



Comparing Dimensionality and Local Independence Parameters of Lagos State Unified Mock and West African Senior School Certificate Examinations in Biology

Olufemi Abiodun Ajayi, Risikat Iyabode Salam & Olawale Yussuf Saka

Department of Counselling Psychology & Educational Foundations,

Tai Solarin University of Education, Ijagun, Nigeria

Corresponding author: ajayioa@tasued.edu.ng

Abstract

This study compared the Lagos State Unified Mock and West African Senior School Certificate Examinations in Biology in terms of dimensionality and local independence. Descriptive research design and multistage sampling procedure were adopted. The six (6) Lagos Education Districts were enumerated, with two local government areas (LGA) selected from each district and five (5) schools selected from each LGA. The instruments adopted were the 2021 LSUME and WASSCE multiple-choice biology. The sample comprised 2,400 SSS 3 students. Data were subjected to Stout's Test of Essential Unidimensionality and Yen Q3 statistics. The results showed that LSUME and WASSCE Biology fulfilled unidimensionality ($DETECT < 0.20$, $ASSI < 0.25$, $RATIO < 0.36$) and local independence assumptions (residual correlation < 0.20). The authors concluded that LSUME and WASSCE Biology had comparable item parameter estimates. It is recommended that examination bodies and teachers should devote sufficient efforts to item development and adopt item response theory in students' cognitive assessment.

Keywords: Item local independence, Item response theory, Uni-dimensionality, Biology

Introduction

Biology stems from the Greek words "bios" (life) and "logos" (study). It is the study of organisms and their functioning, including how they interact with one another and their surroundings. It is one of the prerequisites for pursuing any science-related subject at a higher level of education after secondary school.

The Lagos State Unified Mock Examination is written by all SSS III students in all public schools in the state to determine their readiness for the external examination. The items are usually developed by the State Ministry of Education and accompanied with a marking guide to ensure consistent scoring. According to Asuru (2010), a unified examination is an evaluation of a student's performance utilising question cues to identify the quantity of information, abilities, mindsets, or behaviours a student has acquired, the extent to which he may employ them, or their quality and efficacy. Furthermore, Kwasi (2009) sees unified examinations as a method for accountability and raising educational standards, especially if they mirror what

students are taught in the classroom. The West African Senior School Certificate Examination is conducted, and students who pass the exam receive a certificate indicating that they have finished their secondary education. It is used to determine who gets admitted to colleges and universities for future education.

A test is a tool that provides details of a person's ability, expertise, performance, and accomplishment. One of the kinds of tests used to evaluate students' level of learning is the objective test. The objective test is one of the assessment forms used to gauge how well pupils are meeting the behavioural goals. Onuka & Ogbebor (2016) describe a multiple-choice test as a question having three or more answers, one of which is correct; the greater the number of possibilities, the more trustworthy the answer. It is significant because of its potential for assessing an equitable proportion of the universe of relevant content without necessarily increasing testing duration. It is also used to supplement the essay test since it is objective in grading examinee

responds (Onuka & Ogbebor, 2016). When a test is properly developed bearing in mind the test construction guidelines and psychometric features, it offers the teachers substantial insight or knowledge about the learner, enabling them to accurately and reliably measure students' mastery of specific course material (Babatunde & Benson, 2020).

Item Response Theory (IRT) is a measurement framework employed in educational and psychological assessments to investigate individual responses to test items. It provides perceptions on the relationships between respondents' characteristics and test performance. It's also referred to as latent trait theory. IRT attempts to predict a test taker's aptitude and opportunity of accurately answering a question by considering the pattern of test item responses (Ojerinde, 2013). It provides perceptions on the relationships between respondents' characteristics and test performance. According to Ojerinde et al. (2012), the basis of IRT includes measurement assumptions regarding examinee, item, and test performance, as well as how this performance connects to knowledge assessed by particular test items. IRT may determine an item parameter regardless of the test-taker characteristics or the traits of all other items on the test. IRT offers four models: 1-parameter logistic model (1PL), 2-parameter logistic model (2PL), 3-parameter logistic model (3PL) and 4-parameter logistic model (4PL). IRT presupposes that the test items' latent traits are unidimensional or may be represented by a single scale.

Unidimensionality is one of the axioms of item response theory. The concept of unidimensionality states that each item only assesses one underlying characteristic or aptitude. According to Kyung (2013), unidimensionality indicates that a group of items or a test examines just one latent ability. If the unidimensionality criterion is broken, an item may assess a variety of underlying traits or skills, resulting in erroneous estimates of the levels of each attribute. If violated, it reduces the test's validity, reliability, and fairness. It guarantees that the items assess the desired attribute correctly and consistently. According to Ojerinde and Ifewulu (2012), the Cronbach analysis test, factor loading test, vector frequency test, and confirmatory factor analysis (CFA), structural equation modelling (SEM) test, and the Eigenvalue test are all methods for testing for unidimensionality.

Local item independence is the second assumption of item response theory, which stipulates that the students' response to an item should not in any

way be affected by their responses to another item in the test. This indicates that after the effect of the students' underlying trait is partially taken into account, there should be no correlation between any two items. The test's latent trait should be the only way to correlate the items. When upheld, it produces small errors of estimate. Aryadoust et al. (2021) stated that if a student answers an item correctly or incorrectly, it will affect the accuracy of statistical analysis of their performance. Numerous techniques for assessing local item independence include the Wald test, the power-divergence (PD) statistic, the Q3 statistic, Fisher's r-to-z transformed Q3 (Yen, 1993), the likelihood ratio test in logistic regression (LR G2), the absolute value of mutual information difference (AMID; Tsai & Hsu, 2005), the mutual information difference (MID), and the application of the residual correlation from factor analysis (FA) (Metibemi, 2017). Among all these techniques, the most prevalent approach for determining local dependency is the Q3 coefficient, which entails examining correlations between standardised or raw score residuals (Aryadoust et al., 2021).

Statement of the Problem

In practice, concurrent tests are expected to yield comparable results; however, this is not the case between the Lagos State Unified Mock Examination (LSUME) and the West African Senior School Certificate Examination (WASSCE). There is no justification for the variation observed because the same group of students sits for the two examinations, and the domains of learning and behaviours sampled also remain the same. This has raised concern among the stakeholders in education. Thus, this study investigates the concurrent validity of the LSUME and WASSCE biology.

Research Questions

1. To what extent do the LSUME and WASSCE in Biology satisfy the unidimensionality assumption of item response theory?
2. To what extent do the LSUME and WASSCE in Biology satisfy the local item independence assumption of item response theory?
3. How Comparable are the unidimensionality and local item independence assumptions between the LSUME and WASSCE in Biology?

Methodology

The descriptive research design of survey type was adopted for this investigation. The study's population includes all 46,356 examinees who registered for and sat the Lagos State Unified Mock and West African Senior School Certificate tests. Multistage sampling procedure was adopted to cater for adequate representation of the study's population. At first, the six (6) education districts were enumerated for the study. Simple random sampling was adopted to select two (2) Local Government Areas (LGAs) from each District. Secondly, in each LGAs, five (5) public senior secondary schools. Lastly, in each school, 40 students were selected using the ballot method. In all, a total of Twelve (12) LGAs, sixty (60) schools and 2400 students constituted the study sample. The research instruments for the study are the 2021 LSUME and the WASSCE multiple-choice questions in Biology.

Data Analyses Techniques

Stout's test of essential Unidimensionality and Assessment Criteria

A dimensionality evaluation was conducted on the Biology Achievement Test (BAT) from the LSUME and the WASSCE using a supplementary item response theory (IRT) package called "sirt." This package, developed by Robitzsch and Robitzsch (2022) and implemented in the R programming software version 4.0.2 (R Core Team, 2016), was employed to assess the dimensionality of the tests. The BAT comprised 50 items, and the responses of 2400 test-

takers were analysed. To evaluate the dimensionality, Stout's test of essential unidimensionality was employed. This test is well-established and has been used by Ayanwale (2023) and Ayanwale et al. (2020), building upon Hattie et al.'s (1996) work. The "conf detect" function from the *sirt* package was utilised for this purpose. The criteria for assessing the dimensionality of the test data were drawn from the research of Jang and Roussos (2007) as well as Zhang (2007). The following guidelines were employed:

Strong multidimensionality: DETECT > 1.00

Moderate multidimensionality: $0.40 < \text{DETECT} < 1.00$

Weak multidimensionality: $0.20 < \text{DETECT} < 0.40$

Essential unidimensionality: DETECT < 0.20

Maximum value under simple structure: ASSI=1, RATIO=1

Essential deviation from unidimensionality: ASSI > 0.25, RATIO > 0.36

Yen Q3 statistics for Item Local independence and Assessment Criteria

Item local independence was conducted on the test items for both the LSUME and WASSCE, using Yen's Q3 statistic (Yen, 1984). After taking into account the influence of individual person location estimates, the Q3 statistic shows the correlation between residuals for a pair of objects. As suggested by Yen (1993), a cutoff value of 0.20 is recommended. Values beyond this limit suggest that the local independence assumption may not be true, calling attention to this as showing local dependence.

Results Presentation

The responses from the LSUME and WASSCE tests were both subjected to Stout's test of essential unidimensionality and the results are detailed in Table 1 below:

Table 1: Dimensionality assessment of LSUME and WASSCE

Indices	LSUME		WASSCE	
	unweighted	weighted	unweighted	weighted
DETECT	-0.369	-0.369	-0.384	-0.384
ASSI	-0.365	-0.365	-0.228	0.228
RATIO	-0.485	-0.485	-0.354	-0.354

The results of Table 1 above reveals that the Biology Achievement Test (BAT) demonstrated a fundamental unidimensional nature for both the LSUME and WASSCE assessments. In the LSUME context, the highest DETECT value was -0.369, comfortably falling below the critical threshold of 0.20. The ASSI value stood at -0.365, well within the limit of 0.25, while the RATIO value was -0.485, below the 0.36 benchmark. Similarly, in the WASSCE scenario, the maximum DETECT value reached was -0.384 ($< .20$), the ASSI

value was -0.228 (< 0.25), and the RATIO value was -0.354 (< 0.36). Consequently, these findings firmly supported the absence of evidence against the assumption of unidimensionality for both tests (that is, the assumption was not rejected.) This finding distinctly indicates that the variability observed in how test-takers responded to the BAT could be attributed primarily to a single dominant dimension. Thus, the BAT aligns with the foundational premise of unidimensionality as stipulated by the principles of item response theory.

This research question was answered using Yen's Q3 statistics and the results were presented in Tables 2 and 3 respectively.

Table 2: Summary of Q3 statistics for LSUME

Mean	SD	Min.	Max.	10%	25%	50%	75%	90%
-0.020	0.057	-0.389	-0.226	-0.087	-0.055	-0.023	0.012	0.046

Table 2 shows the findings of the estimation of Yen's Q3 statistic for the LSUME test items which is used to evaluate item local independence based on estimated theta score. With a symmetric matrix of 50 by 50 items, the mean Q3 statistic value for all 1200 item pairs is around -0.020, and the standard deviation is 0.057. This number is very close to zero, indicating that after accounting for the estimated theta scores, there is typically very little correlation between the residuals of paired items. This is in line with the expectation for local independence, according to which after the impact of the latent trait (theta) is taken into account, there should be no significant correlation between the items.

The minimum value Q3 value indicate lowest correlation between residuals within any item pair and maximum Q3 value also falls below 0.2. Even at different percentiles, the distribution of Q3 values indicates that most values are negative and relatively modest. This visualisation corroborates earlier findings, as the majority of Q3 values in the plot are clustered closely around zero. Succinctly, the Q3 values are predominantly negative and close to zero, indicating that the correlation between residuals for most item pairs is minimal after accounting for the estimated theta scores. This upholds the local independence assumption of item response theory.

Table 3: Summary of Q3 statistics for WASSCE

Mean	SD	Min.	Max.	10%	25%	50%	75%	90%
-0.027	0.072	-0.296	0.142	-0.107	-0.069	-0.022	0.029	0.069

Table 3 shows information regarding the estimation of Yen's Q3 statistic for WASSCE test items. It provides details on how the Q3 statistic for the WASSCE exam items is estimated. With a standard deviation of 0.072, the average Q3 value across the 1200 pairings of items is roughly -0.027. This closeness to zero suggests that, on average, local independence on the test items is attained. The smallest Q3 value observed is -0.296, which represents the

weakest residual correlation among any item pair, but this value remains relatively close to zero, implying that the principle of local independence is generally upheld. On the other hand, the highest Q3 value observed is -0.461, which represents the strongest residual correlation among any item pair. The negative value at different percentiles also shows minimally no correlation between the items. This shows that both

instruments adhere to the local independence assumption of IRT.

The results of the Stout's test of essential unidimensionality employed shows that the Lagos State Unified Mock examination in Biology shows that the highest DETECT value was -0.369, comfortably falling below the critical threshold of 0.20. The ASSI value stood at -0.365, well within the limit of 0.25, while the RATIO value was -0.485, below the 0.36 benchmark. Similarly, in the WASSCE scenario, the maximum DETECT value reached -0.384 ($< .20$), the ASSI value was -0.228 (< 0.25), and the RATIO value was -0.354 (< 0.36). This shows that the Lagos State Unified Mock and West African Senior School Certificate examinations obey the unidimensionality assumption of IRT. This means that the variation in students' responses to the Lagos State Unified Mock and the West African Senior School Certificate examinations could be predominantly related to one dominant factor.

Also, the estimation of Yen's Q3 statistic for the LSUME test items shows that the mean Q3 statistic value across all 1200 item pairs (symmetric matrix of 50 by 50) is roughly -0.020, with a standard deviation of 0.057 while the average Q3 value for the West African Senior School Certificate examinations (Biology) across the 1200 pairs of items is approximately -0.027 with a standard deviation of 0.072. An examination of percentiles (10%, 25%, 50%, 75%, 90%) in the distribution of Q3 values in the Lagos State Unified Mock and West African Senior School Certificate examinations indicates that most values are negative and relatively modest. Although both instruments showed positive value at 75% and 90% but the value is still below the recommended value of 0.2. This implies that both instruments obey the local independence assumption of IRT.

Discussion of Findings

The result indicated that LSUME and WASSCE Biology are essentially unidimensional and also obey the item local independence assumption. This is because both had DETECT values less than the critical value of 0.20, ASSI values within the limit of 0.25 and RATIO values below the benchmark of 0.36. The implication of the study is that one dominant ability accounted for the variances in performance of the examinees who took the LSUME and WASSCE Biology Multiple Choice Test.

This result is consisted with the findings of Ojerinde and Ifewulu (2012) who report that the UTME Mathematics complies with the assumption of unidimensionality. Also, Orim (2015) looks into the unidimensionality of WAEC and NECO Biology items and found out that unidimensionality exists in 47 out of 50 WAEC items and 39 out of 50 NECO items. In the same vein, Oribhabor (2019) findings shows that some of the items showed DIF and the items are unidimensional. Another study by Adediwura and Kasali (2022), establish that item response theory's unidimensionality and local independence assumptions are met by NECO 2016 Mathematics' multiple-choice questions.

The result of research question two indicates that Lagos State Unified and West African Senior School Certificate examinations obey the item local independence assumption. This is because the mean Q3 statistic value for all the item pairs is around -0.020 which falls below a recommended cutoff of 0.20. The finding is in agreement with the work of Jimoh and Abdulsalam-Nuhu (2022) who reported that Unified Tertiary Matriculation Examination Mathematics items for 2000, 2001, and 2003 years were locally independent.

In contrast, Nasela and Adewale (2020) in their study, reported that some of the items of Basic Education Certificate Examinations (BECE) and National Examinations Council (NECO) violates the local independence assumption. Also, Alaba and Kasali (2022) in their study on dimensionality and local independence of IRT, reported that some of the examine items violated item local dependence.

Conclusion

This study's findings lead to the conclusion that the biology multiple choice items that make up the Lagos State Unified and West African Senior School Certificate Examinations adhere to Item Response Theory's unidimensional and item local independence assumptions.

Recommendations

1. Sufficient efforts and time should be invested in item development by item writers.
2. Classroom teachers and examination bodies should adopt Item Response Theory in cognitive assessment of students

References

- Adediwura, A. A. & Kasali, J. (2022). Conformity of National Examinations Council mathematics multiple choice test items to the assumption of Item Response Theory. *Al-Hikmah Journal of Arts & Social Sciences Education*, 4 (1), 1-4.
- Asuru, V. A. (2010). Beyond cognitive assessment: Mechanism for reducing system generated causes of examination malpractices. *Journal of the Association for Educational Assessment in Africa*, 4(1), 105 - 110.
- Aryadoust, V., Ng, L. Y. & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: recommendations and guidelines for research. *Lang. Test.* 38, 6–40. doi: 10.1177/0265532220927487
- Babatunde K, O. & Benson A. A. (2020), Using test theories models to assess senior secondary students ability in constructed-response Mathematics tests. *Journal of Education and Practice.* 11 (7)
- Jimoh, M. I., & Abdulsalam-Nuhu, R. (2022). Investigating local item independence in Civic Education multiple-choice items of Joint Mock Examination. *Journal Plus Education*, 31(2), 114-128.
- Kasali, J., & Adeyemi, A. (2022). Estimation of item parameter indices of NECO Mathematics multiple choice test items among Nigerian students. *Journal of Integrated Elementary Education*, 2(1), 43-54. doi:<https://doi.org/10.21580/jieed.v2i1.10187>
- Kwasi, A. S. (2009). An evaluative study of a distance teacher education programme in a university in Ghana. *Journal of International Review of Research in Open and Distance Learning*, 10(4), 45 – 56.
- Metibemu, M. A. (2017). Appraisal of unidimensionality and test component local independence. *189 workshop paper presented at the 3rd International Conference of Educational and Research Network in Africa, Makurdi, Nigeria.* 3-8th December 2017.
- Ojerinde, D. (2013) Classical Test Theory VS Item Response Theory: An evaluation of the comparability of item analysis results. Retrieved from: [http://ui.edu.ng/sites/default/files/PROF%20OJERINDE%27S%20LECTURE%20\(Autosaved\).pdf](http://ui.edu.ng/sites/default/files/PROF%20OJERINDE%27S%20LECTURE%20(Autosaved).pdf)
- Ojerinde, D., & Ifewulu, B. C. (2012). Item unidimensionality using 2010 unified tertiary matriculation examination mathematics pre-test [paper presentation]. *2012 international conference of iaea, Kazakhstan*
- Onuka, A.O.U. & Ogbebor, U.C (2016). An Introduction to assessment. public examining in Sub-Sahara Africa: issues, challenges and prospects. Vol 1. 145-245. A book in honour of Professor Dibu Ojerinde published by JAMB, Nigeria
- Orim, S. N. (2015). Unidimensionality and classical test parameters of 2007 WAEC and NECO biology tests in Cross River State, Nigeria. *Unpublished Dissertation.* University of Calabar, Calabar, Nigeria.
- Pila, N. & Adewale, J. G. (2020). Assessment of dimensionality and local independence of state and national basic education Certificate Examinations in National values Education in Benue State. *The African Journal of Behavioural and Scale Development Research*, 2(1). <https://10.58579/ajbsdr.v2i1.43>

Declaration

I, Prof. Olufemi Abiodun Ajayi, do hereby declare that my contribution to the study titled “comparing dimensionality and local item independence parameters of Lagos State Unified Mock and West African Senior School Certificate Examinations in Biology” was the writing of the statement of the problem, purpose of the study, research questions, the methodology and collection of data.

Declaration

I, Risikat Iyabode Salam, hereby declare that my contribution to the study titled "Comparing Dimensionality and Local Item Independence Parameters of Lagos State Unified Mock Examination and West African Senior School Certificate Examinations in Biology" was in the areas of collection of data, writing background to the study, discussion of findings, conclusion and recommendations.

Declaration

I, Olawale Yussuf Saka, hereby declare that my contribution to the study titled "Comparing Dimensionality and Local Item Independence Parameters of Lagos State Unified Mock and West African Senior School Certificate Examinations in Biology" was in the areas of data collection, data analysis, results and interpretation of findings.