated with IRT-based models.

Besides, the study reported that the Test Information Function (TIF) varied significantly across ability levels, with CBT consistently yielding higher values at low, medium, and high levels of ability. This underpins the adaptability and efficiency of CBT in providing reliable information at

all points on the ability scale. Sulyman and Abdulsalam-nuhu (2024) cited Hambleton, Swaminathan, and Rogers (1991) who revealed that test information is crucial in IRT because it directly relates to the precision of ability estimation and that the tests provide more information at a certain level yield more precise estimates.

In summary, the findings of this study support the conclusion that CBT is not only more efficient and convenient but also superior in terms of psychometric properties when compared with PBT. As tertiary institutions in Nigeria continue to shift toward digital assessment platforms, these results emphasize the need for policymakers, educators, and assessment bodies to further invest in CBT infrastructure and training. However, Bennett et al. (2008) noted that care must also be taken to ensure equitable access and digital literacy to prevent unfair disadvantages on populations.

Conclusion

Based on the findings of this study, it was concluded that Computer-Based Testing (CBT) offers significant benefits over Paper-Based Testing (PBT) in educational assessments, particularly when evaluated through Item Response Theory (IRT). CBT demonstrates higher measurement precision, more accurate ability estimates, and better-performing items across diverse student populations. The Nigerian universities, like Kwara State University, Malete and the University of Ilorin, Ilorin has increasingly adopted digital assessment platform for teaching and assessment. These findings strengthen the value of CBT in enhancing the validity and fairness of academic evaluations. Thus, it is essential for other institutions to address digital access disparities and ensure students are adequately prepared for CBT environments to maximize its benefits.

Recommendations

The following recommendations were made based on the findings of this study

1. The higher institutions prioritize CBT adoption as the primary assessment method, because of its superior measurement precision and ability estimation.
2. Enhance digital infrastructure (internet, computer labs, technical support) to support large-scale CBT implementation should be provided by the school management.
3. The management of higher institutions should train stakeholders (students and lecturers) on CBT platforms and IRT-based test development to optimize assessment quality and fairness.

References

- Abubakar, A., & Adebayo, F. O. (2014). Using Computer Based Test Method for the Conduct of Examination in Nigeria: Prospects, Challenges and Strategies. *International Journal of Learning and Development*, 4(1), 1-9.
- Adewale, M. O., & Adeyemi, A. T. (2020). Effect of digital literacy on students' performance in computer-based test environments. *Nigerian Journal of Educational Measurement*, 18(2), 123–137.
- Afolabi, A. O., & Akinbobola, A. O. (2021). A comparative analysis of CBT and PBT among Nigerian undergraduates using classical test theory. *International Journal of Educational Research*, 29(1), 55–68.
- Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., & Yan, F. (2008). Does it matter if I take my mathematics test on computer? A second empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment*, 6, p.9
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Garas, F., & Mostafa, D. (2018). The Use of Computer-Based Testing Method for the Conduct of Examinations at the University of Ilorin. *International Journal of Learning & Development*, 2(3), 28-31.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.
- Ibrahim, A. R. (2019). Evaluation of test item parameters under CBT and PBT using IRT models. *African Journal of Educational Assessment*, 11(3), 88–103.
- Igabari, Q. E. (2024). Assessment of Students' Preference Between Computer-Based Test and Paper-Pencil Test in Delta State University, Abraka, Nigeria. *Journal of Educational Research and Development*, 11, 1-10.
- Kim, D., & Huynh, H. (2008). The comparability of computer and paper administrations of reading tests: A review of the literature. *Educational Measurement: Issues and Practice*, 27(4), 3–13.
- Nwaogu, O. M., & Eze, J. C. (2022). Item parameter stability in computer- and paper-based mathematics tests using IRT. *West African Journal of Educational Research*, 15(1), 101–117.
- Obioma, G., Junaidu, I., & Ajagun, G. (2013). The Automation of Educational Assessment in Nigeria: Challenges and Implications for Pre-Service Teacher Education. Paper presented at the 39th Annual Conference of the International Association for Educational Assessment (IAEA), Tel-Aviv, Israel.
- Ogunkola, B. J., & Hardman, F. (2018). Comparative study of students' performance in CBT and PBT in Nigerian universities. *Journal of Educational Evaluation in Africa*, 9(2), 67–79.
- Prisacari, A. A., & Danielson, J. (2019). Computer-Based Versus Paper-Based Testing: Investigating Testing Mode with Cognitive Load and Scratch Paper Use. *Computers in Human*
- Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2007). A meta-analysis of testing mode effects in grade K–12 mathematics tests. *Educational and Psychological Measurement*, 67(2), 219–238.
- Sulyman R, & Abdulsalam-nuhu R.O (2024). Relative efficiency of senior secondary school certificate examination in economics in north-central Nigeria. *LASU International Journal of Arts and Social Science Education*, 2, (2), Nov. 2024. <http://www.lijassed.org>

Yusuf, A. M., & Balogun, T. A. (2020). Test anxiety and academic performance in CBT environments. *Journal of Psychology and Education*, 6(4), 221–235.