

LINEAR AND SEPARATE CALIBRATION METHODS OF EQUATING 2018 AND 2019 ENGLISH LANGUAGE BASIC EDUCATION CERTIFICATE EXAMINATION SCORES

Victor Abiodun Lijofi¹, Oluwasola Damilola Owoeye (Ph.D.)² & Bolarinwa, Adeolu Kolade³

*Department of Educational Foundations and Counselling, University of Ilesa, Osun State.
lijofivictor@gmail.com/07062176715*

*Department of Educational Foundation and Counselling, University of Ilesa, Osun State.
owoeye06@gmail.com/08067849134*

*Dr. Lawrence Omole Library, University of Ilesa, Osun State.
koladeolu@gmail.com/08066847659*

ARTICLE INFO

Article No.: 0123

Accepted Date: 03/12/2025

Published Date: 12/01/2026

Type: Research

ABSTRACT

The study examines efficient techniques for the comparison of students' performance in the English Language Basic Education Certificate Examination (BECE) scores. The study adopted an Ex-Post facto research design. The population for the study comprised 123,441 Junior Secondary School 3 (JSS 111) students who sat for the 2018 and 2019 English Language BECE conducted by the Ondo State Ministry of Education. The study sample consisted of 4,320 Junior JSS 111 students, selected from the population using a multistage sampling technique. The instruments for the study consisted of the students' responses to 60 multiple-choice items from the 2018 and 2019 English language Ondo State BECE, extracted from the Optical Marks Record (OMR) sheet. Data was coded and analyzed using (BILOG-MG and Linear Equating framework implemented through the R language package environment for statistical packages. The research showed that the standard error of equating for the linear method was 0.6, which was lower than the separate calibration method of 0.7 across the obtainable scores. The study further revealed that the 2019 OSBECE English test was more difficult than the 2018 version by about 0.4 units. The research concluded that the linear method was better for equating 2018 and 2019 OSBECE English language tests. Also, the 2018 and 2019 OSBECE English language tests were not equivalent. The research recommended that the Ondo State government and Nigerian examination organizations should utilize the most suitable test theory model to verify the statistical equivalence of their examination test items before administration.

Keywords: Basic Education Certificate Examination, Test, Test score equating, Linear Equating, Separate Calibration Equating

Introduction

The Basic Education Certificate Examination (BECE) is a public examination held in Nigeria by each State Ministry of Education for final year students in Junior Secondary School three (JSSIII) at the end of their third year. Without uniformity in terms of test administration and scoring, all pupils enrolled in public and private schools within a state's jurisdiction are subject to the development, administration, marking, and awarding of grades and certificates. The Basic Education Certificate Examination is considered in this study because it determines pupils who will be promoted from Junior Secondary School to Senior Secondary School, Technical Schools and enhance the awards of Basic Education Certificates in Nigeria. Faleye & Afolabi, (2005) argued that the standard of this BECE will vary from state to state in Nigeria depending on its human and material resources, level of educational development, and the general state of its schools. However, it is expedient to examine the consistency of the scores obtained from the test in order to determine its dependability.

A test (also known as an exam) is a measurement of a person's knowledge, skills, and aptitude. According to Kissi, Baidoo, Anane & Annan (2023), it is a structured situation in which an individual is expected to respond to a series of items, each of which has a preferred answer. Jimoh, I. M., Akinboboye & Sheu, (2021) Further noted that a test is a single-occurrence, unidimensional, timed activity that often uses free-form or structured response items. The size of the construct one holds can be determined both qualitatively and quantitatively using this method. One of the tools used to gauge and assess pupils' academic success is this one. The BECE Test is divided into two main sections. Its first section comprised of objective items. The objective test items include a list of possible options as well as a list of questions. Students must select the most appropriate option that best represents the answer as it relates to the question. The essay is the second section, and it contains items to which students are allowed to respond using own expressions. Essay items allowed for a wide range of responses, with varied degrees of quality or correctness. As stated in the National Policy on Education document (FRN, 2013). Twenty-five subjects are administered at BECE level, and a candidate is expected to register and sit for at least ten and up to thirteen of them. If a candidate passes at least six subjects, including English and Mathematics, he or she is regarded to have passed and due to obtain the Junior Secondary School Certificate (JSSC). Despite possible variations in test difficulty or examinee demographics, equating the 2018 and 2019 BECE English exams can help guarantee that scores are comparable across years. In the end, this can help students and stakeholders by improving the certification process's validity and impartiality.

der Linden, (2022) defines test score equating as the technique of statistically comparing test results. It is a process of comparing the test scores of multiple tests given to examinees or groups of examinees. Test equating is a measurement procedure that includes test formulation, administration, analysis, scoring, reporting, and assessment. It can be done using either Classical Test Theory (CTT) or Item Response Theory (IRT). The researcher has a variety of options when it comes to equating test scores. Mean Equating, linear equating, Levine equally reliable linear equating, and Levine unequally reliable linear equating are some of these methods. Tucker linear equating, chained linear equating, equipercentile equating (frequency estimation, equipercentile equating, chained equipercentile equating), one-parameter logistic (Rasch) model equating (Concurrent calibration, Fixed-based procedure, equating constant procedure, major axis procedure), two-parameter logistic model equating (Concurrent calibration, Fixed-based procedure, equating constant (separate and concurrent calibration)).

The linear and separate calibration Equating Test procedures are the core of this study. Linear equating considers that, aside from differences in means and standard deviations, the distributions of scores on two different types of tests are the same; this leads to difficulty disparities varying over the scale of the two assessments. Based on this assumption, the z-scores of the two versions can be matched. The mean and standard deviation of the two scores (X and Y) are used to define the linear conversion. The slope of the linear conversion is A, and the intercept is B in this case. It enables the relative difficulty of two or more types of tests to be varied along the score scale. When the standard deviations are equal, linear equating becomes equivalent, according to Kolen and Brennan's (2014). If the only differences between the score distributions are the mean and standard deviation, linear equating is appropriate. If there are variations in the distributions beyond these two moments, or if the shapes of the distributions differ, linear equating may be impossible. When presuming that score distribution will be generally consistent over multiple assessment populations, caution is required. Since the equating procedure would differ from other datasets, a significant score distribution variation can invalidate a linear equating (Eiji, Catherine & Yong Won, 2000). Moment methods (mean-mean, mean-sigma, and strong mean-sigma) and characteristic curve methods are two types of calibration methods that can be studied separately (Stocking, Lord & Haebara). The moments approaches, which utilize more data from item parameters, are hence expected to yield better scalability results (Baker & Al-Karni, 1991; Hanson & Beguin, 2012). Different criteria and challenges in comparing scores may result from the inconsistent administration and scoring of the BECE among Nigerian states. In order to guarantee fairness and comparability of scores, the study intends to close this gap by investigating test score equating techniques, particularly linear and separate calibration equating test procedures.

Statement of the Problem

The BECE is crucial to educational development and it is expected that such test be comparably standard. Hence, its psychometric properties must be determined. Previous researches revealed that the BECE tests are being developed without consideration to uniformity in terms of tests psychometric properties, time of test administration and scoring. Hence, the comparability of students' scores obtained from test such as BECE across multiple formats and occasions has been in doubt by various researchers and stakeholders. This appears to be caused by differences in the quality of assessment instruments and the crises of getting a pool of items, which has further allowed test takers to be at an advantage or disadvantage when taking the test. Furthermore, the validity of comparing scores from different years is called into question by the absence of control for cohort ability variations. This absence may result in unjust outcomes, such as incorrect assessments of educational programs or incorrect student classifications. However, Data available at the examination management department, Ondo state ministry of education, revealed that a total of 55,766 and 67,675 sat for the 2018 and 2019 English Language test in the Basic Education Certificate Examination conducted by the Ministry of Education in public and private junior secondary schools across 18 local government areas in Ondo State. Furthermore, it was revealed that a total of 55,292 and 64,366 passed, while a total of 474 (0.9%) and 3,309 (4.9%) failed the BECE English Language in 2018 and 2019. As a result, it is challenging for test administrators to guarantee fair comparability and the consistency of scores across multiple formats and occasions. Moreover, some conceptual solutions have been proposed in relation to the issues of standard comparability in testing; however, they do not appear to be adequate. This is most likely because they lack a robust statistical foundation. As a result, it is necessary to investigate how scores from multiple evaluation standards are psychometrically equivalent and comparable using test score equating methods.

Purpose of the Study

The main objective of the study is to determine the efficient techniques for the comparison of student's performance in English Language Basic Education Certificate Examination scores.

The specific objectives of this study are to:

- a. Compare the equating function used in placing scores of 2018 and 2019 English Language BECE tests using Linear and Separate Calibration Methods; and
- b. Determine the difference of the equated scores of 2018 and 2019 English Language BECE obtained using Linear and Separate Calibration Methods.

Research Questions

From the objectives of the study, the following research questions were raised.

1. How comparable are the equating functions used in placing the scores of 2018 and 2019 English BECE tests using Linear and Separate Calibration Method?
2. Will there be a significant difference in the mean equated scores of 2018 and 2019 English language BECE tests using Linear and Separate Calibration Methods?

Methodology

The study adopted an ex-post facto design. The design was adopted for the study because it gives a descriptive explanation after an event has occurred, with no manipulation, because the data were taken from the responses of students who sat for the 2018 and 2019 Ondo state Basic Education Certificate Examination; the researcher did not have complete control over the data. The population for the study comprised the 123,441 junior secondary year three (JSS3) students who sat for the English Language test in the 2018 and 2019 Basic Education Certificate Examination conducted by the Ondo State Ministry of Education in public and private junior secondary schools, which includes 18 local government areas. There were 55,766 candidates who took the BECE English Language exam in public and private schools in 2018. Whereas 67,675 took the test in 2019. The study sample consisted of 4,320 students selected from the population using a multistage sampling technique. From each of the three senatorial districts of Ondo State, three local government areas (LGAs) were selected using a simple random sampling technique to make a total of 9 local governments. Six schools (three public and three private schools) were selected using a stratified random sampling technique to make a total of 54 schools. From the 54 Junior Secondary Schools selected, an intact class of 50 public and 30 private school students who sat for the Ondo State Basic Education Certificate Examination in all selected public and private schools to make a total of 2,160 for each of 2018 and 2019 were selected as sample for this study. Secondary data was used for the study. This consisted of students' responses to multiple-choice English Language items of the Basic Education Certificate Examination (BECE) for 2018 and 2019, which consisted of 60 multiple-choice items. The Optical Marks Record (OMR) sheet, which contains students' response records, was collected from the Department of Curriculum Development and Evaluation (CD&E) of Ondo State Ministry of Education, Science and Technology. The data were extracted from a soft copy of the BECE OMR sheets; the extraction contains all details of BECE candidates, including their responses and scores. The Cronbach's Alpha coefficient for the instrument was found to be 0.85, indicating high levels of consistency and dependability in the data. The data collected were coded and analyzed using BILOG-MG and Linear Equating framework implemented through the R language package environment for statistical packages.

Results

Research question One: How comparable are the equating functions used in placing the scores of 2018 and 2019 English Language BECE tests using Linear and Separate Calibration Method? To answer this research question, the equating function of the linear and separate calibration method were compared for English language test. The results are presented as follows.

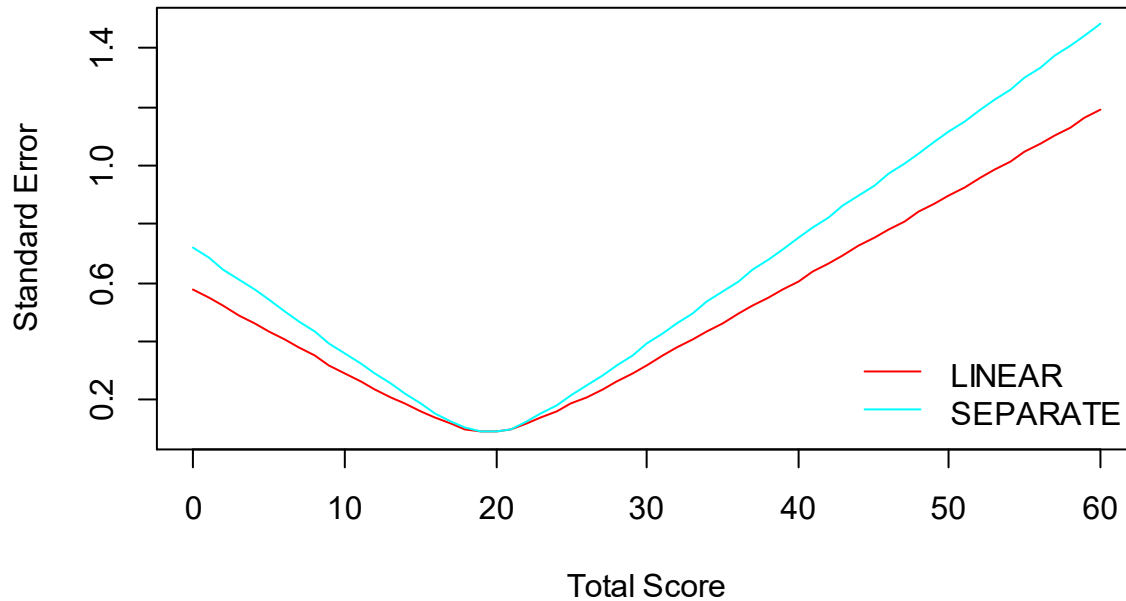


Figure 1: Comparability of the equating functions of the linear and separate calibration method of equating for 2018 and 2019 OSBECE English language test

Figure 1 shows the comparability of the equating functions of the linear and separate calibration methods of test equating. The Figure shows that the Standard Error of Equating for linear equating is 0.6. Also, the SEE for separate calibration method is 0.7. This showed that the equating functions for the linear method was better than the separate calibration method. The result implies that the linear method of equating was a better method for equating 2018 and 2019 OSBECE English language tests.

Research question Two: Will there be a significant difference in the mean equated scores of 2018 and 2019 English language BECE tests using Linear and Separate Calibration Methods?

To answer this research question, the significant difference in the mean equated scores of the 2018 and 2019 English language tests, using linear equating and separate calibration methods, was examined using an independent t-test. The result for the linear method of equating was used because the method was the best as found in the study. The results are presented as follows:

Table 1: Independent t-test of equated scores of 2018 and 2019 English language tests.

	N	Mean	Std. Deviation	T	df	Sig. (2-tailed)
2018_EQUATED	2160	41.2739	4.41519	3.552	4318	0.000
ENG 2019	2160	40.8792	4.95735			

The table shows that 2019 OSBECE test was more difficult than the 2018 version by about 0.4 units. The independent t-test showed the observed difference was statistically significant (t (df =

4318) = 3.552, $p < 0.05$). The result showed that 2018 and 2019 OSBECE English language test were not equivalent.

Discussion of Findings

The result of the First objective of the study which compare the equating function used in placing the scores of 2018 and 2019 English Language BECE tests using Linear and Separate Calibration Method revealed that equating functions for the linear method was better than the separate calibration method. This is because the standard error of equating for the linear method was lower than the separate calibration method across the obtainable scores. The finding conforms to the opinion of Kolen and Brennan (2014) that the most effective method is the equating method that produces the smallest standard error of equating. In a related study, Atsua, Uzoeshi, Oludi, and Wagbara (2018) compare the Equating functions of 2015 and 2016 basic education certificate examination on civic education using classical test theory and item response theory in Oyo state. The results indicated that the equating functions used in the conversion of the 2015 Civic Education BECE test scores to the scale of 2016 Civic Education BECE test under the CTT framework and IRT framework were not the same. However, CTT and IRT approaches produced similar results. It was concluded that both theories are good approaches for test equating. Both CTT and IRT approaches were recommended for test equating.

Finally, to determine the comparability of the equated scores of 2018 and 2019 English Language ODBECE using Linear and Separate Calibration Methods. The result for the linear method of equating was used because of the fact that the method was the best as found in the study. Hence, the result showed that the linear equating method was more efficient than the separate calibration method in the estimation of students' scores. In a similar study, Olanigan, Adediwura and Ogunsanmi (2022) compare linear and separate calibration methods of equating continuous assessment scores of public and private elementary schools. The study concluded that linear equating was more efficient. The results also revealed that 2019 ODBECE English language test was more difficult than the 2018 version. The result further showed that 2018 and 2019 ODBECE English language test were not equivalent. This study is in line with the study carried out by Babatimilehin (2021) while investigating the equivalence of WAEC and NECO SSCE chemistry items using Linear and equating approach of CCT and IRT, the study concluded that there is no statistical equivalence between WAEC and NECO Chemistry examination.

Given that the linear equating has a lower standard error of equating than the separate calibration approach, the study's results support the use of linear equating for connecting the 2018 and 2019 ODBECE English language examinations. This supports the idea put out by Kolen and Brennan (2014) that the best equating technique reduces the equating standard error. Moreover, issues regarding the premise of parallel forms are raised by the study's conclusion that the 2018 and 2019 examinations were not equal. Differences in test content, difficulty, or administration circumstances could be the cause of this non-equivalency.

Conclusion

The results of the study indicate that a better option for connecting the 2018 and 2019 OSBECE English language examinations is the linear method of equating. This suggests that the linear approach provided a more accurate comparison of scores across years by successfully accounting for variations in test form complexity. The 2018 and 2019 examinations' non-equivalency emphasizes the necessity of equating to guarantee score comparability and fairness. The performance of the linear approach justifies its application for equating upcoming OSBECE English language exams, which may influence examination procedures and policies. These results have consequences for how OSBECE English language scores are interpreted and applied,

especially when making important choices. Stakeholders can be confident in the comparability of scores by employing the linear equating method, promoting impartial and knowledgeable decisions.

Recommendations

Based on the findings of this study, the recommendations are given subsequently by the researcher.

- i. Nigerian examination organizations should use the most suitable test theory model to verify the statistical equivalence of their examination test items before administration.
- ii. In Nigeria, public examinations should be constructed and examined using test equating procedures and designs. By doing this, test results from this examination body can be compared between years and the geopolitical regions in which they are administered.
- iii Test scores equating should be used to standardize ODBECE students' scores. The linear equating method should be used for equating the BECE results of public and private secondary schools in Ondo state.
- iv. The various forms of test should be equated when the item parameters of the test are different

References

- Atsua, T. G., Uzoeshi, I. V., Oludi, p., & Wagbara, E. S. (2018). Equating 2015 and 2016 Basic Education Certificate Examination on Civic Education using Classical Test Theory and Item Response Theory in Oyo State, Nigeria. *Journal of Pristine*, 14(1).
- Babatimilehin (2021), *Equivalence of WAEC and NECO SSCE Chemistry Items using Linear and Equating Approach of CCT and IRT* (Unpublished M.A.Ed. Thesis). Department of Educational Foundations and Counselling, Obafemi Awolowo University, Ile-Ife, Nigeria.
- Baker, F. B., & Al-Karni, A. (1991). A Comparison of Two Procedures for Computing IRT Equating Coefficients. *Journal of Educational Measurement*, 28: 147-162. <https://doi.org/10.1111/j.1745-3984.1991.tb00350.x>
- Der Linden, W. J. V. (2022). What Is Actually Equated in “Test Equating”? A Didactic Note. *Journal of Educational and Behavioral Statistics*, 47(3), 353362.
- Eiji, M., Catherine, M. H. & Yong-Won, L. (2008). Equating and Linking of Performance Assessment. *Applied Psychological Measurement*. 24(4): 325-337. Available at: <http://upm.sagepub.com/cgi/content/abstract/24/4/325>
- Faleye, B. A., & Afolabi, E. R. I. (2005). The Predictive validity of Osun State Junior Secondary Certificate Examination. *Electronic Journal of Research in Educational Psychology*. No 5 -3 (1), pp. 131-144.
- Hanson, B. A. & Beguin, A. A., (2002). Obtaining a Common Scale for IRT Item Parameters using Separate Versus Concurrent Estimation in Common Item Nonequivalent Groups Equating Design. *Applied Psychological Measurement*. 26(1): 3-24. <https://doi.org/10.1177/0146621602026001001>
- Jimoh, I. M., Akinboboye, J. T., & Sheu, A. L. (2021). Investigating unidimensionality of Mathematics Multiple-Choice Test Items used for Joint Mock Examination in Kwara State Public Schools. *SPEKTA (Jurnal Pengabdian Kepada Masyarakat: Teknologi dan Aplikasi)*, 2(2), 103-116.
- Kissi, P., Baidoo-Anu, D., Anane, E., & Annan-Brew, R. K. (2023). Teachers’ Test Construction Competencies in Examination-oriented Educational System: Exploring teachers’ multiple-choice test construction competence. In *Frontiers in Education* (Vol. 8, p. 1154592). Frontiers. <https://doi.org/10.3389/educ.2023.1154592>
- Kolen, M. J. & Brennan, R. L. (2014). *Test Equating, Scaling and Linking* (Third Edition). USA: Springer.
- Olanigan, N. A., Adediwura, A. A., & Ogunsanmi, O. A. (2022). Linear and Separate Calibration Methods of Equating Continuous Assessment Scores of Public and Private Elementary Schools. *Journal of Integrated Elementary Education*, 2(2), 117-129.