E-ISSN: 1115 - 8425 P-ISSN: 3027-012X
MPRI-JAEE | INTERNATIONAL JOURNAL OF ASSESSMENT & EVALUATION IN EDU.
VOL. 09 NO. 8 – JULY. 2025

TESTING THE ASSUMPTIONS OF DIMENSIONALITY, LOCAL INDEPENDENCE AND MONOTONICITY OF CHRISTIAN RELIGIOUS STUDIES MULTIPLE CHOICE QUESTIONS

AYEK BASSEY BASSEY; & BASIL C. E. OGUGUO

Department of Science Education, University of Nigeria, Nsukka

Corresponding Author: basil.oguguo@unn.edu.ng

DOI: https://doi.org/10.70382/mejaee.v9i8.060

Abstract

This study tested the assumptions of Item Response Theorey (IRT) (dimensionality, local independence and monotonicity) of WAEC Christian religious studies multiple choice questions. Instrumentation Research Design was adopted for the study involving 1500 participants. The participants were drawn through multistage sampling procedure. Christian Religious Studies Multiple Choice Questions (CRSMCQ) was the instrument used for data collection. The internal consistency reliability coefficient of the instrument was determined to be 0.81 using Kudder Richardson formula ($K\text{-}K_{20}$). The data collected was analyzed using Normal Ogive Harmonic Analysis Robust Method (NOHARM) and Root Mean Square Reduction (RMSR). The findings of the study showed that the test items of West African Senior School Christian Religious Studies Examination have three dimensions. The test items have good guessing factor index as such the correct response to one does not provide clue to correct response to other items. The study shows that some of the items did not violate the Item Response Theory assumption of item local independence. Teachers and examination bodies should determine the number of dimensions underlying test to understand the number of traits a test is designed to measure.

Keywords:
Dimensionality, Item Response Theory, Local Independence, Monotonicity, Test Validation

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

Introduction

Test development process is a systematic procedure adopted by test developers with the aim of developing test items that are valid, reliable and as objective as possible to ensure fairness among examinees. Nworgu (2015) outlined fourteen steps in test development process thus; content analysis, review of instructional objectives, development of test blue print, item writing, face validation, item review, trial testing, item analysis and so on. Chester (2008) also identified five steps in test development process, namely; test conceptualization, test construction, test try-out, analysis, and revision. These processes are geared towards developing test items that have sound psychometric qualities. Ali (2006) explained that test can only achieve its objective when such test is reliable and valid. Reliability of a test deals with the consistency of result obtained from a test. A test is reliable if the measures obtained from repeatedly using the test do not vary, but the values are consistently accurate (Ali, 2006). Reliability of a test is the confidence that on administering an instrument that such a test yields the same set of scores consistently (Adejoh & Obinne 2013). The second quality of a test is the validity of the test.

Validity of a test is the degree of trustworthiness of the test. When a test measures exactly what it was purports to measure the test is described as a valid test (Soura, 2016). Any test that measures only what it was designed to measure is referred to as a valid test (Nworgu, 2015). De Ayala (2009) pointed out that validity of a test is when measurements obtained are manifestation of the dormant variables. There are several validity indices that establish the validity of a test, these validity indices are determined through a process known as item analysis. Item analysis according to Nwani (2021) is a statistical procedure of determining the psychometric qualities of test items. Lee (2019) maintained that item analysis is the act of analysing students' responses to examination questions with the intention of determining examination quality. Donald et al. (2010) sees item analysis as the statistical method adopted by test item developers to select items for inclusion in a test. To effectively determine the validity indices as well as the reliability of a test which are major components of psychometric quality of a test, two major theoretical frameworks are usually employed; the item Response Theory (IRT) framework and the Classical test Theory (CTT) framework. IRT was pioneered by Fredrick Lord in 1953. IRT is centred on unobserved traits of the examinee. It is a group of mathematical models that attempt to explain the relationship between latent traits and their manifestations. They establish a link between the properties of items on an instrument, individuals responding to these items and the underlying trait being measured.

This study adopted IRT framework for the item analysis because it is only IRT framework that make provision for the testing of dimensionality, item local independence and

monotonicity as one of its major assumptions. Dai-Trang (2013) pointed out that IRT is the standard and effective method of carrying out psychometric assessment of new established measures. The theory allows the correct model specification that fits the data to be used as well as the Item Characteristics Curve (ICC) which represents the model expectation. CTT framework have been used by WAEC and NECO to determine the psychometric qualities of public examinations (Okwilagwe & Ogunrinde, 2017), as such, does not determine the number of dimensions underlying a test, does not separate the examinees characteristics and the test characteristics, this study therefore, adopted IRT framework to test the dimensionality, local independence and monotonicity of West African Senior School Certificate Christian Religious Studies Examination Multiple Choice Items. Item Response Theory as a framework has three basic assumptions which must be considered and tested for item analysis, these assumptions are; Dimensionality assumption, Local independence assumption and monotonicity assumption.

Dimension of a test is the number of ability(s) explaining the performance of examinee in a test; it is the number of traits required by examinees to answer correctly the items in a test. Such that a test requiring only one ability to correctly respond to the item is termed unidimensional test and test that requires two or more distinctive abilities is called multidimensional test. The number of traits or attribute or psychological constructs underlying a test is the dimensionality of a test (De Ayala, 2009). Dimensionality, according to De Mars (2010), is the number of trait(s) explaining the performance of examinee in a test; it is a number of traits required by examinees to answer correctly the items in a test. Reckase (2009) opined that a test that requires only one trait is termed unidimensional test and the test that requires two or more distinctive traits are called multidimensional test. To calibrate a unidimensional test under IRT, the unidimensional IRT (UIRT) models have been developed and for multidimensional test, multidimensional IRT (MIRT) models have been developed as well. Hence, the choice of IRT model for test calibration depends on the number of dimensions that underlie the test.

The dimensionality assumption of IRT is the most basic of the IRT assumptions among others. This is because the correct assessment of the other assumptions and other applications of IRT depend largely on the number of trait(s) underlying a test. The implication of this is that if the number of trait(s) underlying a test is not correctly determined, the functional form (Item Characteristics Curve, - ICC), local independence of the test items and other applications of IRT such as differential item functioning and test equating will be incorrectly conceptualized. Assessing the dimensionality of test items is very important in interpreting scores obtained from that test. Test items are constructed to assess a particular construct or another, in a situation where the items in a test are not homogenous, the interpretation of that test based on one construct using the scores obtained from such a test will be grossly deficient. The interpretation of scores as

representing the construct hypothesized to be measured might be wrong. If an item assesses more than the desired construct, the total score will be influenced by the extra information as well. The score will not be interpreted as the standing of the person on the latent variable. Therefore, dimensionality of items comprising a test score is specifically for soundness of the assessment. Without testing dimensionality, interpretation of test score as representing one dimension is risky (Ziegler & Hagemann, 2015). If Christian Religious Studies Examinations must be appropriately reported based on the score obtained from the examination, there is need to test for the dimensionality of the examination.

Item local independence (ILD) assumption states that responses to an item are statistically independent of one another after conditioning them on their latent trait(s). Item responses are independent of one another given ability, once you know a person's ability level; the student responses to item(s) are independent of one another. This is one of the hallmark assumptions in IRT, and it makes many things possible (it will also be important for estimating examinee trait levels). Conditional independence provides us with statistically independent probabilities for item. The assumption of local independence implies that responses to an item are independent of responses to another item once controlling for the underlying variable measured by the scale. This concept is related to that of dimensionality assumption, if one trait determines success on each item, then examinee ability is the only thing that systematically affects item performance. Local independence means that if the trait level is held constant, there should be no association among the item responses. Local independence assumption holds a firm grip on the fact that items in a test should not be related to each other. Sharing a common passage, which is prevalent in reading comprehension tests and cloze tests, can be a potential source of local item dependence (LID). Lee, (2004) argued that LID results in biased parameter estimation and affects the unidimensionality of the test. Determining the item local independence of a test provides a justification that no item in the Christian Religious Studies Multiple Choice Test Items provides a clue to correct answer to another item. If items in the Christian Religious Studies Multiple choice test items provides clue to correct response to another item, the item parameters estimation will be affected, and if this is done, the quality of the entire test is affected and the test will not correctly express the various abilities of the students. It will therefore be difficult to differentiate or identify students with respect to their ability levels.

Monotonicity assumption is also known as model fit assumption. The fit between the model and the data can be assessed to check for model misspecifications. For instance, if 1PLmodel is used and the data follow a model with different slopes or a non-zero lower asymptote, many of the items will not fit 1PL model. If the function is not monotonically increasing, none of the common models will fit. Typically, IRT practitioners focus on the

fit of individual items, not the overall fit of the model across all items. For item fit, the concept of a residual, or the difference between an observed and model-predicted (expected) proportion is key. This residual is conditional on ability, meaning it is calculated for groups of examinees with approximately the same ability. The Item Characteristics Curve (ICC) represents the model expectation. The item characteristics curve provides graphical properties of an item. It provides information relation to the relationship between the ability of the examinees and the probability of providing correct response to an item.

ICC is a chart showing the properties of an item in terms of its difficulty index, discriminating index and guessing factor (for a 3-parameter-model). It gives comprehensive information about test item. It proffers solution to some questions like: Does the item discriminate well between the good and the poor testees? How do the low/high ability testees respond to the item?"

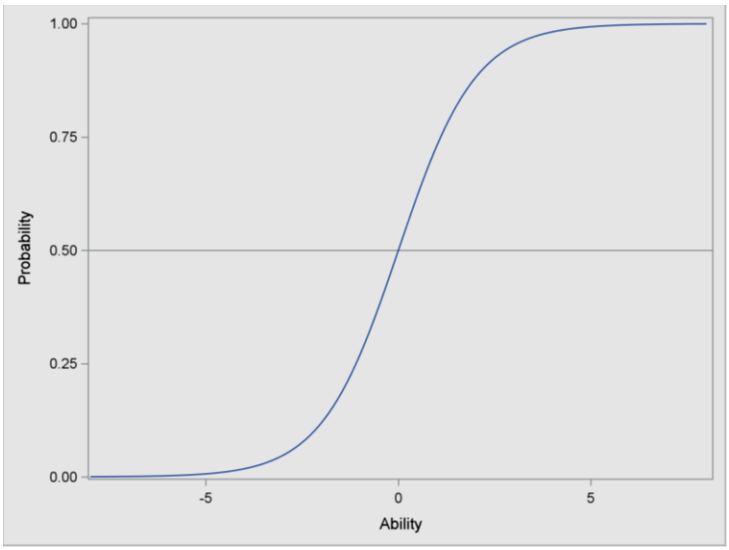


Fig. 2 Item Characteristics Curve
Adopted from Xinming An and Yiu-Fai Yung (2014)

"Under the item response theory, the standard mathematical model for the item characteristic curve is the cumulative form of the logistic function. The theoretical range of ability is from negative infinity to positive infinity. Practical consideration usually limits the range of values from -3 to +3." Most of the studies reviewed adopted either 1PLM, 2PLM, or 3PLM in determining the nature of the item characteristic curve of Christian Religious Studies Examination items but the present study adopted the 4PLM in the determination of the nature of item characteristic curve of West African Senior School Christian Religious Studies Examination for 2020. In order to present a clear graphical

representation of item level difficulty and how items discriminate among examinees with high and lower abilities of Christian Religious Studies Multiple Choice Items, the monotonicity assumption is very important.

Determining the dimensionality of Christian Religious Studies Multiple Choice Items reveals the number of abilities needed by an examinee to correctly respond to the items of the examination. Determining the item local independence provides information of how dependent or independent correct response to item are in Christian Religious Studies Multiple Choice Items, and determining the monotonicity of Christian Religious Studies Multiple Choice Items provides information about the two major properties of Christian Religious Studies Multiple Choice Items which are item facility and item discrimination.Interpreting students' scores and making judgement on the performance of students with respect to an examination such as Christian Religion Studies cannot be effective if the psychometric properties of such examination are not established. Establishing the psychometric properties of entails determining if the test is a unidimensional test or a multidimensional test, assessing if the items in such a test is statistically dependent on each other, that the response to an item gives a clue to the correct response to the corresponding item. In determining these, the model that best fit the data obtained from such instruments must be applied which will present the level of difficulty of the items of the examination, and the ability of the items to identify the low ability and high ability students.

Most students with lower ability may also respond correctly to an item which they should have responded wrongly given to their ability levels which now yield an asymptote parameter. Since Christian Religious Studies Multiple Choice Items is used to assess students' knowledge, morals and level at which learning has taken place and the outcome of the assessment will be used to take decision and recommendation; when the quality of the examination is not ascertained the outcome and decisions would be faulty. It is on this premise that this study tested the Dimensionality, Item Local Independence and Monotonicity of Item Response Theory on West African senior School Certificate Christian Religious Studies Multiple Choice Items as well as establish the item parameters since item response theory framework is more efficient in doing it. The study tested the assumptions of dimensionality, local independence and monotonicity of 2020 multiple choice test items of West African Senior School Certificate Examination in Christian Religious Studies. In order to gain more insight into the assumptions of IRT in Christian Religious Studies Question, four research questions were formulated to give direction to the study:

1. What is the dimensionality of the test items of CRSMC Question?
2. What is the item local independence of the test items of CRSMC Question?
3. What is the monotonicity of the multiple-choice items of CRSMC Question?

Materials and Methods

Design and Participants

This study adopted instrumentation research design. Instrumentation research design according to Binitio (2014) is a design that deals with all the processes involved in constructing research instrument that could be used appropriately in gathering data on the study. The design is suitable for the study because the researchers are interested in using the Item Response Theory to assess the properties of 2020 West African Senior School Christian Religious Studies Examination with respect to the dimensionality, local independence and monotonicity. The study participants comprised of 1,500 students. The decision for the number of participants was on the basis of the findings of De Ayala, (2009) who opined that on the basis of research, it appears that for Marginal Maximum Likelihood Estimation (MMLE) a sample of 1,000 or 2,000 persons will lead to reasonably accurate item parameter estimates with the 3PL model under normal conditions. The researchers adopt multi-stage sampling procedure in arriving at the sample size.

In the first stage, simple random sampling by balloting with replacement was used to draw two education zones out of six education zones in the State. Under this method, the names of each education zone was written on a piece of paper and folded into a container, the container was shuffled and the researcher picks the first paper, the paper was unfolded and an Education Zone emerge. The same paper was folded again and put back into the container. The process was repeated and the second Education Zone also emerge. In the second stage, the simple random sampling with replacement was repeated to draw four local governments from each of the two education zones that emerged. In the third stage, simple random sampling technique was used to select six schools from each of the four local governments that emerged from the second stage. The total of twenty-four intact SS3 classes of the selected schools formed the sample that was used for the study.

Measurement

The instrument for data collection was Christian Religious Studies Multiple Choice Questions (CRSMCQ) which was adopted from West African Senior Schools Certificate Examination for 2020. The instrument is made up of 50 items. Each item has four response options: A, B, C and D in which only one among the options is the key for each item. All the responses of the students were coded in Microsoft Excel, each item that was correctly responded to was scored as one (1) while the items that were not correctly responded to was scored as zero (0). The instrument was validated by three (3) experts' and they were asked to vet the items in terms of clarity, suitability, relevance and total coverage. Their comments and suggestions helped to improve the quality of the instrument. In order to determine the internal consistency of the instrument, thirty instruments were

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

administered to SS3 students outside the main study area but share similar characteristics in terms of the same curriculum content was used to teach them and the sample to be used for the study. The instruments were collected and scored. The instrument was further subjected to internal consistency reliability estimate and a reliability coefficient of 0.81 was obtained using Kudder Richardson 20 ($K\text{-}R_{20}$). The instruments was administered to 1,500 students in the selected schools using on the spot method to ensure the return of all the instruments. The time that was allocated for the response to the instrument was one hour (1 hour). This time was adopted because the same time is apportioned for the multiple-choice section of the examination. The question papers were retrieved by the researchers and the research assistants. The retrieved instruments were coded and marked in Microsoft Excel using the marking guide.

Data Analysis

The data collected was analyzed with respect to the research questions that guided the study. Research question one was analyzed using Normal Ogive Harmonic Analysis Robust Method (NOHARM) and percentage reduction value was used for research question one. In which the Root Mean Square Reduction (RMSR) percentage of less than 10 ($R\% < 10$) is considered to be unidimensional whereas percentage reduction equal to or greater than 10 ($R\% \geq 10$) is considered to be multidimensional. Research questions 2 was answered using Yen Q3 statistics to determine the item local independence of the test items; in which a correlation coefficient of pairs of items greater than 0.2 was considered as violating the local independence assumption. The Akaike Information Criterion (AIC) of each of the IRT model was compared, any model with the lowest AIC value was considered the model that best fits the data, and multidimensional four parameter logistic model of item response theory using R programming software was used to answer research questions 3. The acceptable values for students' ability (ϑ) theoretically range from $-\infty$ to $+\infty$, practically values between -3 and +3 were considered acceptable.

Ethical Approval

The ethics committee where the research was conducted granted ethical approval. Before the commencement of the study, the respondents were presented with informed consent forms to fill and sign. The informed consent forms were properly filled and signed. The data collection from the entire sample was in accordance with the ethical standards of the researchers' institutional research committee and with the 1964 Helsinki declaration.

Results

Table 1 shows the result of the dimensionality of 2020 multiple choice items of West African Senior School Christian Religious Studies Examination. The Sum of square Residual of 0.1089126, Tanaka's Goodness of fit Index of 0.9736329, Root Mean Square Criterion of 0.1264911, Root Mean Square Residual of 0.0094291 was obtained with a percentage reduction of 10. This is conclusive that the 2020 multiple choice items of West African Senior School Christian Religious Studies Examination have three dimensions. Hence, the 2020 multiple choice items of West African Senior School Christian Religious Studies Examination is a multidimensional test.

Table 1. Result of the Dimensionality of the Test Items

No of Dim	Sum of Square Residual	Tanaka's index of GFI	RMS Crit.	RMSR	Diff. in RMSR	Red. in RMSR	% Reduction
1	0.1824677	0.9558256	0.1264911	0.0122046			
2	0.1356097	0.9671697	0.1264911	0.0105215	0.0016831	0.137907019	14
3	**0.1089126**	**0.9736329**	**0.1264911**	**0.0094291**	**0.0010924**	**0.1038255**	**10**
4	0.0913251	0.9778907	0.1264911	0.0086343	0.0007948	0.084292244	8
5	0.0789102	0.9808963	0.1264911	0.008026	0.0137217	1.589208158	159

SSR= Sum of Square Residual, GFI= Goodness of Fit Index, RMSR= Root Mean Square Residual

Table 2 show the result of the correlation item residual of the local independence of 2020 multiple choice items of West African Senior School Christian Religious Studies Examination using Yen Q3 statistic for item 1 to item 12. By comparing the items, the table showed that of all the fifty items, items 12, 15, 29, 36 and 37 fulfilled the assumption of local independence because their correlation item residual are less than 0.2 used as benchmark for taking decision, while the other items correlation item residual are greater than 0.2, hence they all violates the assumption of local independence.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

Table 2. Result of the Local Independence of the Test Items

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item11	Item12
Item1	1.000											
Item2	0.206	1.000										
Item3	0.063	0.265	1.000									
Item4	0.329	0.195	0.137	1.000								
Item5	0.134	0.261	0.202	0.155	1.000							
Item6	0.262	0.165	0.184	0.443	0.104	1.000						
Item7	0.237	0.101	0.106	0.396	0.084	0.393	1.000					
Item8	0.214	0.148	0.092	0.348	0.125	0.339	0.380	1.000				
Item9	0.296	0.095	0.155	0.447	0.133	0.451	0.464	0.420	1.000			
Item10	0.254	0.142	0.124	0.404	0.164	0.392	0.379	0.362	0.443	1.000		
Item11	0.247	0.206	0.095	0.323	0.144	0.291	0.338	0.325	0.312	0.358	1.000	
Item12	-0.029	-0.081	-0.087	-0.054	-0.129	0.000	-0.065	-0.093	-0.020	-0.079	-0.076	1.000

Fig 1 shows the monotonocity of 2020 multiple choice items of West African Senior School Christian Religious Studies Examination. Monotonicity of the examination item reflect the true relationship between the trait and the responses to the item and as the ability level increases, the probability of getting the item correct increases monotonicially. The result shows that all the items with exemption of items 10, 11, 18, 20, 23, 27, 28, 33 and 41 have the ability to differentiate between the higher ability students and the lower ability students and they have good difficulty abilities. But items 10, 11, 18, 20, 23, 27, 28, 33 and 41 have good discrimination abilities but does not have good difficulty abilities.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

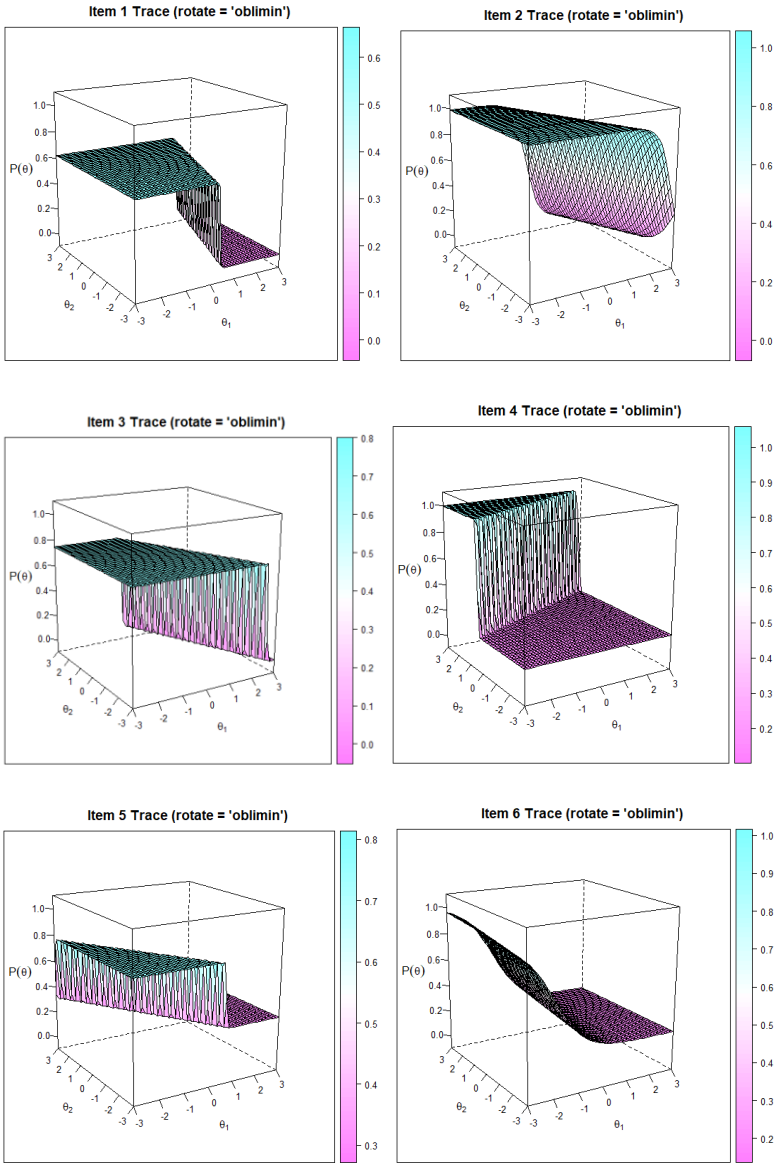


Fig 1. Monotonicity of the Multiple-Choice Items of West African Senior School Christian Religious Studies Examination.

Discussion

Dimensionalities of the Test Items

The study revealed that the Multiple-Choice Test Items is a multidimensional test, this is because the test has its Root Mean Square Reduction (RMSR) percentage greater than 10 (R% > 10). This gives an indication that the Multiple-Choice Test Items is underlined by more than one trait. The examinees to whom this examination is developed for require more than one ability level to respond correctly to the items in such examination. When this happens, an examinee may not have the required abilities to respond correctly to the

items in the test because such items do not only measure one underling construct but also other ability levels of the students. When this happens, the students may tend to score low in such examinations owing to the fact that these students are exposed to examination more than their ability levels. This means that the examination cannot be used to make decision based on the particular construct because it does not measure only that specific construct. This agrees with Reckase, (2009) that a test that requires only one trait is termed unidimensional test whereas, the test that measures two or more distinctive traits are called multidimensional test. In the same vein, Collins et al. (2018) noted that in assessing the dimensionality of test items, some tests tend to measure only one trait or attribute, hence the test is one-dimensional test. The finding also supported the view of Kose and Demirtasli, (2012) that when test measure only two, three or four traits or attributes they are called multidimensional test. The result is also in agreement with the findings of Ubi et al. (2012) who argued that examinations designed for selections of candidates might not be purely unidimensional, especially when items are fielded from a wide syllabus. There is also an agreement on unidimensionality of examination as reported by (Moyo, 2013) which revealed that there was no evidence of unidimensionality in Agricultural examination of Botswana.

Local Independence of the Test Items

The inter-item correlations of all the items of 2020 West African Senior School Certificate Examination in Christian Religious Studies fulfilled the assumption of local independence because their correlation item residual is less than 0.2 used as benchmark for taking decision. Items that meet the assumption of local independent establishes the reliability of the information obtained from the examination. Items that violated the local independence assumption contaminates the derive measures from the examination, it results in overestimation of reliability and consequently leads to small standard error estimate (SEE) which is a very severe problem in computerized adaptive testing where SEE is the criterion for terminating the test. Dependency among items can inflate reliability and give a fake impression of the precision and quality of the test. The finding of this study is in line with the study of Ayanwale et al. (2020) that established the violation of the 2018 Osun state unified multiple choice mathematics achievement test items using Non-linear factor analysis, Stout's Test of Essential unidimensionality, Factor Analysis, Full Information Factor Analysis and Bootstrap Modified Parallel Analysis Test. The study also aligns with the findings of Ubi et al. (2012) that reported that all the University Matriculation Examination in Nigeria from 2000 to 2003 met assumption of Item Response Theory on item local independence. The finding of this study is in agreement with the findings of Monseur et al. (2011) that PISA test format assessment violated the assumption of local independence.

Monotonicity of the Multiple-Choice Items

The result reveals nature of the monotonicity of the multiple-choice items of West African Senior School Christian Religious Studies Examination items. Monotonicity of the examination item reflect the true relationship between the trait and the responses to the item and as the ability level increases, the probability of getting the item correct increases monotonicially. Out of all the items except items 10, 11, 18, 20, 23, 27, 28, 33 and 41 have the ability level increases as the probability of getting the item correct increases monotonicially. This implies that as students' ability increases their ability to provide correct responses to the items also increases. This is in agreement with the findings of (De Mars, 2010) that the underlying assumption of monotonicity is that every response to an item on an instrument provides some inclination about the individual's level of the latent trait or ability and that the ability of the person (θ) in simple terms is the probability of endorsing the correct answer for that item. As such, the higher the individual's ability, the higher is the probability of a correct response. The result was also in agreement with Wind (2017) that assessed the monotonicity of test items by checking for non-decreasing patterns over increasing values of the latent trait score. The result of the study is also in tandem with the report of Kula-Kartal and Mor-Dirlik (2021) that monotonicity of items may affect predictions of different dimensionality analysis techniques concerning the size and sign of the inter-item correlations for instance, the factor analysis accepts that all scale items have linear and monotonically increasing item response curves.

Conclusions

The result revealed that the Multiple-Choice Test Items is a multidimensional test. The inter-item correlations of some of the items of 2020 West African Senior School Certificate Examination in Christian Religious Studies fulfilled the assumption of local independence. The Multiple-Choice Test Items of 2020 West African Senior School Certificate Examination in Christian Religious Studies has three underlying factors, indicating that it is a multidimensional test. Majority of the items of Multiple-Choice Test Items of 2020 West African Senior School Certificate Examination in Christian Religious Studies were suitable to the level of students it was designed for. All the items in the Multiple-Choice Test Items of 2020 West African Senior School Certificate Examination in Christian Religious Studies have good guessing factor index as such the correct response to one does not provide clue to correct response to other items. The study shows that some of the items did not violate the Item Response Theory assumption of item local independence. This indicated that the correct response to items were not dependent on each other. The implication of items violating assumption of local independence is that the derive measures from the examination is contaminated and this result in overestimation of reliability and consequently leads to small standard error estimate

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

which is a very severe problem in computerized adaptive testing where SEE is the criterion for terminating the test. Dependency among items can inflate reliability and give a fake impression of the precision and quality of the test.

Acknowledgements

The researchers are very grateful to all their colleagues and the authors whose studies were made use of during the course of this study.

Conflict of Interest statement

The authors declare no conflicts of interest regarding the publication of this paper.

References

Adejoh, M. J. & Obinne, A. D. E. (2013). *Basic issues in test, Measurement and Evaluation.* Markurdi: Asaofad printing limited.

Ali, A. (2006). *Conducting Research in Education and Social Sciences*; Nigeria: Tashiwa Networks Ltd.

Ayanwale, A. I., Isaac-Oloniyo, A. R. & Abayomi, I. M. (2020). Assessing item local independence and unidimensionality of BECE 2017 mathematics Multiple Choice Examination in Benue state. *African Journal of Theory and Practice of Educational Assessment (AJTPEA), 7*(3), 38 – 48.

Chester, E. N. (2008). *Item local independence and students mathematics ability in Calabar*, Unpublished M.Ed Thesis, University of Lagos, Nigeria.

Collins, M., Minobe, S., Barreiro, M. *et al.* (2018). Challenges and opportunities for

improved understanding of regional climate dynamics. *Nature Clim Change 8*, 101–108 (2018). https://doi.org/10.1038/s41558-017-0059-8

Dai-Trang L, (2013). *Applying item response theory modeling in educational research;* Graduate Theses and Dissertations. 13410. Retrieved from: https://core.ac.uk/download/pdf/38926389.pdf

De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory*; London: The Guilford press.

De Mars, C. (2010). *Item Response Theory: Understanding statistics measurement*. New York: Oxford University press

Donald, A; Lucy C. J. and Chris, S. (2010). *Introduction to research in education*. Canada: Wadsworth Cangage learning.

Kose, I.A. & Demirtrasli, N.C. (2012). Comparison of unidimensional and multidimensional

models based on item response theory in terms of both variables of test length and sample size. *Procedia-Social and Behavioral Sciences,* 46(4):135–140 DOI:10.1016/j.sbspro.2012.05.082

Kula-Kartal, S. & Mor-Dirlik, E. (2021). Examining the Dimensionality and Monotonicity of an Attitude Dataset based on the Item Response Theory Models. *International Journal of Assessment Tools in Education, 8*(2). 296 – 309.

Lee, C. (2019). *What is item analysis? And other important exam design principles.* Retrieved from https://www.turnitin.com/blog/what-is-item-analysis-and-other-important-exam-design-principles

Monseur, C, Baye, A., Lafontaine D. &Quitre, V. (2011). PISA test format assessment and the

local independence assumption. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments,4*(6). Educational Testing Service and International Association for the Evaluation of Educational Achievement https://orbi.uliege.be/bitstream/2268/103137/1/IERI_Monograph_Volume04_Chapter_6.pdf

Moyo, W. (2013). Causes and Effects of Poverty on Academic Achievements of Rural

Secondary School Students: Case of Tshazi Secondary School in Insiza District. International *Journal of Asian Social Science*, 3, 2104-2113. https://archive.aessweb.com/index.php/5007/article/view/2568

Nwani, S. U. (2021). *Trends in the psychometric qualities of Basic Education Certificate Examination in Basic Science in Enugu State*. An unpublished M.Sc project submitted to the Department of Science Education, University of Nigeria, Nsukka.

Nworgu, B.G. (2015). *Educational measurement and evaluation: theory and practice.* Nsukka: University Trust Publishers.

Okwilagwe, E. A. & Ogunrinde, M. A. (2017). Assessment of unidimensionality and local independence of WAEC and NECO 2013 Geography achievement tests. *African Journal of Theory and Practice of Educational Assessment,* 5(3), 31 – 45.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer- Verlag.

Soura, B. D., Bastien, R., and Fallu, J. S. (2016). Evaluability assessment of an intervention to prevent the use of psychoactive substances during the transition from primary to secondary school. *Canadian journal of program evaluation, 31, (2), 211-231.*

Ubi, U. E., Umoinyang, P. A., & Joshua, L. I. (2012). The relationship between multiple choice and essay response questions in assessing economics understanding. *Economic Record,75*(4), 348-357.DOI: 10.1111/j.1475-4932.1999.tb02571.x

Wind, S. A. (2017). Examining the psychometric quality of multiplechoice assessment items using Mokken Scale Analysis. *Journal of Applied Measurement, 17*(2), 142 – 165.

Xinming An & Yiu-Fai Yung (2014). Item Response Theory: What It Is and How You Can Use

the IRT Procedure to Apply It. https://support.sas.com/resources/papers/proceedings14/SAS364-2014.pdf

Ziegler, M., & Hagemann, D. (2015). Testing the Unidimensionality of items. *European Journal of Psychological Assessment, 31*, 231-237