Journal of Agricultural Economics, Extension and Rural Development: ISSN-2360-798X, Vol. 5(1): pp, 590-599, January, 2017.

Full Length Research Paper

Factor and IRT Analysis for 2013 BGCSE Agricultural Examination

Sello E. Moyo

Research, Statistics and Demography, Kanye SDA College of Nursing

E-mail: sellmoyo@gmail.com*

Accepted 11th January, 2017

The deterioration of students' performance in the Botswana General Certificate of Education (BGCSE) examination results is a disturbing trend that bothers parents, teachers, policy makers and government. This problem prompts this study on dimensionality analysis of students' performance in 2013 BGCSE Agriculture Examination, to determine its dimensionality. The population of the study was all the 12784 students' responses who sat for the 2013 BGCSE Agriculture examination. The students' responses were analysed using factor analysis and IRT (1PL, 2PL and 3PL) models to examine the psychometric parameter estimates of the forty test items; dimensionality analysis and the chi square test for each test item that fitted in the three IRT models. The findings revealed that examination was not unidimensional. None of the 40 items fit the 1PL. Only one fitted the 2PL and 8 items fitted the 3PL. In conclusion, the results of this study, as it explored the national assessment tool, showed that 2013 BGCSE Agriculture Examination was not unidimensionality. It was, therefore recommended that test developers and examination bodies should consider improving the quality of their test items by conducting IRT psychometric analysis for item validation.

Keywords: Agriculture education, unidimensionality, factor analysis, item response theory, invariance, psychometric analysis.

INTRODUCTION

There is not yet an art through which educationalists can look at a learner on the face or open up his/her brain to know if he/she has learned or how much he has learned. So educational measurement, that is the art and science of quantifying or qualify the cognitive, affective and psychomotor behaviour of learners is inevitable. A valid measure for behaviours analysis in particular cognitive is well classified by the Bloom taxonomy. Thus in education, items are developed as challenges or task which when encountered by a testee will provoke an action appropriate for the trait under measurement latent in the testee (Nenty, 2004).

In return, it is also imperative to evaluate the dimensionality evidence of any given examination items as a standard practices in educational measurement. Such evaluation of items is also very important to teachers, examiners and the public in general for decision making. According to Siamisang and Nenty (2012), a test is never better than the quality of its items, so to identify problematic items is only possible through item analysis. Many scholars (Nenty, 2004; Nworgu, 2011; Umoinyang, 2011) have recommended that item response theory (IRT) approach be applied to analyze dimensionality of examination as a means of contributing to test fairness.

According to Hattie (1985) in Hill (2007) "one of the most crucial and basic assumptions of IRT is that a set of items forming an instrument of all measure just one thing in common" (p. 139). A good measurement model should provide a guide to the process of constructing and administrating a test in such a way that the trait interaction with the tasks implied by items depend on, and only on, the values of the trait and that of each item designed to measure the trait. Hence unidimensionality assumption is fundamental for a valid operationalization

of all IRT models applied to data from dichotomously scored achievement items. If a test is truly unidimensional, then the variance common to all the items represents. Given the decline in students' achievement in public examinations in Botswana, there is need to assess unidimensionality to enable fair interpretation.

Theoretical Foundation

Assessment of Test Dimensionality using Factor Analysis

Factor analysis (FA) is analytic tool used to provide evidence regarding fit and unidimensionality, such as scree plots and eigenvalue-based indices (Reckase, 1979). FA is used to determine the underlying structure of a measuring instrument and at times it is used to investigate the nature of the underlying factors in an existing scale. Principle component analysis (PCA) together with eigenvalue plots is a common way to evaluate test dimensionality and has been used for decades (Lord & Novick, 1968; Hattie, 1985). The percentage of total variance explained by the first principle component is often regarded as an index of unidimensionality. The higher percentage of total variance of the first principle component accounts for the closer the test is to unidimensionality. However there is one downside with eigenvalue plots, that is, no statistical index available to decide the number of underlying dimensions.

Various criteria have been proposed to solve the problem of eigenvalue plots. Reckase (1979) recommended that a percentage of 20 or more of the total variance explained by the first principle component is necessary for the data to be viewed as unidimensional. Similarly Lord (1980) suggested checking the ratio of the first to the second eigenvalue, and compares that with the ratio of the second to any of the other eigenvalues. Kaiser (1970) suggested retaining any components with eigenvalues larger than 1.

In addition, PCA provides analytic tools for exploring model-data fit used to explore hypotheses regarding invariant measurement. Conversely single model-data fit index can detect all of the possible sources of misfit (Reckase, 1979). Model-data fit is sample-dependent, and the key question in judging fit is: How good is good enough? There is no definitive statistical answer to this question, but various indices (including FA) can provide evidence to support inferences regarding invariance within a particular context (Randall & Engelhard, 2010).

Assessment of the Model Fit

According to Duong (2004), the judgment of the suitability of a model for solving particular measurement problems can be based on three kinds of evidence: the appropriateness of the model assumptions; the invariance of the obtained model parameters, and the accuracy of the model predictions. For unidimensional IRT models, the checking of the model assumptions should focus on four fundamental assumptions: unidimensionality, equal discriminating power, minimal guessing, and non-speeded test administrations (Hambleton, 1989). Evidence on the appropriateness of model assumption can be used to select IRT models.

To check the invariance of ability, one method is by giving examinees more than one set of items where items in each set have various levels of difficulties. The score of examinees in all the tests should be correlated because the expected ability score for each examinee does not depend on the choice of item if the model fits the test data. Moreover one of the useful methods in checking model prediction is to analyse item residuals. In this method, after a model is chosen and parameters are estimated, predictions about performance of various groups of examinees are made. Predicted results are then compared with actual results. If the residuals are small, it is reasonable to accept the accuracy of the model predictions (Duong, 2004). The benefit of IRT models is to improve the development of test items and guide the examiner to balance items not just by testing higher order thinking. Rather, if an item misfit is diagnosed, it is due to poor item quality. For example confusing distractor in the multiple choice paper and hence such item is removed from test form or replaced (Mellenbergh, 1994).

According to Royal and Puffer (2013), evaluation of dimensionality of the multiple choice examination involves tests of fit, principal components analysis (PCA) of standardized residual correlations and data-to-model fit both overall and by individual item analysis. Generally, chi-square fit statistics are required to be non-significant (Bonferroni adjusted). Residual fit statistic are expected to be within a given range ± 2.5 for individual items and with a mean fit residual value close to 0.0 and standard deviation approaching 1.0 (usually < 1.4) for summary statistics (Velde, Beaton, Hogg-Johnston, Hurwitz & Tennant, 2009). This is helpful to discern if multiple dimensions are present and exactly where these dimensions might be in the dataset. In this study detailed analysis and interpretation of the 2013 BGCSE Agriculture examination results will go a long way in performing an exploratory diagnosis.

Statement of the Problem and Purpose of the Study

Educational standard for the evaluation instruments for the public examination purposes in Botswana like other Africa countries has for years been dominated by

Classical Test Theory (CTT) despite its weaknesses (Nworgu & Agah, 2012).

Existing evidence shows that BGCSE results for students in public secondary schools are not as good as they used be. In the past, the results at times deteriorate yearly across all schools and levels (Ministry of Education, 2013). Such deterioration may be rooted from many distinct factors which are significant variance. Thus some could be either items variance or non-items variance (for example too worded items). Hence if such score is put to use in any examination-score-based decision, such decision might be unfair and biased. In Botswana public examinations, there be might items with significant probability neither fit for testing nor unidimensional. There is also a worrisome observation that Botswana Examination Council (BEC) does not seem to subscribe to the modern way of analysing students' results; instead the council seems to opt for analysis of raw scores (Thobega & Masole, 2008).

With IRT, calibration of items without reference to the items by a test of fit of the model is possible. Once items have been shown to fit the model, such items are chosen for test construction. Item calibration can be sample-free through controlling the influence of the ability level on the sample-bound item scores (Duong, 2004).

In an attempt to contribute a solution to this problem, the current study is purposeful in examining the dimensionality of Agriculture examination as a means of generating information with which contribution could be made to the improvement of test development. The specific questions of the study are:

Research Questions

1. What is the dimensionality of 2013 BGCSE Agriculture multiple choice items?
2. What are the items of 2013 BGCSE Agriculture multiple choice items that fit the 1PL, 2PL and 3PL model?

Significance of the Study

Examination results in Botswana are used as input in various decisions. This study contributes to stakeholders in educational measurement; among them are measurement specialists, classroom teachers, policy makers and BEC. The findings of this study will be of immense importance to examination council like BEC in the evaluation of the items for agricultural examinations and improving the quality of achievement examinations. In like manner, the students to whom this study is targeted would either directly or indirectly benefit. This is because if the orientation of their examinations achievement is towards fairness they stand the chance of benefitting maximally.

LITERATURE REVIEW

Ubi, Joshua and Umoinyang (2012) sampled 800 candidates' scripts from a pool of examination scripts of candidates who sat for the Joint Admissions and Matriculation Board's University Matriculation Examination (JAMB-UME) in Cross River State, Nigeria for the years 2002 and 2003. The purpose of the study was to assess the dimensionality of mathematics items using factor analysis. Results showed that JAMB-UME test revealed five significant dimensions and they concluded that examinations designed for selection of candidates might not be purely unidimensional, especially when items are fielded from a wide syllabus.

They recommended, among others things that, since it might not be possible to set tests, particularly mathematics, that are purely unidimensional, test practitioners especially those in charge of selection examinations should endeavour to meet the principles of item construction like ideal item difficulty, high discrimination and high option distraction indices to compensate for violating unidimensionality requirement. The analysis of unidimensionality in this study would have been better if Rasch model was used. Despite that, the study paved way to have an insight into the obstacles to attaining unidimensionality in some examination, like mathematics.

In a similar study by Adedoyin (2006) of Botswana Examination Council (BEC) 2004 junior secondary school final examination in mathematics, it was found that the first factor in an exploratory factor analysis accounted to only 15.05% of the variance of the entire 38 items used. Only 2 of the 38 items were found to fit 1-parameter, 11 were found to fit the 2-parameter and 16 were found to fit the 3-parameter IRT models. Based on repeated measure analysis of data from the 11 items that were found to fit the 2-parameter model, IRT person- and item-parameter estimates were invariant, but the discrimination-parameter as well all the CTT parameters was found to be variant.

METHODOLOGY AND RESEARCH DESIGN

This is descriptive analytic research design conducted to diagnose the underlying dimension of 2013 BGCSE Agriculture Examination and this provides the researcher's direction on how to improve and monitor the instrument from one test administration to the next (Boone & Scantlebury, 2005). This study target 12784 Form 5 candidates' responses to items in Paper 1of BGCSE Agriculture Examination administered to the 32 public senior secondary students both government

and government aided schools in Botswana. The multiple choice component (Paper 1) of the examination carries the same 40 percent weight as for Paper 2 (constructed response items) contribution to the whole BGCSE Agriculture Examination.

In this study, every student's responses to multiple choice items in BGCSE Agriculture were given equal chance to be selected and this enhanced the external validity of the study. In effect, students' academic records in Agriculture examinations for 2013 were available. The researcher retrieved the entire students' responses to every item for 2013 Agriculture multiple choice examination. Permission from BEC was requested to retrieve students' academic records on Agriculture examination for 2013 final year. The scores for BGCSE Agriculture are assumed to be valid, on the basis that BEC has intensive panel-base who deals with content analysis and face validation for every subject. It also assumed that the instrument was reliable in which the examination scores for students were attained.

ANALYSIS OF DATA AND INTERPRETATION OF THE RESULTS

Q1: What is the dimensionality of 2013 BGCSE agriculture multiple choice items?

To answer this question, the responses of the students on the 40 multiple-choice items of BGCSE Agriculture Examination were subjected to factor analysis. Factor analysis was performed to determine whether or not a dominant factor existed among all items as it was expected that the BGCSE agriculture examination would come up with one dominant factor. This factor would represent the construct underlining the agriculture learning domains measured by the examination. A Principal Component Factor Analysis (PCFA) was conducted to determine the underlying structure of the data. The initial eigenvalues were greater than 1, which are considered significant.

Table 1 shows the percentage variance accounted for by each of the variables. Nine factors had eigenvalues over Kaiser's criterion of 1 and in combination explained 10.05 %. Thus, the first eigenvalue was 3.82 greater than the next eight eigenvalues (1.358, 1.146, 1.115, 1.102, 1.069, 1.040, 1.025, and 1.018) respectively. The first factor explained only 10.05 % of the variance in the data set. The second factor explained 3.40 % of the remaining variance. The rest of the variance was explained by the other 38 factors with 7 factors each having a percentage of variance between 2.80 and 2.50, then 30 factors each having a percentage of variance of between 2.49 and 1.80. These last 30 factors were eliminated because they did not contribute to a simple factor structure and failed to meet minimum criteria of having a primary factor loading of greater than 1 eigenvalues.

A scree plot was produced to determine whether unidimensionality could be inferred. The scree plot should provide a convenient way of visualising a dominant factor from principal component analysis. An inspection of the scree plot of Figure 1 showed a high visual representation of relatively the first factor, but which accounts for only 10.05% of the total items variability.

In effect, the overall analyses indicated that nine distinct factors with eigenvalues bigger than 1.0 underlay in 2013 BGCSE Agriculture Examination and they accounted only for 32.34 % cumulative variance (see Table 1).

Q2.What is the items of 2013 BGCSE Agriculture multiple choice items that fit the one-parameter, two-parameter and three-parameter logistic model?

To answer the question of whether 2013 BGCSE Agriculture examination items do fit IRT models as a means to assessing the evidence of unidimensionality, the utility of the IRT model is dependent upon the extent to which the given responses fit these models.

To determine whether the test item fitted the model, a chi-square test was run on the data set using BILOG-MG V3.0 item analysis computer programme to establish whether the items fitted the 1PL, 2PL and 3PL models. Table 2 showed the results of the chi-square statistics. The chi-square goodness of fit analysis showed that none of the items fitted the 1PL model because their residuals variances were statistically significant. With a 2PL model, only 1 item fitted model that is Item 27 in which, its residual variance was not statistically significant.

From the chi-square values from the 3PL model, it is evident that thirty-two items representing 80% of the total items in the test were statistically significant through their residuals variances and hence do not fit 3PL because their residuals variances were statistically significant. The table also indicated that 8 items representing 20% of the total test were not statistically significant and this means that they fit the 3PL because their residuals variances were not statistically significant. For the 3PL model Item 18 and 32 were omitted from the calibration as its initial slope was less than - 0.15.

Table 1: Total Variance explained of the 2013 BGCSE Agriculture Examination

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% Variance	ofCumulative %
1	4.021	10.053	10.053	4.021	10.053	10.053
2	1.358	3.395	13.447	1.358	3.395	13.447
3	1.146	2.866	16.313	1.146	2.866	16.313
4	1.115	2.787	19.099	1.115	2.787	19.099
5	1.102	2.755	21.855	1.102	2.755	21.855
6	1.069	2.674	24.528	1.069	2.674	24.528
7	1.040	2.599	27.127	1.040	2.599	27.127
8	1.025	2.563	29.691	1.025	2.563	29.691
9	1.018	2.545	32.236	1.018	2.545	32.236
10	1.000	2.499	34.735			
11	.995	2.486	37.221			
12	.984	2.459	39.681			
13	.974	2.434	42.114			
14	.962	2.405	44.520			
15	.957	2.393	46.912			
16	.942	2.356	49.268			
17	.939	2.347	51.616			
18	.934	2.336	53.951			
19	.924	2.311	56.262			
20	.919	2.298	58.560			
21	.909	2.274	60.834			
22	.905	2.262	63.096			
23	.897	2.242	65.337			
24	.887	2.217	67.555			
25	.882	2.204	69.759			
26	.876	2.190	71.949			
27	.870	2.174	74.123			
28	.856	2.140	76.263			
29	.847	2.118	78.381			
30	.830	2.075	80.456			
31	.826	2.066	82.522			
32	.820	2.049	84.571			
33	.815	2.037	86.607			
34	.808	2.020	88.627			
35	.795	1.989	90.616			
36	.778	1.945	92.561			
37	.773	1.931	94.492			
38	.754	1.885	96.378			
39	.730	1.824	98.202			
40	.719	1.798	100.000			

Extraction Method: Principal Component Analysis

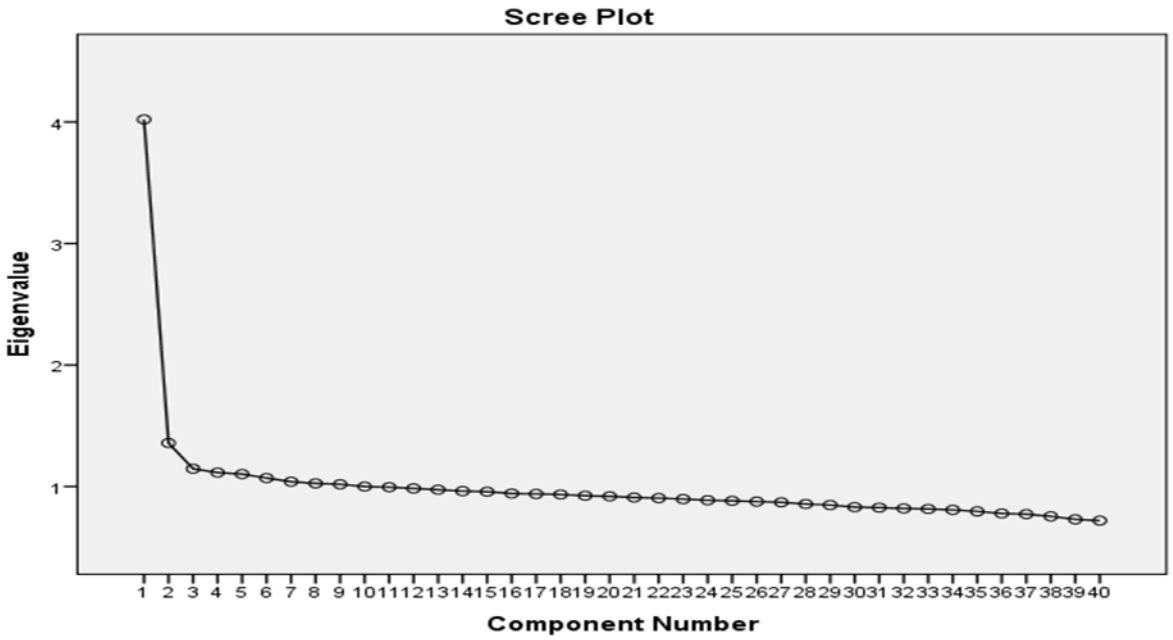

Figure 1: Scree plot showing dimensionality of 2013 BGCSE agriculture items

DISCUSSION

Dimensionality of 2013 BGCSE Agriculture Examination

In assessing dimensionality of a set of 2013 BGCSE Agriculture items, it was discovered that the examination is not unidimensional. The first factor did not meet Reckase (1979) recommendation that a percentage of 20 or more of the total variance

explained by the first principle component is necessary for the data to be assumed to be unidimensional. That is, a factor analysis on the inter-item correlation matrix should show that the first factor accounts for at least 20 of the variance of the unrotated factor matrix or second the eigenvalue of the first factor should clearly exceed that of the second factor (Reckase, 1979). The answer to the research Question 1 as revealed in Table 1 and subsequently Figure 1 showed that there was no evidence of unidimensionality. It appears that the Agriculture examination was multidimensional rather than unidimensional.

Even though the total variance was very small, it confirms the findings of Ubi, Joshua and Umoinyang (2012) who stressed that examination design for selections of candidates might not be purely unidimensional, especially when items are fielded from a wide syllabus. Like the aforementioned researchers alluded, agriculture in general is an applied science subject. It has a wide breadth of syllabus. For instance the BGCSE Agriculture multiple choice items are constructed on the contents which range from mathematics, chemistry, physics and biological concepts as reflected in the assessment objectives for the syllabus (Republic of Botswana, 2001). These have contributed to the items to measuring different things and not only one thing. In addition, since the 2013 BGCSE Agriculture was multidimensional, it is not appropriate to analyse it using IRT models. However, it was further subjected to IRT analysis to see if the findings did corroborate the exploratory factor analysis already made.

Mode Fit for One-Parameter, Two-Parameter and Three-Parameter Logistic Model

The result presented in Table 2 for Question 2 showed the level to which the agriculture examination items fit the 1PL, 2PL and 3PL models. The chi-square goodness of fit statistics revealed that none of the items fitted the 1PL model and only one fit the 2PL model. This implied for 1PL and 2PL models, all the 2013 BGCSE Agriculture items were not invariant in measuring what the test was intended to measure except only Item 27 for 2PL model, for which there was a fit. In other words, items for agriculture examination were neither unidimensional nor were they locally independent as

Table 2: Chi-Square Test of Fit in 2013 BGCSE Agriculture Examination using 1PL, 2PL and 3PL IRT Models

ITEMS	1PL Chi-square	df	*P*	2PL Chi-square	df	*p*	3PL Chi-square	df	*p*
1	151.30	9	.000	42.20	9	.000	25.10	9	.003
2	543.30	9	.000	117.90	9	.000	46.10	9	.000
3	70.20	9	.000	35.80	9	.000	45.70	9	.000
4	394.60	9	.000	130.00	9	.000	263.80	9	.000
5	20.50	9	.015	39.50	9	.000	38.10	9	.000
6	114.70	9	.000	76.50	9	.000	14.40	9	.109**
7	207.50	9	.000	122.80	9	.000	14.90	9	.094**
8	150.50	9	.000	80.50	9	.000	117.80	9	.000
9	440.60	9	.000	55.20	9	.000	60.30	9	.000
10	459.30	9	.000	60.70	9	.000	33.90	9	.000
11	204.10	9	.000	18.80	9	.027	21.90	9	.009
12	69.300	9	.000	43.60	9	.000	53.00	9	.000
13	281.90	9	.000	74.40	9	.000	59.40	9	.000
14	83.40	9	.000	25.60	9	.002	12.00	9	.211**
15	74.80	9	.000	76.40	9	.000	17.20	9	.046
16	292.00	9	.000	86.90	9	.000	92.60	9	.000
17	341.80	9	.000	24.30	9	.004	23.90	9	.005
18	581.10	9	.000	158.20	9	.000			
19	25.50	9	.003	22.50	9	.007	37.10	9	.000
20	172.30	8	.000	29.30	8	.000	55.40	8	.000
21	256.70	9	.000	80.00	9	.000	109.40	9	.000
22	30.00	9	.000	32.60	9	.000	10.00	9	.353**
23	114.40	9	.000	56.40	9	.000	23.00	9	.006
24	25.00	9	.003	19.80	9	.020	6.70	9	.664**
25	191.30	9	.000	67.80	9	.000	99.70	9	.000
26	112.40	9	.000	81.70	9	.000	113.10	9	.000
27	68.20	9	.000	8.00	9	.538**	3.90	9	.918**
28	354.70	9	.000	60.90	9	.000	53.60	9	.000
29	146.90	9	.000	55.30	9	.000	46.10	9	.000
30	83.30	9	.000	100.20	9	.000	12.70	9	.174**
31	682.60	9	.000	148.20	9	.000	117.30	9	.000
32	809.30	9	.000	92.50	9	.000			
33	154.30	9	.000	51.90	9	.000	56.70	9	.000
34	122.70	9	.000	67.20	9	.000	12.80	9	.171**
35	137.20	9	.000	34.60	9	.001	36.00	9	.000
36	189.60	8	.000	111.40	9	.000	18.80	9	.027
37	349.50	9	.000	58.10	9	.000	32.20	9	.000
38	90.20	9	.000	36.50	9	.000	39.60	9	.000
39	795.50	9	.000	228.30	8	.000	235.40	9	.000
40	552.10	9	.000	167.80	9	.000	70.90	9	.000

** The item selected with probability great than the alpha level of .05 significant level

confirmed earlier through exploratory factor analysis. Thus, it appeared that through 1PL and 2PL models analysis items in the agriculture national examination multiple choice items were not measuring one and only one trait. This can also be explained through the infringement of local independence because the items performances across the examination were related. Therefore, trait level was not the only influence being measured by agriculture examination (Nenty, 2004). This finding collaborates that of Adedoyin (2006) who found that only two of the 38 mathematics items in BEC's JS final examination in 2004 fitted the 1-PL model, 11fitted the 2- and 16 fitted the 3-PL models.

Despite the unfitness of items on the 1PL and 2PL models, other scholars do consider to opt for other models which are less stringent when exploring model fit of items regarding unidimensional. This is also corroborated by Reckase, (1979) who attested that no single model-data fit index can detect all of the possible sources of fit or misfit. To respond to that the same 40-items were subjected to 3PL model analysis and this revealed that 20 % of the total items fitted and 80 % did

not fit the model. Gruijter and Kamp (2000) suggested that, item(s) that do not fit a chosen model should be dropped from a given instrument or revise for subsequent use. With the 3PL model, only 8 items fit the model and hence given this were appropriate items in measuring students' ability items in agriculture. Even with the use of less stringent model for the 2013 BGCSE Agriculture, the fit analysis results remained unsatisfactory; hence one is tempted to speculate that the 2013 BGCSE Agriculture assessment instrument has a lot to be desired as far as IRT scrutiny is concerned.

In effect, the remaining 32 items represented by 80% of the total items were required to be dropped or revised from the agriculture examination. The unfit items were indicative of bad items and hence not suitable for national examinations unless revised critically to correct their fault. This finding is in line with that of Nworgu and Agah (2012) and Ene (2005) who applied chi-square test with probability greater than the alpha level of .05, selected items fit models they used in their studies respectively.

CONCLUSION AND RECOMMENDATIONS

Through the dimensional analysis of the examination, it was found that the agriculture examination was not unidimensional and very few items fitted IRT Models. This means that during the ability-by-task interaction, there were some demands on some items that provoked behaviour or trait other than that under measurement (agriculture achievement); hence those were a source of multidimensionality.

While effort by African examination bodies in constructing items for public examinations is appreciated, most of such items are not fit for objective and valid measurement of what the test was intended to measure. It is breathtakingly surprising that no item out of forty designed to measure level of knowledge of agriculture actually 'selfishly' measured that knowledge when the influence of extraneous factors, including guessing, was disallowed.

On the basis of the above conclusion we recommend that Botswana Examination Council's (BEC) department of the Directorates of Product Development and Standards should commit itself to constructing items that fit objective and modern measurement models like Rasch, 2- or 3-parameter models and hence reduce unfair and bias measurement through national examinations.

REFERENCES

Adedoyin, O. M. (2006). *Invariance of parameter estimates by classical test theory and item response theory based on 2004 Botswana JSS certificate examination in mathematics.* Doctoral dissertation, University of Botswana.

Boone, J. W., & Scantlebury, K. (2005). The role of Rasch analysis when conducting science educational research utilizing multiple choices test. Wiley InterScience. www.interscience.wiley.com.

Duong, M. (2004). *Introduction to item response theory and its applications*, research development paper. https://www.msu.edu/~dwong/StudentWorkArchive/CEP900F04-RDP/Mihn-Item ResponseTheory.htm

Ene, C. U. (2005) Application of Rasch model in assessing the attitude of students towards biology in senior secondary schools in Enugu Education Zone. Unpublished M.ED Project. Department of Science Education, University of Nigeria, Nsukka.

Gruijter, D, N. & Kemp J. T. C. (2002). *Statistical test theory for education and psychology*. Boston:Kluwer.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139-164.

Hambleton, R.K. (1989). Item response theory. Three parameter-logistic model. *Centre for Study of Evaluation*. Retrieved from http://www.cse.ucla.edu/products/reports/R220.pdf

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, 35, 401-415.

Lord, F.M . (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Mellenbergh, G. J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research*, 29, 223-236.

Nenty, H. J. (2004). From CTT to IRT: An introduction to a desirable transition. In O. A. Afemikhe & J. C. Adewale (Eds.), *Issues in educational measurement and evaluation in Nigeria (in honour of Professor Wole Falayajo)* (pp. 371–383). Ibadan: Institute of Education, University of Ibadan, Nigeria.

Nworgu, B. G., & Agah, J.J. (2012). Application of three-parameter logistic model in the calibration of a mathematics achievement test. *Journal of Educational Assessment in Africa*, 7, 162 - 172.

Nworgu, B. G. (2011). Differential item functioning: A critical issue in regional quality assurance. *Journal of Educational Assessment in Africa*, 6, 112-123.

Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207-230.

Republic of Botswana. Ministry of Education (2013). *Request for quotation for the evaluation of declining*

results in basic education sector (primary, junior and senior secondary) since 2007 to date- 2013. Gaborone, Botswana: Government printer.

Royal, K. D., & Puffer, J.C. (2013).Dimensionality of the maintenance of certification for family physicians examination: Evidence of construct validity. *Am Board Fam Med*, 11(3), 286-288.

Siamisang, F. T., & Nenty, H. J. (2012). Analysis of gender-based differential item functioning (dif) in 2007 TIMSS examination among students from Botswana, Singapore and USA. *Journal of Educational Assessment in Africa,* 7, 043-054

Thobega, M., & Masole, T. M. (2008). Predicting students performance on agricultural science examination from forecast grades, *US-China Education Review*, 5 (10), 45-52.

Ubi, I. O., Joshua, M. T., & Umoinyang, I. E. (2012) Assessment of dimensionality of mathematics tests of university matriculation examination in Nigeria: Implications for regional development. *Journal of Educational Assessment in Africa,* 7, 122- 130

Umoinyang, I.E. (2011). The challenge of removing consistent errors in achieving tests using differential item functioning (DIF) detection methods. *Journal of Educational Assessment in Africa,* 6, 120-132

Vedle, G. V., Beaton, D., Hogg-Johnston, S., Hurwitz, E & Tennant, A. (2009) Rasch analysis provides new insights into the measurement properties of the neck disability index*, American College of Rheumatogy,* 61 (4), 544 - 551.