



DIROSAT

Journal of Education, Social Sciences & Humanities

Journal website: <https://dirosat.com/>

ISSN : 2985-5497 (Online)

DOI: <https://doi.org/10.58355/dirosat.v4i2.214>

Vol. 4 No. 2 (2026)

pp. 193-203

Research Article

The Test of Testing: Evaluating Testee Perception of Ethical Compliance and Social Responsibility in High-Stakes Examinations

Afeez Tunde Jinadu

Nigerian Institute of Social and Economic Research Ibadan, Nigeria;
afeezjinadu95@gmail.com



Copyright © 2026 by Authors, Published by **DIROSAT: Journal of Education, Social Sciences & Humanities**. This is an open access article under the CC BY License <https://creativecommons.org/licenses/by/4.0/>

Received : February 06, 2026
Accepted : April 14, 2026

Revised : March 29, 2026
Available online : May 16, 2026

How to Cite: Jinadu, A. (2026). The Test of Testing: The Test of Testing: Evaluating Testee Perception of Ethical Compliance and Social Responsibility in High-Stakes Examinations. *DIROSAT: Journal of Education, Social Sciences & Humanities*, 4(2), 193-203. <https://doi.org/10.58355/dirosat.v4i2.214>

Abstract. The study investigated the examination bodies' compliance with ethics and social responsibility in high stake examinations from the testees' perspectives. The study used ex-post facto design. Simple random technique was used to select two thousand and five hundred respondents drawn from two states in South-Western Nigeria. Testees' Response to Ethics and Responsibility Questionnaire ($\alpha = 0.862$) was the instrument used to collect data. Frequency count and Analysis of variance ANOVA were used to analyse the data collected. The level of compliance varies across the examination bodies as responded by the testees [$F(2,2498) = 13.256; p < 0.05$] in favour of WAEC testees ($M=15.32; SD=2.37$) followed by NECO testees ($M=12.99; SD=2.62$) and lastly NABTEB testees ($M=12.98; SD=2.55$). It was found out that the difference was only significant for WAEC using pairwise comparison while the difference between NECO and NABTEB was not significant. Three public examination bodies were studied and the compliance with ethics and social responsibility may not be generalized beyond this context. The type of examination undertaken by testees varies in the way and

manner in which they comply with ethics and social responsibilities in assessment. For high ethical and social responsibility in high stake examination, there is need for total compliance. Therefore, examination bodies that are non-compliant with ethics and social responsibility in assessment should seek ways of improving their development and administration of test items along that line.

Keywords: Compliance, Examination, Testee, Ethics and Responsibility

INTRODUCTION

As countless children take tests of various forms across the globe every day, the terminal and certification tests at the secondary school level always have been important tools in assessing learners' academic achievement for that particular level of education. The mandates for these tests originate outside of the classroom with different examination bodies conducting such tests in Nigeria at state, federal and of course regional levels. These are the respective examination bodies, the National Examination Council (NECO), the National Board for Technical and Business Education (NABTEB) and the West African Examinations Council (WAEC). The results of these assessments are used primarily to inform decisions made by the policymakers and administrators who shape and direct the provision of education. They often help decision-makers monitor the ability and achievement of students and determine eligibility for access to the next level of educational programs and resources.

Away from the secondary school level, in some cases, test results become the principal if not only measure by which institutions and educators are judged in this "age of accountability." They often play a pivotal role in important decisions about the future of programs, the level and allocation of funding, salary increases, and whether administrators or teachers are praised or sanctioned as the case may be. Even when the results do not form the foundation of official decisions, the results are likely to be taken by the public to apportion blame or praise, to distinguish schools or teachers as either good or bad, and in other situations to advance political and educational courses.

Despite the motivation by a sincere desire to increase the quality of education, the infusion of high stakes into the administration of large-scale statewide or nationwide testing programs and into the use of the results has put intense pressure on our educators to improve the test scores of students. This pressure has created a climate in which test performance drives what is taught, what is not taught, and how it is taught. This atmosphere may distort the teaching process, restrict the scope of the education that schools provide, and influence the practices and conducts of the tests as well as the results of the test itself. This is a question of ethics and social responsibilities in assessment by the examination bodies which will ultimately have implications on the integrity, credibility as well as usability of such test results. Ethics in assessment refers to striving to maintain focus on the fundamental reason for any testing program or the education of the students being assessed (Jinadu, 2024; Reuben & Eremie, 2020; Abe, 2012). Airasian (2005) suggests that the ethical standards for assessment refer to some aspect of a teacher's fairness in dealing with his or her pupils. Similarly, Taylor and Nolen (2005) point out that because poor assessment can

significantly affect students, the ethical responsibility of educators is 'first, Do No Harm'. The most effective way to achieve this goal is to ensure that the tests used are reliable and that no practices used in preparing for and administering the tests detract from their validity (Gregor, 2014).

Of course, any practice contrary to the fundamental reason for any testing program is termed unethical practice in assessment. By using such practices, a number of laid down protocols are distorted which may inappropriately raise test scores. In doing so by the parties involved, teachers may believe that they are "saving their jobs," administrators may be attempting to "promote" their schools or districts, a politician may be attracting attention, or others may be acting out of self-interest. However, the roots of inappropriate or unethical testing practices can be more complex. The causes can originate with anyone involved in assessments, including those who develop the tests, the policymakers and administrators who choose assessments and interpret or act upon the results, and the educators who prepare students for tests and administer them (Jinadu, 2024; Reuben & Eremie, 2020; Washington Educational Research Association- WERA, 2001; Haladyna, Nolen & Haas, 1991).

Similarly, when testing programs are selected, administered, and used appropriately, they can make a valuable contribution to the nation's educational system. The misuse of testing is probably unintentional in many instances. Individuals involved in the preparation and conduct of examinations may not completely understand their proper roles or the range of appropriate practices expected in standardized testing. The appropriate roles are the responsibilities in assessment. These responsibilities include selecting an assessment instrument, preparing students for testing, administering the test itself, and interpreting or using the results as derived from various codes, guidelines, and suggestions advanced by the professional assessment community and others (WERA, 2001; Haladyna, Nolen & Haas, 1991).

Developing and selecting assessments require developers, policymakers, and administrators to make important choices that can have a significant impact on the usefulness of the assessments and the validity of the results (Murchan and Siddiq, 2021). Those involved in developing and selecting assessments are responsible for ensuring that the instruments are well crafted and suited to the students being assessed. Hence, an assessment should be selected only when it satisfies the specific purposes for which it is to be used and is appropriate for the intended population(s). Therefore, the first responsibility of those who select tests is to define clearly the purpose(s) for testing and understand the characteristics of the population that they wish to assess. Others are to preserve the security of the assessments being reviewed to undermine efforts to raise test scores through inappropriate preparation practices. For the same reason, tests should be kept secure during the development process (National Council on Measurement in Education-NCME Task Force, 1991).

Similarly, during administration, disclosure of purposes of testing to all parties involved is essential. Monitoring the administration of the test, dealing with breaches of protocols and a good testing environment is key. This supervision can include unannounced observations and interviews of those administering the assessment.

Also, appropriate security precautions should be taken before, during, and after assessments are administered. It is important to follow the conditions or procedures prescribed by the developer of an assessment. The test also should be administered in such a way that sources of potential bias are eliminated. In other words, student evaluations should be ethical, fair, useful, feasible, and accurate (Joint Committee on Standards for Educational Evaluation-JCSEE, 2003).

Popham (1991) offered two other general standards to determine whether a particular practice is appropriate (1) Test preparation should not violate the ethical standards of the education profession. Educators should not violate general ethical standards concerning theft, cheating, lying, and the like. In addition, educators must realize that because they act in place of the parent they have an ethical obligation to serve as models of behaviour for their students. This ethical foundation should inform choices made when preparing students. (2) Test preparation should have "educational defensibility." Under this concept, a test preparation activity that raises student test scores is inappropriate unless it simultaneously increases student mastery of the content domain tested. Test preparation, as with any instructional activity, should be employed in the best interest of the students. Accordingly, because inappropriate test preparation practices deprive students of a portion of their education and deceive them (and others) about their true mastery of a subject, such activities are educationally indefensible.

Murchan and Siddiq, (2021) investigated an overview of how ethical and regulatory requirements interface with process data from assessments in primary and secondary education (K-12), by conducting a systematic literature review. Initial results showed that few studies considered ethical, privacy and regulatory issues in K-12 assessment, prompting a widening of the search criteria to include research in higher education also, which identified 22 studies. The literature that was relevant to the research questions represented an approximate balance in the number of theoretical and empirical studies. The studies identified as relevant interpret issues of privacy largely in terms of informed consent and the research pays little attention to ethical and privacy issues in the use of process data in assessment. The implications for the field of educational assessment and the use of process data were discussed to include the need to develop a specific code of ethics to govern the use of file data in educational assessment.

Reuben & Eremie (2020) investigated shaping the future and ethics of psychological testing. It considered some of the professional concerns that continually play an essential role in the present and future states of psychological testing such as theoretical concerns, test adequacy, human rights, labelling, privacy invasion, test constructors/test users' responsibility, dehumanization, the usefulness of tests, and current fashion in psychological testing like the development of new tests, increase in public awareness and influence on psychological testing, and the hope of new and improved tests. With the increase in the level of awareness of test users and the psychological needs of humans, especially with the changes in human body chemistry which sometimes may have some psychological implications on human personality, Psychologists should be more creative in building new tests that

will meet future testing needs of the fast-growing population and be persistent in modifying the existing tests while accomplishing the goals of psychological testing.

Trina, Yong, Jiayi and Danqing (2019) synthesized the major concerns that have been raised in the literature concerning large-scale assessments (LSAs) in education. The study completed a deep review of the literature on LSAs to synthesize the reported side effects. The review was synthesized thematically to understand and report the consequences of the ongoing push for the use of LSA in education. The study's thematic analysis indicated overarching side effects of LSA in education and discussed why negative side effects exist and presented evidence of the most commonly observed side effects of LSA in education, including distorting education, exacerbating inequity and injustice, demoralization of professionals, ethical corruption, and stifling of innovation in education. Originality/Value: While concerns about the use and misuse of LSA in education are not new and have been discussed widely in the literature, rarely have they been discussed as inherent qualities and consequences of LSAs that can harm education. Similarly, Susan, Robert, Do-Hong, and Nakia (2007) investigated ethical behaviour and examined educators' ethical judgments concerning assessment. It describes the results from a web-based survey of educators in which they read a brief scenario and indicated whether the student evaluation practice in the depiction was ethical or unethical. Results showed strong agreement among the educators on fewer than half of the scenarios presented in the study. The findings of the study suggest that assessment is currently an educational realm without professional consensus.

The assessment literature provides some guidance for teachers and other test administration officials in terms of ethical and unethical practices in standardized testing as put forth by previous studies; however, few scholarly resources in assessment directly address the comparison of the ethics of assessment practices by various examination bodies. There appears to be a dearth of literature on the effect of the type of examination body on compliance with ethics and social responsibility. Therefore, the study investigated the examination bodies' compliance with ethics and social responsibility in assessment from the testees' perspective in southwest Nigeria.

Research Questions

1. What is the profile of testees in South-Western Nigeria?
2. Is there any significant difference among the testees' responses to examination bodies' compliance with ethics and social responsibility in assessment?

RESEARCH METHODS

The study used *ex-post facto* design. This design became necessary because it allowed the researchers to compare the different testees and draw inferences. The multi-stage sampling procedure was adopted to select the sample. In the first stage, South-Western Nigeria was stratified along the existing six states and simple random sampling was used to select two states (Oyo and Osun). In the third stage, testees were stratified into WAEC, NECO and NABTEB testees. In the fourth stage, population proportion to size was used to select 61 WAEC centres, 40 NECO centres and 24 NABTEB centres. In the fifth stage, simple random sampling was used to select

twenty testees from each of the centres. A total of 2,500 testees comprising one thousand two hundred and twenty (1220) WAEC candidates, eight hundred (800) NECO candidates and four hundred and eighty (480) NABTEB candidates were selected across one hundred and twenty-five examination centres participants for the study.

Testees' Response to Ethics and Responsibility Questionnaire (TREREQ) was used to collect data. TREREQ was developed by the researcher to measure testees' responses to examination bodies' compliance with ethics and social responsibility in assessment. It consists of two sections A and B. Section A sought participants' demographic information such as the name of the school, gender, age, type of examination and number of examinations to be taken. Section B is on examination bodies' compliance to ethics and social responsibility in assessment based on seven indicators: communication about grading, multiple assessment opportunities, confidentiality, standardized test preparation, standardized test administration, bias and grading practices. The initial test contain forty-eight items in which participants were asked to respond on a four-point scale of always-4, sometimes-3, rarely-2 and never-1 however, the scoring was reversed for negative items. These items were subjected to pilot testing using testees who were not part of the final sample for the study.

The content validity was established by giving the draft to psychometricians in the field of assessment and testing, where irrelevant items were deleted or modified and others subsequently retained. To determine the reliability of the instrument, the internal consistency of the instrument was obtained using Chronbach's Alpha method of reliability which yielded a value of 0.862. The researcher monitored the data collection exercise. Four research assistants were co-opted for the study who were trained on the administration of the instrument before the commencement of the study. They were intimated with the objectives of the study and the reasons they were selected. Frequency count and one-way analysis of variance: 1-Way ANOVA were used to analyze the data.

RESULTS AND DISCUSSION

Research Question 1: What is the profile of testees in South-Western Nigeria?

Table 1: Profile of Testees

S/N	Testees' Examination	Gender		Total
		Male	Female	
1	NABTEB	284 11.4%	196 7.8%	480 19.2%
2	NECO	386 15.4%	414 16.6%	800 32.0%
3	WAEC	516 20.6%	704 28.2%	1220 48.8%
4	TOTAL	1186 47.4%	1314 52.6%	2500 100.0%

Table 1 shows the descriptive statistics of the profile of testees in South-Western Nigeria. The result revealed that West African Examination Councils WAEC testees were the highest respondents for the study having one thousand two hundred and twenty 1220 (48.8%) testees who participated in the study comprising 516 (20.6%) male and 704 (28.2%). The table also showed that this is followed by the National Examination Council NECO having 800 (32.0%) of their testees as respondents for the study comprising 386 (15.4%) male and 414 (16.6%) female. The National Board for Technical Business Education had 480 (19.2%) of their testees participate in the study comprising 284 (11.4%) male and 196 (7.8%) female. In all, there were more females 1314 (52.6%) than males 1186 (47.4%) participants in the study.

Research Question 2: Is there any significant difference among testees' responses to examination bodies' compliance with ethics and social responsibility in assessment?

Table 2a: Descriptive Statistics
Descriptive Statistics

Testees' Examination	N	Mean
NECO	800	12.99
NABTEB	480	12.98
WAEC	1220	15.32
Total	2500	

Table 2a shows the descriptive statistics of mean scores of each of the examination body's ethics and social responsibilities as indicated by the testees. The table revealed that WAEC had the highest mean (15.32) score indicating the examination body upholding ethics and social responsibilities in assessment followed by NECO (12.99) and lastly by NABTEB (12.98).

Table 2b: 1-Way Analysis of Variance of Compliance to Ethics and Responsibility by Examination Bodies

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	172.599	2	86.299	13.256	0.000*
Within Groups	16261.98	2498	6.510		
Total	16434.579	2500			

*= significant at $p < 0.05$

Table 2b shows that there is a significant statistical difference among external examination bodies' compliance to ethics and social responsibilities based on testees' responses. This is shown by the p-value being lower than the critical value 0.005 [$F_{(2,2498)} = 13.256; p < 0.05$].

Table 3: Post-hoc Multiple Comparison

(I)Ethics/Responsibility(J)	Ethics/Responsibility	Mean Difference (I-J)	Std. Error	Sig.
WAEC	NABTEB	2.341*	0.481	0.000
	NECO	2.335*	0.492	0.000
NABTEB	WAEC	-2.341*	0.481	0.000
	NECO	-0.007	0.349	1.000
NECO	WAEC	-2.335*	0.492	0.000
	NABTEB	0.007	0.349	1.000

Table 3 shows the post-hoc multiple comparisons of the external examination bodies’ compliance to ethics and social responsibilities based on testees’ responses. The table reveals that WAEC testees have the highest means score (M=15.32; SD=2.37) followed by NECO testees (M=12.99; SD=2.62) and NABTEB testees (M=12.98; SD=2.55). The further showed in the multiple comparisons that the difference between WAEC mean score in ethics and social responsibilities in assessment and that of NECO and NABTEB were statistically significant while the difference between the mean score in ethics and social responsibilities of NECO and NABTEB is not statistically significant. This is further shown in Figure 1.

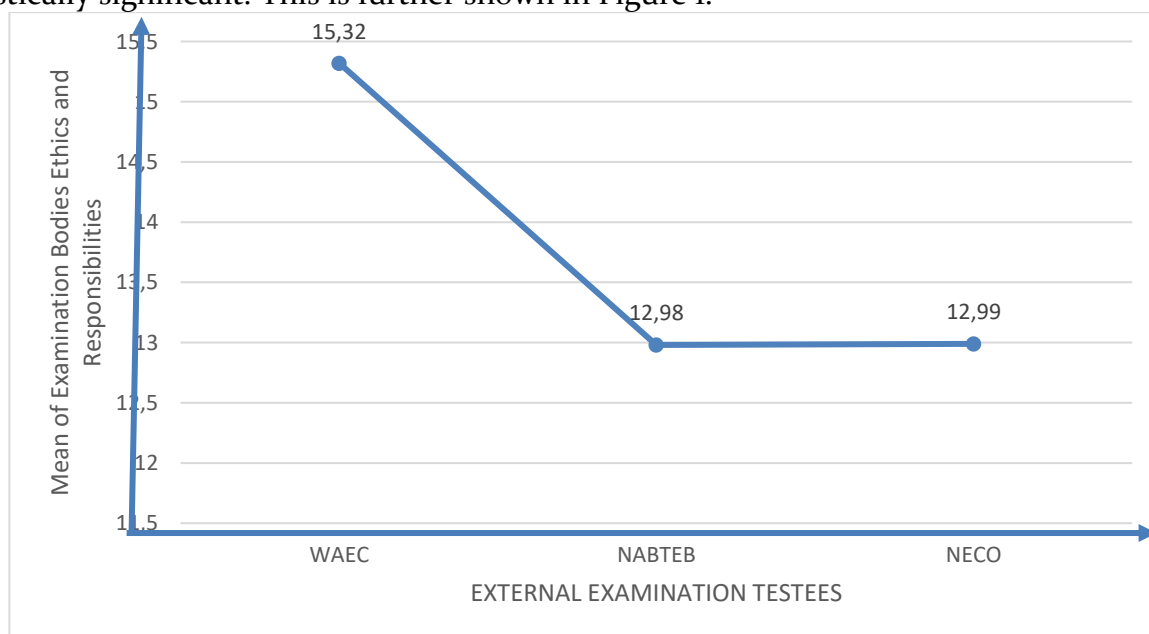


Figure 1: Mean Plot of External Examination Bodies’ Ethics and Social Responsibilities

DISCUSSION OF FINDINGS

The finding on the profile of the testees in terms of external examination sat indicates that there are more WAEC testees than NECO and NABTEB testees. This finding is in line with that of Reuben & Eremie (2020) who found out psychologists should be more creative in building new tests that will meet future testing needs of the fast-growing population of testees and be persistent in modifying the existing

tests while accomplishing the goals of psychological testing. The finding of this study also agrees well with that of Trina, Yong, Jiayi and Danqing (2019) who found out that the major concerns that have been raised in the literature concerning large-scale assessments (LSAs) in education rigour of meeting up the ethical standard in assessment arising from the vastly growing number of examinees. The study's thematic analysis indicated overarching side effects of LSA in education and discussed why negative side effects exist and presented evidence of the most commonly observed side effects of LSA in education, including distorting education, exacerbating inequity and injustice, demoralization of professionals, ethical corruption, and stifling of innovation in education.

The finding concerning significant differences among examination bodies' compliance to ethics and social responsibility in assessment based on testees' responses, indicated that WAEC had the highest means score followed by NECO testees and lastly NABTEB testees. Meanwhile, the finding of this study is limited to seven variables under the ethical and social responsibilities of three examination bodies (communication about grading, multiple assessment opportunities, confidentiality, standardized test preparation, standardized test administration, bias and grading practices) as rated by testees drawn from two south-western part of Nigeria. Also, the findings relied on the testees' rating of examination bodies which may reflect the views of testees used only and not the generality of the testees as well as the views of other education stakeholders.

This finding is in tune with that of Murchan and Siddiq, (2021) who found that relevant interpretation of issues of privacy is largely in terms of informed consent and the research pays little attention to ethical and privacy issues in the use of process data by examination bodies in assessment. The implications for the field of educational assessment and the use of process data were to include the need to develop a specific code of ethics to govern the use of process and log file data in educational assessment.

Similarly, the finding of this study is in tandem with that of Susan, Robert, Do-Hong, and Nakia (2007) who found out that assessment is currently an educational realm without professional consensus as born out of the investigation on ethical behaviour and examining educators' ethical judgments about assessment. It describes the results from a web-based survey of educators in which they read a brief scenario and indicated whether the student evaluation practice in the depiction was ethical or unethical.

CONCLUSION AND RECOMMENDATIONS

It is deduced from the results that the West African Examinations Councils (WAEC) testees were the highest respondents for the study followed by National Examination Councils (NECO) lastly National Board for Technical and Business Education (NABTEB) with more female than male participants in the study. However, the type of examination body affects ethics and social responsibilities in assessment. This implies that the type of examination undertaken by testees varies in the way and manner in which they comply with ethics and social responsibilities in assessment. This portends great danger for the authenticity, acceptability, integrity, credibility as

well as generalizability of the test results within the ambit of the three examination bodies considered in this study. Therefore, the West African Examinations Councils should continue to use their training in the pursuit and improvement of their ethics and social responsibilities in assessment while National Examination Councils and National Board for Technical and Business Education should seek ways of improving the quality of their ethics and social responsibilities in assessment.

BIBLIOGRAPHY

- Abe, C.V. (2012). *Monitoring and evaluation: A panacea of quality assurance in higher education*. Ibadan: Stirling Holden Publishers Ltd.
- Airasian, P. (2005). *Assessment in the classroom: A concise approach* (2nd ed.). Boston, MA: McGraw-Hill Company.
- Gregor, B. (2014). *The Test of Testing: Making Appropriate and Ethical Choices in Assessment*. Oak Brook: North Central Regional Educational Laboratory. pp 1-30.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20, 2-7.
- JCSEE. (2003). *The student evaluation standards*. Arlen Gullickson, Chair. Thousand Oaks, CA: Corwin.
- Jinadu, A. T. (2024). Causal Modelling of Testing and Assessment Ethics among Proctors of Public Examinations in Oyo State. *Arab Gulf Journal of Scientific Research*, 42(3), 660-669. DOI: <http://doi.org/10.1108/AGJSR-01-2023-0017>.
- Murchan, D. and Siddiq, F. (2021). A call to action: a systematic review of ethical and regulatory issues in using process data in educational assessment. *Large-scale Assessments in Education* (2021) 9:25 <https://doi.org/10.1186/s40536-021-00115-3>.
- NCME Task Force. (1991). *Regaining trust: Enhancing the credibility of school testing programs*. Report by a task force of the National Council on Measurement in Education.
- Popham, W.J. (1991). Appropriateness of teachers' test-preparation practices. *Educational Measurement: Issues and Practice*, pp12-15.
- Reuben, N. & Eremie, M. D. (2020). Issues Shaping the Future and Ethics of Psychological Testing. *International Journal of Innovative Psychology & Social Development* 8(4):1-8.
- Susan K. G., Robert L.J., Do-Hong K., and Nakia S. P. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education* 23, pp 999-1011.
- Taylor, K., & Nolen, S. (2005). *Classroom assessment: Supporting teaching and learning in real classrooms*. Upper Saddle River, NJ: Pearson Education, Inc.
- Trina E. E., Yong Z., Jiayi D. and Danqing Y. (2019). Side Effects of Large-Scale Assessments in Education. *ECNU Review of Education*. Vol. 2(3) 279-296. DOI: 10.1177/2096531119878964. <http://www.journals.sagepub.com/home/roe>.

Afeez Tunde Jinadu

The Test of Testing: Evaluating Testee Perception of Ethical Compliance and Social Responsibility in High-Stakes Examinations

WERA (2001). Ethical Standards in Testing: Test Preparation and Administration.
Washington Educational Research Association Professional Publications
Volume 1. Pp 1-11.