British Journal of Education
Vol.7, No.4, pp.74-83, April 2019
Published by European Centre for Research Training and Development UK (www.eajournals.org)

INVESTIGATION OF NATIONAL BUSINESS AND TECHNICAL EXAMINATION BOARD (NABTEB) 2013 TO 2015 CHEMISTRY MULTIPLE CHOICE TEST ITEMS' PARAMETER INVARIANCE

Oguguo, Basil C.E
Department of Science Education, Faculty of Education, University of Nigeria, Nsukka

Regina Awele LOTOBI
Department of Educational Evaluation and Counseling Psychology
University of Benin City, Nigeria

ABSTRACT: *The purpose of the study was to analyze NABTEB 2013, 2014 and 2015 Chemistry multiple choice test items for item parameter invariance using item response theory. An evaluation research design was adopted for the study. Unidimensionality test was conducted to see to the IRT statistical assumption being satisfied before further analysis was carried out. The IRT statistical software -eirt - Item Response Theory Assistant for excel was used to determine the item difficulty parameter, item discrimination parameter and guessing parameter estimates. The findings from the study revealed that item difficulty and discrimination parameters were only invariant by sex in 2013 and 2015 but guessing parameters were invariant by sex in 2013, 2014 and 2015. Item difficulty parameters were invariant by school location in 2013, 2014 and 2015 while item discrimination and guessing parameters were invariant by school location in 2014 and 2015 only. Item difficulty and guessing parameters were invariant by school ownership in 2013, 2014 and 2015 but item discrimination parameter was invariant by school ownership in 2013 and 2015. Item difficulty and guessing parameters were invariant by school type in 2013 but item discrimination parameter was invariant by school type in 2013 and 2015. It was recommended that examination bodies such as NABTEB should use any of the subgroups within the population for item analysis provided such subgroups include respondents of the different levels of ability.*

KEYWORDS: Investigation, unidimentionality, Item parameter invariance, NABTEB chemistry multiple choice test items, IRT

INTRODUCTION

Chemistry as a school subject is a requirement for the study of science related courses in institutions of higher learning. The role of Chemistry towards national development is highly significant in the school curriculum. Chemistry develops scientific literacy and help to build up necessary knowledge and skills required for lifelong learning in science and technology. According to Ruth (2012) students' lives are enriched with these knowledge, skills and attitudes to become scientifically capable members of society. The knowledge gained further provides a fanlight on what it takes to be a scientific researcher. Many people through their careers and day-to-day pursuits of chemistry develop an applied knowledge of chemistry (Ruth, 2012).

Print ISSN: ISSN 2054-6351 (print), Online ISSN: ISSN 2054-636X (online)

The awareness Chemistry creates makes the subject interesting to scientists while carrying out their work. The acquired knowledge serves as guide in developing capacities that will enable students to assess the strengths and limitations of science. Chemistry curriculum attempts to make the study of chemistry exciting and relevant. The aim is to enable students to carry out handy investigations, examine data, report evidence, and present conclusions. Chemistry becomes relevant in education as it enable the demonstration of the students' capability level and understanding in different areas of science and technology. Science students are expected to offer and pass it at credit level before they can proceed to study any science course at the university level (Kolawole, Oginni & Fayomi, 2011) or else be denied admission due to inability to meet the basic entry requirements.

The National Business and Technical Examinations Board (NABTEB) is a public examination body established under Act 70 of 1993. The body is primarily concerned with the conduct of quality tests for evaluating student's learning, promotion, placement and selection. Test quality is determined by its psychometric properties which include item difficulty, item discrimination and guessing. These are the parameters employed during measurement. The responses to the items are essential for analyzing these psychometric properties. These items are generated to check how capable those being examined can perform in a test once effective teaching has taken place. The knowledge gained is expected to reflect in the examinees' response to the test items presented before them for examination purpose. The properties contained in the item determine to a large extent the trait in every examinee. The implication of this is that the examinee gets the item right through relying on the examinee's ability and the characteristics of the item. The examinee's estimated ability level and item difficulty level are put on the same scale allowing for the illustration of item difficulty and item discrimination simultaneously using the Item Characteristic Curve graph to depict the characteristics of each item. This approach is conceptualized in Item Response Theory (IRT) which is used in measurement.

The Classical Test Theory (CTT) is the foundation of modern testing upon which IRT builds on. The CTT which most examination bodies in Nigeria and Africa use provides item difficulty and item discrimination parameters only in the item selection process.All along, and even up till now, it's like many of the examination bodies are still interested in the CTT. According to literature, CTT lacks some vital psychometric properties. One of them is item parameter invariance. It is believed that examination properties have been determined before the advent of IRT using CTT. Item difficulty and item discrimination alone can no longer be the basis of item selection as other psychometric properties are now being considered: more test item parameters such as the issue of item invariance, guessing parameter, local item independence, differential item functioning, item parameter drift, test equating have become needed parameters in order to determine the quality of test items. One of these psychometric properties that are of interest is the item parameter invariance.

Item response theory is a modern testing theory that studies test and items scores based on assumptions concerning the mathematical relationship between abilities and item responses. IRT is also a general statistical theory about examinee item and test performance and how performance relates to the abilities that are measured by the items in the test. Within IRT context, are parameters which determine the function of items and the ability of the examinees. An examinee's ability is a function of the item parameters. In item response theory, the probability that an examinee will produce a correct response to an item is dependent on the ability of the examinee and the item parameters.Item Parameter (IP) means those parameters that interpret the quality of the items such as item difficulty parameter (b), item

Print ISSN: ISSN 2054-6351 (print), Online ISSN: ISSN 2054-636X (online)

discrimination parameter (a), and guessing parameter (c). Item difficulty parameter refers to an examinee having attained adequate ability level that can equate the mean to which examinees are likely going to get a particular item right. Item discrimination parameter determines how items can distinguish between highly proficient examinees and less proficient examinees. Item discrimination parameter takes full advantage of items that exhibit high discrimination power in the selection of items, while the guessing parameter depicts the likelihood that an examinee with low ability performs a difficult task above his actual capability. When guessing parameter is considered, it reveals that in many items, no matter the impact of difficulty, the examinees still have some probability greater than zero to endorse the item right.Every item usually possesses features of difficulty, discrimination and chance of a guess. The chance of a guess will be higher because it favours the low ability group. It is obvious that the properties of the items with their discrimination power and difficulty level sort out the examinees in their respective ability levels. Guessing becomes an option for those who cannot cope with the impact of difficulty and discrimination. The purpose of testing is to discriminate examinees with high ability and examinees with low ability. An examinee with high ability tends to answer an effective item correctly while examinee with low ability tends to answer it incorrectly. Faced with that type of situation, the impact of difficulty and discrimination are felt depending on the nature of the task. The impact of the discrimination and difficulty is to sort examinees of the two groups into bright and lower ability at the extremes of the distribution. For those that are with lower ability, the fact that they cannot contend the challenge of discrimination and difficulty, their only way out of the situation will be to guess. The examinees, who are likely to attempt the task before them by virtue of trying to attempt and not by guessing, will be the bright ones. Whether the examinee actually opts for guessing or attempt depends on the impact of difficulty, discrimination and guessing. Naturally, those that are discriminated into the groups of higher and average ability are those that are not likely to favour guessing. It is expected that when examinees are subjected to tests, certain things happen. The items will discriminate based on the item difficulty.

Item Parameter Invariance (IPI) implies that the value of item parameter ought to be alike irrespective of the group that is chosen (i.e., subgroups). IPI is a desirable characteristic of a statistical model that implies identical statistical decisions made across different measurement decisions. The strictness in item parameter invariance represents a state that is errorless and nearly not achievable in practice (Rupp & Zumbo, 2004). IPI is an abstract ideal state because the parameters are either invariant or not. If a parameter varies across a different measurement conditions, then a lack of invariance is present. An invariance assumption is that the item parameters are not subjective by the sample characteristics.Item parameters are only accepted to be invariant when its parameter estimates and statistics do not vary across samples. This is the reason for IRT's superior advantage. Item parameter invariance depends completely upon the closeness of fit connecting a set of test data and the item response model that is fit to it (Hambleton & Swaminathan, 1985).In theory, when an item response model fits the data, it indicates item parameters are independent of the abilities of respondents (Hambleton, Swaminathan, and Rogers, 1991).IRT assumptions must be equally met by items which if not met; the items definitely cannot be imbued with invariance property by the parameter estimation method used. According to the invariance property of item response theory (IRT), item parameter values should be the same for all samples from a population. In practice, however, it is not always possible to satisfy the invariance property. Research has found that item parameters can change for different subgroups of examinees and across different testing occasions. The absence of the invariance property of item parameters could render ability scores not comparable which implies violation of the invariance property. Any violation made about this property of invariance would put at risk model parameter estimation, person ability scoring, and interpretation.

Print ISSN: ISSN 2054-6351 (print), Online ISSN: ISSN 2054-636X (online)

Change in parameter values for different subgroups is called differential item functioning. Change across time is called item parameter drift. Item Parameter Drift (IPD) or differential change in item parameters over time is expected to occur if invariance does not hold. IPD is a change that is not expected to occur within the estimated parameter values existing between testing forms or between any testing occasions. Item parameter drift (IPD) can as well be exhibited when digression occurs from the true value in item parameters. If parameter values fluctuate more than would be expected due to measurement error alone, then it can no longer be assumed that parameter values are invariant over testing occasions.

The National Business and Technical Examinations Board is a public examination body in Nigeria. The board conducts different examinations and issues results and various certificates. These examinations are in series. The in-school May/June examination series (National Business Certificate (NBC)/ National Technical Certificate (NTC) and the private candidates` November examination series (National Business Certificate (NBC)/ National Technical Certificate (NTC)) of which Chemistry is a subject examined. Chemistry as a school subject matter is of immense importance to education and for the positive advancement of society. However, researchers (Agbodeka, 2002; Adeyemi, 2011 and Ruth, 2012) points to insufficient performance in Chemistry in school curriculum. According to Agbodeka (2002) the performance of students has remained low and unimpressive despite the importance of Chemistry as a discipline.
Introducing item parameter invariance in our educational system is to put to test whether the items generated and selected by examination bodies satisfy the conditions for item parameter invariance using IRT. In this study, the researchers analyzed NABTEB 2013 to 2015 Chemistry multiple choice test items for item parameter invariance using item response theory to see test items that are not coinciding through these known parameters. The purpose is to find out whether Item Response Theory, item parameter invariance property is stable over the 3 years across different groups. In aggregate the study should highlight the significance of Chemistry in addition to the primary purpose for the study. The subgroups of population (sex, school location, school ownership and school type) are attended by final year business and technical students and were used to determine IRT, Item Parameter Invariance estimates. Candidates from these different groups of population write National Business and Technical Examinations Board Examination each year. The certificate awarded is based on scored responses of the candidate. Knowledge of Chemistry test items were sampled in each of these examinations in each year.

Hence, this study looked at testing for item parameter invariance. In the past, a framework which had been used is classical test theory (CTT). The National Business and Technical Examinations Board (NABTEB) do standardization of items using Classical Test Theory.But there has been a criticism against this classical test theory due to IRT's theoretical advantage. The property of item parameter invariance is one of main theoretical advantage IRT has over CTT. IRT provides invariance properties of test items. Even though NABTEB generated and selected their items with CTT in the past, this is an attempt to see whether those items administered in their standardized form by NABTEB satisfy item parameter invariance using Item Response Theory. Therefore, the question addressed is: Would the NABTEB test items satisfy the item parameter invariance property of IRT across sex, school location, school ownership and school type samples?

The main purpose of the study was to ascertain if NABTEB 2013 to 2015 Chemistry multiple choice test items satisfy item parameter invariance using Item Response Theory. Specifically, an attempt was made to establish whether the 2013, 2014 and 2015 NABTEB Chemistry multiple choice test items are not

Print ISSN: ISSN 2054-6351 (print), Online ISSN: ISSN 2054-636X (online)

significantly invariant by sex, school location, school ownership and school type in the difficulty, discrimination and guessing parameters.

The following research questions were raised to guide the study:

1. What proportion of the NABTEB's 2013, 2014, and 2015 Chemistry multiple choice test items difficulty, discrimination and guessing parameters are invariant with respect to gender?
2. What proportion of the NABTEB's 2013, 2014, and 2015 Chemistry multiple choice test items are invariant with respect to school location in the difficulty, discrimination and guessing parameters?
3. What proportion of the NABTEB 2013, 2014, and 2015 Chemistry multiple choice test items are invariant with respect to school ownership in the difficulty, discrimination and guessing parameters?
4. What proportion of the NABTEB's 2013, 2014 and 2015 Chemistry multiple choice test items are invariant with respect to school type in the difficulty, discrimination and guessing parameters?

MATERIALS AND METHOD

An evaluation research design was adopted for the study. The population of the study comprised the NABTEB 2013 to 2015 Chemistry multiple choice test items (MCTI) numbering 1 – 50 and SSS year III students offering Chemistry. According to the statistics from NABTEB statistical unit at the time of this study, there were 127,585, NABTEB Chemistry students in Nigeria for the three years under study. These students consisted of males and females, mixed and single schools, urban and rural schools, private and public schools. The following data show the total population per year; from NABTEB: May/June 2013: 54,085; May/June 2014: 39,523 and May/June 2015: 33,977 in Nigeria. The statistical population of items for this study is 150 items (50 items for each year: 2013, 2014 and 2015) for the three years of study. The sample for this study consisted of the entire students who sat for the NABTEB 2013, 2014 and 2015 May/June Chemistry multiple choice examinations numbering 127,585. The focus of item analysis in this study was according to the statistical population and statistical sample. The 50 items for each of the years: 2013, 2014 and 2015 were used in the course of the analysis.

The three research instruments used in this study were obtained from National Business and Technical Examinations Board (NABTEB) May/June 2013 – 2015 Chemistry Multiple Choice Objective Test Items. The instruments contain 50 items each and were adopted whole and entire without modifications. The soft copies of candidates' responses as well as their demographic information were obtained from the examination body. The electronic copy of the dichotomously scored responses (1 for correct, 0 for wrong answer) was obtained from NABTEB in the form of a spreadsheet Microsoft Excel showing the examinee by items matrix for each of the subgroups (Male and female, urban and rural, mixed and single, private and public). The validity of the instrument was established by the NABTEB Board. Hence the instrument contained standardized items by originality. The Board has the credit and merit of adopting different approaches of determining validity. As such, the items were appropriate in terms of subject content and instructional objectives. The researchers therefore adopted the instrument as valid and appropriate for this study. The test items were considered to be reliable being a standardized test. The data for this study were collected by the researchers. The desired data on the candidates' responses; item by item matrix, the candidates' sex, school location, school ownership and school type collected through Microsoft Excel sheet by NABTEB examination unit, were duly obtained by the researchers. Unidimensionality assumption of IRT was determined for the 2013 – 2015 Chemistry multiple choice items. Principal Component Extraction Method (PCEM) was used to determine the unidimensionality of the test items. This was done by analyzing item responses from examinees by separately establishing dimensionality for

Print ISSN: ISSN 2054-6351 (print), Online ISSN: ISSN 2054-636X (online)

each group; sex (male and female), school location (urban and rural), school ownership (private and public) and school type (mixed, all boys and all girls). Eigen values were determined to identify the number of dominant factors that existed among the test items. The formula of the psychometric properties was entered into the Microsoft Excel data sheet for computation. Item parameters were estimated with IRT statistical software known as IRT Assistant for Excel (EIRT). The formulated hypotheses 1 to 4 were tested. Test of invariance was then carried out using Statistical Package for Social Sciences (SPSS) version 21. In evaluating Item Parameter Invariance, the item difficulty, item discrimination and guessing parameters were expected to be statistically identical across the different groups and different populations after being subjected to more statistical analysis. This is to find out if NABTEB 2013, 2014 and 2015 Chemistry multiple choice test items were not significantly invariant by sex, school location, school ownership and school type in difficulty, discrimination and guessing parameters. The stated hypotheses were subjected to paired sample t-test statistics at an alpha level of 0.05 significance respectively. A t-value that was not statistically significant indicated the parameters being questioned were invariant across subgroups but when it was a significantly statistical t-value, it indicated that the parameters being questioned lacked invariance across groups.

RESULTS

Unidimensionality of the test items were determined as required for the further computations in the Item Response Theory. Using the Principal Component Analysis for the different subgroups, the existence of a dominant factor provided by the first factor as well as the noticeable drop in eigenvalues between the first and second factor shows the test satisfies unidimensionality. The determination of unidimentionality of the various subgroups are as follows:

The subgroup of male 2013 yielded the first eigenvalue of 13.587, second eigenvalue of 5.266 and third eigenvalue of 2.819. The subgroup of female 2013 yielded the first eigenvalue of 8.578, second eigenvalue of 3.755 and third eigenvalue of 3.436. The subgroup of urban 2013 yielded the first eigenvalue of 13.231, second eigenvalue of 4.412 and the third eigenvalue of 2.671. The subgroup of rural 2013 yielded the first eigenvalue of 7.082, second eigenvalue of 3.764 and third eigenvalue of 2.848. The subgroup of private 2013 yielded the first eigenvalue value of 8.661, second eigenvalue of 4.497 and third eigenvalue of 2.751. The subgroup of public 2013 yielded the first eigenvalue of 7.443, second eigenvalue of 3.383 and third eigenvalue of 2.640. The subgroup of mixed 2013 yielded the first eigenvalue of 12.933, second eigenvalue of 4.389 and third eigenvalue of 2.699. The subgroup of all girls 2013 yielded the first eigenvalue of 9.512, second eigenvalue of 5.321 and third eigenvalue of 4.261. The subgroup of all boys 2013 yielded the first eigenvalues 8.861, second eigenvalue of 6.723 and third eigenvalue of 4.415.

The subgroup of male 2014 yielded the first eigenvalue of 11.469, second eigenvalue of 4.316 and third eigenvalue of 3.146. The subgroup of female 2014 yielded the first eigenvalue of 7.639, second eigenvalue of 4.281 and third eigenvalue of 3.073. The subgroup of urban 2014 yielded the first eigenvalue of 10.279, second eigenvalue of 4.642 and the third eigenvalue of 3.062. The subgroup of rural 2014 yielded the first eigenvalue of 9.020, the second eigenvalue 3.645 and the third eigenvalue of 2.992. The subgroup of private 2014 yielded the first eigenvalue of 6.983, second eigenvalue of 4.185 and the third eigenvalue of 3.337. The subgroup of public 2014 yielded the first eigenvalue of 8.844, second eigenvalue of 3.173 and the third eigenvalue of 3.134. The subgroup of mixed 2014 yielded the first eigenvalue of 9.690, second eigenvalue of 4.446 and the third eigenvalue of 2.624. The subgroup of all boys 2014 yielded the first

Print ISSN: ISSN 2054-6351 (print), Online ISSN: ISSN 2054-636X (online)

eigenvalue of 6.660, second eigenvalue of 5.615 and third eigenvalue yielded 3.627. The subgroup of all girls 2014 yielded the first eigenvalue of 8.493, the second eigenvalue of 7.419 and the third eigenvalue of 5.191.

The subgroup of male 2015 yielded the first eigenvalue of 10.581, second eigenvalue of 3.219 and the third eigenvalue 2.298. The subgroup of female 2015 yielded the first eigenvalue of 8.174, second eigenvalue of 4.325 and third eigenvalue of 3.753. The subgroup of urban 2015 yielded the first eigenvalue of 8.283, second eigenvalue of 3.238 and the third eigenvalue of 2.902. The subgroup of rural 2015 yielded the first eigenvalue of 7.532, second eigenvalue of 4.354 and the third eigenvalue of 3.331. The subgroup of private 2015 yielded the first eigenvalue of 7.845, the second eigenvalue of 3.265 and the third eigenvalue of 3.106. The subgroup of public 2015 yielded the first eigenvalue of 7.485, second eigenvalue of 5.062 and third eigenvalue of 3.698. The subgroup of mixed 2015 yielded the first eigenvalue of 9.197, the second eigenvalue 3.414 and the third eigenvalue of 2.911. The subgroup of all girls 2015 yielded the first eigenvalue of 16.293, the second eigenvalue of 7.666 and the third eigenvalue of 2.774. The subgroup of all boys 2015 yielded the first eigenvalue of 6.419, second eigenvalue of 5.393 and third eigenvalue of 3.509.

In the above reports, the condition of unidimensionality was satisfied for NABTEB Chemistry multiple choice test items in most of the 2013, 2014 and 2015 and was therefore fit for the Item Response Theory analysis, except the subgroup of all boys 2013, 2014, 2015 and all girls 2014 did not satisfy the assumption of unidimensionality conditions and was not fit for further Item Response Theory analysis.

DISCUSSION OF FINDINGS

The finding revealed that the 2013 NABTEB Chemistry multiple choice tests were invariant by gender in the difficulty, discrimination and guessing parameters. The finding also revealed that the 2014 NABTEB Chemistry multiple choice tests were not invariant in discrimination and difficulty parameters but were invariant in the guessing parameter. The 2015 NABTEB Chemistry multiple choice tests were invariant with respect to gender in difficulty parameter, discrimination parameter and guessing parameter. This finding is in line with the findings of MacDonald and Paunonen (2002) who investigated the invariance of item parameters under various conditions using the 1PL- IRT and 2PL- IRT models. The finding of the present study agrees with the result of Wiberg (2004) who examined parameter invariance and found out that invariance was fulfilled in only item difficulty parameter in 1PLM but not in 2PLM as well as 3PLM and guessing parameter was invariant in 3PLM. On the contrary, Galdin and Laurencelle (2010) found out that no invariance occurred with the parameters estimated across a change of estimation situation.

The 2013 NABTEB Chemistry multiple choice tests items were invariant by school location in the difficulty parameter but not invariant in discrimination and guessing parameters. The study also revealed that the 2014 and 2015 NABTEB Chemistry multiple choice tests were invariant with respect to school location in the difficulty, discrimination and guessing parameters. This finding is in line with the result of Mallikarjuna and Natarajan (2012) who utilized real data to test for parameter invariance. The result showed invariant property of item response theory. The finding is equally supported by the finding of Adedoyin (2008) who tested the invariance of each item parameter across different samples of examinees. Their finding showed that the guessing parameters are invariant across the examinees with the IRT framework. In the same vein, the finding contradicts the finding of Nenty and Adedoyin (2013) who

Print ISSN: ISSN 2054-6351 (print), Online ISSN: ISSN 2054-636X (online)

conducted a study testing for significance of the invariance between the item parameter estimates across models.

The 2013 and 2015 NABTEB Chemistry multiple choice tests items were invariant with respect to school ownership in the difficulty, discrimination and guessing parameters. The study also revealed that 2014 NABTEB Chemistry multiple choice tests items were invariant by school ownership in the difficulty and guessing parameters but not invariant in discrimination parameter. This finding is supported by the finding of Curtin (2007) whose dissertation investigated three methods for assessing parameter invariance of item difficulty parameter. The finding proved that difficulty parameter estimates were invariant over measurement occasions when the rating scale model was appropriate for the data. The finding differs in the direction of Progar and Socan (2008) who conducted item and person parameters invariance. The finding showed parameter invariance in the item. However, this finding contradicts Fan (1998) who empirically tested parameter invariance for IRT model. The result revealed that item discrimination parameters were less invariant with greater differences in the score distribution.

The finding of the study revealed that 2013 NABTEB Chemistry multiple choice tests items were invariant by school type in the difficulty, discrimination and guessing parameters. The 2015 NABTEB Chemistry multiple choice tests item were invariant with respect to school in the discrimination parameter but not invariant in difficulty and guessing parameters. The finding corroborates the result of Weiss and Yoes (1991) whose study compared the rank order of calibrated item difficulty estimates across different groups of examinees and found that groups drawn from the same examinee pool can be used to evaluate the degree of item parameter invariance for a set of unidimensional items. Also, the finding agrees with the result of Fan and Ping (1999) who conducted a study to examine item parameter invariance. The result showed a high degree of item parameter invariance. This finding is not in agreement with the study of Dogan and Yurtcu (2015) who investigated the invariability of item parameters using real data. Their finding revealed parameter invariability in the item parameters.

Research Implication

This study contributed to knowledge in test psychometrics in that it has confirmed that the items of NABTEB satisfied Item Parameter Invariance even though they have not been using IRT in the past years for the standardization of their examination items. The main purpose of the study was to analyze National Business and Technical Examinations Board, Chemistry multiple choice test items for item parameter invariance using Item Response Theory. So Examination bodies should take into their account the effect of using Item Response Theory in analyzing their items.

CONCLUSION

Based on the findings of the study, the following conclusions were made

The NABTEB 2013, 2014 and 2015 item parameter invariance by sex is satisfied for item difficulty in 2013 and 2015, satisfied for item discrimination in 2013 and 2015 and satisfied for guessing in 2013, 2014 and 2015. Also, the NABTEB 2013, 2014 and 2015 item parameter invariance by school location is satisfied for item difficulty in 2013, 2014 and 2015, satisfied for item discrimination in 2014 and 2015, satisfied for guessing in 2014 and 2015. Furthermore, the NABTEB 2013, 2014 and 2015 item parameter invariance by school ownership is satisfied for item difficulty in 2013, 2014 and 2015, satisfied for

Print ISSN: ISSN 2054-6351 (print), Online ISSN: ISSN 2054-636X (online)

discrimination in 2013 and 2015, satisfied for guessing in 2013, 2014 and 2015. It also shows, the NABTEB 2013, 2014 and 2015 item parameter invariance by school type is satisfied for item difficulty in 2013, satisfied for item discrimination in 2013 and 2015, satisfied for guessing in 2013. Therefore, the item invariance property of item response theory is true as confirmed by the findings of this study.

REFERENCES

Adedoyin, O. O (2008). Investigating the invariance of item difficulty parameter estimates based on CTT and IRT. *Educational Research and Review,* 3(2): 83-93.

Adeyemi, T. O. (2011). A Comparative Study of Students' Academic Performance in Public Examinations in Secondary Schools in Ondo and Ekiti States, Nigeria. *Current Research Journal of Economic Theory* 3(2): 36-42.

Agbodeka, F. (2002). *The West African Examinations Council (1952 – 2002). Half A Century of Commitment to Excellence and Regional Co-operation*. Accra, Ghana: Woeli Publishing Services.

Curtin, J. A. (2007). Testing the assumptions of sample invariance of item difficulty parameters in Rasch rating scale model. *Dissertation Abstracts International,* (UMI No. 3281484).

Dogan, N. & Yurtcu, M. (2015). Invariability of Item Parameters in The Sample with Restricted Real Data Results. *Elsevier; Procedia - Social and Behavioral Sciences,*197: 2207 – 2214.

Fan, X. (1998). Item response theory and classical test theory: An Empirical Comparison of their Item/Person Statistics. *Educational and Psychological Measurement*, 58(3): 357-381.

Fan, X., & Ping, Y. (1999). *Assessing the Effect of Model-Data Misfit on the Invariance Property of IRT Parameter Estimates*. Paper presented at the annual meeting of the American Educational Association, Montreal, Canada.

Galdin, M. & Laurencelle, L. (2010). Assessing Parameter Invariance in Item Response Theory`s logistic two parameter models: A Monte Carlo investigation: *Tutorials in Quantitative Methods for Psychology*. 6(2): 39-51.

Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory principles and applications*. Boston: Kluwer- Nijhoff.

Hambleton, R. K., Swaminathan, H. & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*: Newbury Park, CA: Sage Press.

Kolawole, E. B., Oginni, O. I., & Fayomi, E. O. (2011). An Ordinary Level as Predictors of Students‟ Academic Performance in Chemistry in Nigerian Universities. *Educational Research and Reviews* 6(14): 824-827.

Mallikarjuna, G. & Natarajan, V. (2012). Investigate into Invariance Properties of Item Response Theory (IRT) By Two and Three Parameter Models: *International Journal of Information Technology and Business Management.*1 (1): 28-34.

MacDonald, P., &Paunonen, S. V. (2002). A Monte Carlo Comparison of item and person parameters based on item response theory versus classical test theory. *Educational and Psychological Measurement* 62: 921-943.

Nenty, H. & Adedoyin, O. O. (2013). Test for invariance: inter and intra model validation of classical test and item response theories. *Asia Pacific Journal of Research*, I(IX): 51-66.

Progar, S. & Socan, G. (2008). An Empirical Comparison of Item Response Theory and Classical Test Theory. *Psiholoskaobzorja / Horizons of Psychology,* 17(3): 5-24.

Print ISSN: ISSN 2054-6351 (print), Online ISSN: ISSN 2054-636X (online)

Rupp, A. A. & Zumbo, B. D. (2004). A note on how to quantify and report whether IRT Item parameter invariance holds: When Pearson correlations are not enough. *Educational and Psychological Measurement*, 64(4): 588-599.

Ruth, L. (2012). Chemistry Accreditation period 2013–2016. Developed and published by the Victorian Curriculum and Assessment Authority. *Chemistry*, pp.94-95.

Weiss, D. J. &Yoes, M. E. (1991). Item Response Theory in R. K. Hambleton and J. N. Zeal (eds.), *Advances in educational and psychological testing*. Boston, MA: Kluwer Academic Publishers, pp. 69-96.

Wiberg, M. (2004). Classical test theory vs Item response theory: An evaluation of the theory test in the Swedish driving-license test: Umea University, Sweden.

Print ISSN: ISSN 2054-6351 (print), Online ISSN: ISSN 2054-636X (online)