Relative Prevalence of Local Item Dependence of Two Major Senior Secondary School National Examinations Conducted in Nigeria

Abimbola Olubukola ALU, Ph.D
Department of Guidance and Counselling,
Adekunle Ajasin University, Akungba-Akoko,
Ondo State, Nigeria
e-mail: bimboalu@gmail.com

Abstract

The assumptions in all Item Response Theory (IRT) models include Local independence of test items, meaning the items in a test should not be related to each other. Sharing common stimuli, which is prevalent in mathematics tests can be a potential source of local item dependence (LID) which is a violation of this assumption. It is argued in the literature that LID results in biased parameter estimation and affects the unidimensionality of the test. In this study, the prevalence of local item dependence in the two major senior secondary school national examinations conducted in Nigeria and the resultant effects are examined. The research design adopted was descriptive survey. The population consisted of secondary school students in Osun State. A sample of 1800 students was selected for the study. The instruments used for the study were 2009 West African Examination Council (WAEC) and National Examination Council (NECO) Senior Secondary School Certificate Examination (SSCE). Data were subjected to confirmatory factor analysis, and t-test. Findings showed that LID occurrence in both examinations but there was no significant difference between its occurrence in both examinations ($t = 0.205$, $p > 0.05$). The study showed that both WAEC and NECO SSCE examinations were comparable in terms of psychometric properties. Based on the findings, it was concluded that assumption of Local independence of test items was disregarded in the process of constructing items for WAEC and NECO SSCE, both WAEC and NECO SSCE exhibited occurrence of LID Therefore, it is recommended in order to ensure test fairness that LID should be evaluated and fully incorporated as one of the steps involved in the process of test construction to deal with occurrence of redundant items in the tests.

Keywords: Prevalence, local item dependence, national examinations, students' academic performance

Introduction

In Nigeria secondary education system, the two major examination bodies which are the West African Examinations Council and the National Examinations Council are responsible for developing items for their examination. Both bodies are saddled with the responsibility of conducting national examinations which make grounds for assessment across the country as uniform as possible. This is in line with the Federal Republic of Nigeria (FRN, 2004) national policy on education which states that the national examinations should be as valid and as fair as possible to all students.

These national examinations are some of the important parts of the Nigerian educational system in that they give indications of the quality of education; calibrate grades for certification which are used for job seeking or admission into higher institutions. Any examination seeking to achieve the purposes stated above is said to be valid and this is the paramount ambition of the National policy on education which states that education in the senior secondary school must be equal for all students (NPE, 2004). A test is valid if the items on the test do not cause systematic errors in the measurement. A valid test should not consist of biased test items. This goes further to state that a test is valid if it is not biased. To ensure that items that constituted an instrument meant for examination of students in accordance to the objectives of FRN (2004) in its policy on education, it appears necessary that these items obey one of the assumptions of IRT models in general which is the local independence assumption. The items that are put to IRT model analysis are required to be independent of each other. That is, a correct or wrong answer to one achievement test item should not lead to a correct or wrong answer to another item within the same instrument. This means that there should not be any correlation between two items after the effect of the underlying trait of testees is partialized. The items should only be correlated through the latent trait that the test is measuring (Lord & Novick, 1968). If there are significant correlations among the items after the contribution of the latent trait is partialized, then the items are locally dependent or there is a subsidiary dimension in the measurement which is not accounted for by the mainIRT (Rasch) model based dimension (Lee & Frisbie, 2004).

However, a major issue in psychometrics is what happens when items continue to relate with each other, after accounting for their contribution to the latent trait. In the context of the Rasch model, this relationship is termed Local Item Dependence (LID) and represents a prerequisite and assumption of the model. Violation of this assumption states that there is still some covariation between items, although the relationship of each item to the latent trait has been accounted for. Every attempt is made by test developers at developing multiple choice items that focus on discrete skills or content areas and at the same time, such items are explicitly written to minimise local item dependence. Despite all these efforts, LID is still being observed. Sireci, Thissen and Wainer (1991) found evidence of LID among reading comprehension items linked to the same reading passage. Also, Bell, Pattison and Withers (1988) found LID to be stronger in Mathematics items than in the verbal items. Ferrari, Huynh and Bagli (1997) also observed the existence of LID in a large scale educational performance assessment in reading and Mathematics. However, the presence of LID in the examinations implies a deficit in the precision and quality of the examinations and the possibility of undeserved test score (Baghaei, 2008).

If the assumption of local item independence is violated, any statistical analysis based on it would be misleading according to Wang and Wilson (2005). Specifically, estimates of the latent variables and item parameters will generally be biased because of model mis specification, which in turn leads to incorrect decisions on subsequent statistical analysis, such as testing group differences and correlations between latent variables. In addition, it is not clear what constructs the item responses reflect, and consequently, it is not clear how to combine those responses into a single test score, whether IRT is being used or not (Wang &

Wilson, 2005). However, in practical testing contexts the local independence assumption gets violated very easily. This happens when several questions are based on a single prompt. This dependence as submitted by Monseur (2009) could result from similarities in the cognitive processes involved in several items or from the specificity of the context. Prior knowledge on the stimulus or the interrelation of the information required to answer different items may certainly be major sources of local item dependence. Ferrara, Huynh and Bagli (1997) posited that if a complex problem is posed in Mathematics, cluster that requires multiple steps to reach a solution, in all likelihood, the clustering of such free-response items would create some level of dependence in the responses in each cluster. When the assumption that an examinee approaches each test item as a new problem without any information gained from responding to any other item is violated, LID which result in failure of proper test score interpretation ensues. Other effect of LID includes biased parameter estimation, giving of a fake impression of the precision of the test and quality of the test. If an item relates to another item on a test, it should be tracked to ensure that the two items are not presented on the same form of an examination since this leads to a redundant measure of proficiency, thereby providing no additional or unique information about the quality of the examinee. According to Wang & Wilson (2005), item dependence can also lead to inaccurate estimation of item parameters, test statistics, and examinee competency and any statistical analysis based on it would be misleading, as estimates of item parameter will be biased. Furthermore, violation of this assumption has major implications regarding the validity of estimates, for instance, on discrimination. It is observed that LID represents a serious psychometric nuisance and should be evaluated at all times, hence it becomes pertinent to ask such question as; do the items of the examinations of the two bodies that conduct examinations which are used to make decisions such as admission to tertiary institutions to pursue academic career of an average Nigerian student contain items that are locally dependent on one another?

The purpose of this paper is to examine the relative prevalence of LID in Senior School 2009 Mathematics examinations conducted by WAEC and NECO in Nigeria. To investigate this, one research question and one hypothesis guided the study.

Research Question

What is the relative prevalence of LID in WAEC and NECO SSCE Mathematics examinations?

Hypothesis

There is no significant difference in the occurrence of LID in WAEC and NECO SSCE Mathematics examinations.

Methodology

The research design adopted for the study was descriptive survey. The population for this study was the senior secondary school class 3 (SSS3) students in Osun State for the year 2017/2018 academic session. This population was chosen for the study because the

instruments used were NECO and WAEC 2009 Mathematics examinations and the content of the syllabus must have been adequately covered in the classroom. It is assumed the students must have been made to attempt similar tests in preparation for the final senior school certificate examination.

The sample for the study consisted 1800 SSS3 Mathematics students of Osun State for the year 2017/2018 academic session. Multistage sampling technique was employed for the study. Stage one involved selecting four Local Government Areas from each of the three senatorial districts in Osun state using simple random sampling technique. Stage two involved purposive sampling of one school reputed to have significant concentration of heterogenous students in terms of age, gender and background as well as densely populated classrooms from each local government area while two other schools, using simple random sampling technique, were selected from each of the selected local government area. Stage three involved random selection of 50 students from each of the 36 schools. In all, a total of 1800 senior secondary school three students were used for this study. The instruments used for the study were the WAEC and NECO SSCE Mathematics examinations items for the year 2009. WAEC SSCE had 50 items while NECO SSCE had 60 items. The items were separately compiled to encourage the students to put in their best. The instruments were administered under conducive examination conditions with the assistant of the subject teachers and trained research assistants under the supervision of the researcher. The responses to the instruments were collected after the administration on the same day. The data were gathered from responses of students to the 50-items and 60-items Mathematics tests by WAEC and NECO 2009 Senior Secondary School Certificate Examinations. The test items were scored 0 for wrong answer and 1 for correct answer. The maximum score obtainable for WAEC was 50 while the maximum for NECO was 60 while the minimum score for both examinations was 0. The confirmatory factor analysis (CFA) of the PRELIS/LISREL computer program was used to analyse the data collected. Using CFA, excess covariation among items in the residual matrix of a single-factor CFA model is indicative of items with LID. This technique was used to answer the research question while the hypothesis was tested using t-test.

Results

The results presented above are used to answer the research question and test the hypothesis. Table 1 shows the covariance matrices of 0.4 and above in the NECO and WAEC Mathematics Examination.

Table 1: Covariation matrix values in NECO and WAEC SSCE Mathematics Examination

Covariance Matrix value	No of occurrence in NECO	Percentage	No of occurrence in WAEC	Percentage
0.4	188	10.5	129	10
0.5	85	4.7	106	8.4
0.6	51	2.3	48	3.8
0.7	18	1	19	1.5
0.8	02	0.1	03	0.2
0.9	06	0.3	Nil	Nil

The results presented above are used to answer the research question and test the hypothesis. The Table 1 reveals the covariance matrices of 0.4 and above in the NECO and WAEC Mathematics Examination. Thetable shows the relative prevalence of LID in the two examinations. The result showed that out of the 1800 pairwise covariation of items in NECO Mathematics examination, 188 (10.5%) had a covariation of 0.4; 85 (4.7%) had a covariation value of 0.5; 51 (2.3%); 18 (1%); 2 (0.1%) and 6 (0.3%) had covariation of 0.6, 0.7, 0.8 and 0.9 respectively while in WAEC Mathematics examination, out of 1266 pairwise covariation of items, 129 (10%) had a covariation value of 0.4; 106 (8.4%) had 0.5, 48 (3.8%) had 0.6; 19 and 3 occurence had respective covariation of 0.7 (1.5%) and 0.8 (0.2%).

In order to test the research hypothesis, an independent t-test was carried out on the relative occurrence of LID in WAEC and NECO SSCE 2009 Mathematics examinations

Table 2: t-test analysis of the occurrence of LID in the examinations

Variables	mean	N	S	t-cal	t-tab	Df	Sig
NECO	58.3333	1800	70.87642	3.5503	1.960	3598	0.05
WAEC	50.8333	1800	54.85769				

Table 2 shows the mean of 58.333 and 50,833 for NECO and WAEC respectively with their standard deviations showed as 70.876 and 54.857 respectively. The t-calculated value obtained was 3.5503 while the table value was 1.960 at p=0.05 level of significance. Since t-calculated value (t=3.5503) is greater than the t-tab (1.960), therefore the hypothesis that says there is no significant difference in the occurrence of LID in the WAEC and NECO SSCE 2009 Mathematics examinations is rejected. This indicates that there was significant difference in the occurrence of LID in the WAEC and NECO SSCE 2009 Mathematics examinations. This result is to the favour of WAEC SSCE.

Discussion

The results revealed that there was presence of LID in the WAEC and NECO 2009 Mathematics SSC examinations despite the efforts put at standardizing the examinations' test items by the examination bodies. This is in agreement with the findings of Bell, et al. (1988) that LID is found to be stronger in Mathematics items. Ferrara, et al. (1997) also observed the existence of LID in a large scale educational performance assessment in reading and Mathematics, the findings which are similar to this present finding. The presence of LID is indicated by excess covariation (high covariance value) among pairwise items in the residual matrix. Items on the same test measuring a single latent trait are expected to covary, however, high covariation is not desirable. Each item in a test is supposed to measure a facet of an examinee's ability. The presence of LID prevents the independence of paired items from carrying out this responsibility such that some items duplicated the effort or achievement of other items. In addition, some facets of the examinee's ability are underestimated or overestimated with the presence of LID in that an examinee that is privy to be familiar with items indicating LID will have a higher score over the less privileged examinees. This implies that any decision made based on this type of test will be biased and unfair. Given the important role of the two certificate examinations conducted by WAEC and NECO, the results of this study have shown that it was important to investigate the prevalence of LID in the examinations.

The results showed that six items demonstrated the presence of LID in WAEC Mathematics examination while eight items demonstrated LID in NECO Mathematics examination using the covariance matrix approach. There was no significant difference found in the occurrence of LID in the two examinations. This may be attributed to the fact that both examinations are: commercial tests constructed by different bodies and may have undergone some test standardization procedures; one of the examination bodies (WAEC) has long been in existence before the other. However, the presence of LID in the examinations implies a deficit in the precision and quality of the examinations and the possibility of undeserved test score (Baghaei, 2008). That is, the proficiency of the examinee will be underestimated or overestimated (inaccurate estimation) to the extent to which LID occurred on the test as the case may be. Item dependence can also lead to inaccurate estimation of item parameters, test statistics, and examinee competency and any statistical analysis based on it would be misleading, as estimates of item parameter will be biased (Wang & Wilson, 2005). The presence of LID on the test probably lead to an increase in the test length, creating redundant items (items that do not contribute any specific and unique information to the test) which may lead to an increase in the test reliability coefficient thereby given a false impression of the test.

Conclusion and Recommendations

Based on the findings of this study, it was concluded that both WAEC and NECO SSCE examinations exhibited the occurrence of LID and were comparable in terms of psychometric properties but this could be improved by dealing with the occurrence of

redundant items in the tests. The occurrence of LID is less in WAEC than NECO in 2009 SSCE Mathematics examinations.

Therefore, it is recommended in order to ensure test fairness that LID should be evaluated and included as one of the steps involved in the process of test construction. Also, efforts should be made in the selection of final test items to expose the items to stringent moderation procedure by test experts through qualitative procedures to improve test quality.

References

Baghaei, P. (2008). Local dependency and Rasch measures. Available at http://www.rasc.org/rmt/rmt213a.htm

Bell, R. C., Pattison, P. E. & Wither, G. P. (1988). Conditional independence in a clustered item test. *Applied Psychological Measurement*, 12(1), 15-26.

Federal Government of Nigeria (2004). National policy on education. Lagos: Nigerian Educational Research and Development Council.

Ferrara, S., Huynh, H. & Bagli, H. (1997). Contextual characteristics of locally dependent open-ended item clusters on a large-scale performance assessment. *Applied Measurement in Education*, 12, 123-144.

Lee, G. & Frisbie, D. A. (1999). Estimating reliability under a generalizability theory model for test scores composed of testlets. *Applied Measurement in Education*, 12, 237-255.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test theories*. Reading, Massachusetts: Addison-Wesley.

Monseur, C. (2009). Effect of local item dependence of country performance estimates. PISA Research Conference, Kiel, Germany from www.pisaresconf09.org.

Sireci, S. G., Thissen, D. & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement,* 28(3), 237-247.

Wang, W. & Wilson, M. (2005). Exploring local item dependence using random-effects facet model. *Applied Psychological Measurement,* 29(4), 296-318.