Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.5, No.9, 2014

Bias in Test Items and Implication for National Development

Caroline Ochuko Alordiah*,Helena T, Agbajor
Department of Educational Psychology, College of Education Warri. Delta State, Nigeria
*E-mail of the corresponding author: carolinealordiah@gmail.com

Abstract
A test that has bias items will not be free from bias which will in turn affect the validity of the test. Tests permit us to make accurate inferences about the knowledge and skills students possess. This can only be possible if the test is valid. The purpose of the study is to identify and eliminate bias items from our tests. Therefore, the study presents the meaning of item bias, the sources, types and ways to detect and eradicate bias from our tests. It was recommended that before any test is put into public use, every effort should be made to detect and do away with biased items.
Keywords: National development, Item bias, Test bias, Differential item functioning and Mantel-Haenszel chi-square.

1. Introduction

National development can be described as the overall attempts to improve the conditions of human existence which involves socio-economic, political, educational as well as religious advancement of a country. Many development plans have been developed in Nigeria, none has been able to yield the development the country is yearning for. Many reasons have been given but one major barrier to the success of these national development plans is the ills in the education sector. The development of any country can only take place if the citizen is well educated. This will help them to develop themselves and the nation at large. When this is done, the living standard of the people will improve and this will foster national unity. The major challenge facing the Nigerian educational system is its ability to meet the needs of the society. An educational system that is geared towards improving and developing the citizenry should be that, which would make sure that all aspects of the educational system are working well. From the implementation stage to the evaluation stage must be geared towards producing an individual that will be able to adapt to his environment and in turn improve the environment. Education consists of three major components: Input (learners), process (learning) and output. The quality of the educational system depends to a great extent on the qualities of these three major components (Alordiah, 2012).
The output has to do with measurement and evaluation. When measurement and evaluation is not properly done, it can affect the humans who come out of the educational system. Test is one of the measurement instruments that are commonly used to assess learners' abilities in all the levels of our educational system, ranging from primary to tertiary institutions.Test is the most widely used measurement instrument in Nigeria. However, the most serious criticism against it is that it sometime show bias against a group of the examinees taking the test. Test bias in measurement has become a heated complex and pronounced issue in the western countries and most developing countries are also becoming conscious of the concept (Joshua, 2005).
The national examinations conducted by West African Examination council (WAEC). National Business and Technical Education Board (NABTEB), National Examination Council (NECO), and Joint Admissions and Matriculation Board (JAMB) cater for candidates from various backgrounds all over the country. This is so because Nigeria as a nation is a heterogeneous state. According to the Federal government in the national policy on education (2004) every Nigerian child should have a right to equal educational opportunities irrespective of any real or imagined disabilities, each according to his or her ability and there shall be the provision of equal access to educational opportunities for all citizens of the country at the primary, secondary and tertiary levels both inside and outside the formal school system. The international Labour Organization (ILO, 2004) R195 Human Resources Development Recommendation of 2004 articles 19(f) and (g) insist on identifying and overcoming barriers to accessing training and education. It requires that for equity in human resources development to be attained, bias in the assessment of competencies must be identified and eliminated. The implication of this policy and recommendation to all educationists and most especially to the evaluators is that items in test should be unbiased to all subgroups in the population. Emaikwu (2012) reported that it has been claimed that some of the national examinations unfairly favour examinees of some particular groups e.g., cultural or linguistic groups to the extent that it is believed that a particular section of the country perform most woefully in these national examinations. This can hinder national development. The purpose of the study is to identify and eliminate bias items from our test. Therefore, the study presents the meaning of item bias, the sources, types and ways to identity and eradicates them from our test.

2.The Meaning and Description of Item Bias

An item is biased if its construction, setting, language, idea or interest portrayed, picture/diagram used, relevance,

illustrations, and administration give an undue advantages or disadvantage to a particular group of testees over the other group (Nenty, 2010). Item bias is defined as the characteristic of an item which causes learners on the same ability levels to perform differently in the test item because they are from different groups (e.g. gender, race, ethnicity, religion, culture, disability or social class). For instance, supposing you had a 30-item mathematics test, if you select one item from the test and found that students from urban schools responses were similar to that of students from rural schools that are of the same latent ability in mathematics, then the item is unbiased. On the other hand, if students from urban schools and students from rural schools with the same latent ability in mathematics responded in different ways to the item, such item is said to be biased.

The term test bias and item bias are not the same but are related. When an item is picked out to be bias, such item is referred to as item bias but when a test has bias items, such test is said to be test bias. Similarly, the term differential item functioning (DIF) is closely related to item bias. Atar (2006) says two meanings can be load to the term item bias: a judgmental meaning and a statistical meaning. While the term item bias indicates that an item provides an unfair advantage to some group of examinees in its judgmental meaning; it indicate that an item exhibits a statistical difference between groups of examinees with comparable ability level in its statistical meanings. Since the item bias leads to confusion in its judgmental and statistical meanings, the term DIF is preferred over the term item bias to imply statistical difference. An item must show DIF before it can be said to be bias. Nevertheless, an item that shows DIF does not necessarily mean it is bias. Yet, all items that show bias are DIF items. DIF help us to sort out items that maybe unfair which are then subjected to further investigation to find out whether they are bias or not (Alordiah, 2013).

A test item is said to be item bias if:There is differential performance for individuals of the same ability but from different groups; It lowers the average score of a particular group; It contains language or content that is differentially familiar for different subgroups of the examinees; It contains sources of difficulty that are irrelevant or extraneous to the construct being tested; The test item, item stem, test instruction or distractor is not good enough or/and can be understood in more than one way by the examinees; Contain clues that would increase the performance of one group over another; There are no equal learning opportunities so much that one group is more exposed to the information being tested than the other group; There are no equal access to relevant textbooks equipment, instruments, laboratories and workshops; There is no equal scoring format for the test takers; It contains offensive elements that would insult any group of examinees on the basis of their personal characteristics.Some of the sources of item bias are inadequate item formulation, language, test wiseness, poor item translation, when an item is invoking additional traits or abilities, when the topic of the test items are not in the curriculum of one of the subgroup, etc.

2.1Effects of Item Bias

A test that is offensive is likely to affect the performance of the students. Take for instance a test item that is constructed thus: “Reverend father EmekaChuku was the man involved with the eating of human being in Ogun state”. This is a biased item. It stands to reason that some catholic students may be offended, on the ground that a reverend father eats human being. Also candidates who are Ibos maybe offended because an Ibo man is mention as an eater of human being. In addition those candidates who have relative in Ogun state may be worried over the safety of their relatives. These various subgroups of the examinees may be disturb emotionally and may pay undue attention to the item which would in turn affect their scores in that item and in subsequent items.

When an item requires mastery of an irrelevant skill to the one that is been tested, it may lead to disparity in performance. For example a mathematics test item that requires a high proficiency in English Language could be biased against students who have the high proficiency in English Language that the item demands. Test item bias also affects the psychometric properties of measurement results from such test. The validity and reliability of such test are greatly affected.

Testing and test scores has very important consequences for people in Nigeria. It is through test that people are promoted, selected for various jobs, placed in various institution/establishment, given awards, scholarships and appointment into various positions ranging from the education sector, economic sector, social sectors, government sector to the political sector. Due to these importance’s of test scores to the development of Nigeria and for the fact that Nigerian is a heterogeneous state. The present of bias item in our test has been an issue of concern. In our legislative assembly, Emaikwu, 2012 observed that the Nigerian senate in 2010 summoned the then minister of education to explain why such massive failure occurred in that year’s national examinations; the issue of test item bias and test-wiseness featured prominently among other reasons given for massive failure in some sections of the country. The Registrar of the JAMB, was summon by the House of Representative over student’s mass failure in the 2013 unified tertiary matriculation examination (UTME) which has led to some candidates taking to the street protesting against JAMB. One of the complaints was concerning the usage of computer for the exam and that most of them have never had contact with computer before or that they only learnt the theoretical aspect without the practical usage of computers (Ndiribe, 2013). This is a clear case of test bias. Obviously, students who filled the manual or paper option but only to find themselves being asked to write the exams using computers may have performed poorly in the exam not because they do not have the latent trait

to answer the questions correctly but because they cannot effectively operate the computer. If this allegation is true, can it be said then that the recently release JAMB result is a valid one?

Decision made from test, programs drawn from such bias test, policies developed based on such bias test and research findings that resulted from scores derived from such bias test are misleading and will not provide solution to educational problems. The implication is that if education is supposed to foster national development and the test which are supposed to be used for assessing who has benefited from the educational system could make some of the examinees to take to the street protesting because the test is bias. Then all decisions, programs, policies and research findings drawn from such test are also biased. Hence, if the issue of test bias is not properly handled, then the statement 'education can foster national development' becomes unattainable.

3.Methods of Detecting Item Bias

The usual thing is to statistically detect items that are bias and then send them to the bias revision panel who examine these items to find out if they are actually biased items. An item may statistically show bias but if the item is measuring something the examinees need to know, and there is nothing in the item that might offend or unfairly penalize students, then such an item should remain in the test. Typically, the panel is made up of educators and non-educators who represent the minority student subgroups who will, in the future, be required to take the test for which the items are being bias-reviewed. There should also be a reasonable representation of males and females on such panels. The panel members, after having received an orientation and some training, will review every potentially bias item by supplying a yes or no answer to a review question (Papham, 2012). Based on their responses the items that are actually biased are taken out. Also, among this bias review panel are experts in language and culture who judge the wording, and social merit of such questions. Those items that are actually bias are either discarded from the test or they are reviewed by rewording/restructuring them. There are several statistical methods that can be used to detect bias items, namely:

3.1Item Discrimination Index

This is done by finding the discrimination index of the item for both groups. If the discrimination indexes are approximately equal, then the item is probably not bias but if the values are not approximately equal, then such item could be bias.

3.2 Factor Analysis

It can be used to evaluate internal structure of a test separately for the two groups. If only one factor is found in each of the group then the test does not contain bias items but if in one of the group more than one factor is found, then the test is bias.

3.3 Rank Order

This is a quick method. Here the test items are rank in order of difficulty for each of the two groups. If the item rank differs across groups, then the test is suspected to be biased.

3.4 Differential Item Functioning (DIF)

This is the best way to evaluate item bias. Examinees are divided into two groups, the focal and reference group. The examinees are matched on ability levels, and then the focal and reference groups are compared to see whether they differ. If they differ significantly the item is said to be bias. Examples of DIF methods are Mantel-Haenszel (M-H), standardization, simultaneous item bias test, transformation item difficulty, logistic regression, scheuneman chi-square, item characteristic curve item response theory-likelihood ratio, comparison method, Lord's chi-square test, log linear model, parameter index and so on. Many of these methods require the use of computer with more sophisticated software. In Nigeria some of these software are not readily available and even where they are available they are difficult to understand. This has discouraged many researches who want to carry out research in this area. However, a very simple but quite reliable method is the M-H which can be calculated manually and it requires small sample size hence classroom teachers at any level can easily use it to detect bias items in their test.

4. Basic Steps for Mantel-Haenszel Chi-square Analysis (M-H)

1. Divide examinees into two groups for comparison say focal and reference groups (e.g. female/male, rural/urban or low socio-economic status/high socio-economic status).
2. Use the total scores of the test whose bias is being studied to split the examinees into about five score groups (matching levels).
3. State the null hypothesis
4. Taking each matching group at a time determine A_k and C_k by counting the number of examinees that got the item correct in the reference and focal group respectively.
5. Taking each matching group at a time determine B_k and D_k by counting the number of examinees that got the item wrong in the reference and focal groups respectively.
6. Add the frequency across reference and focal groups and across right and wrong responses to get the marginal N_{rk}, N_{fk}, N_{1k} and N_{0k}.

7. Determine $E(A_k)$, $E(B_k)$, $E(C_k)$ and $E(D_k)$, which are the expected frequencies. For cell 1: $E(A_k) = (N_{rk}N_{1k})/N_k$ (Repeat this for each of the cells).
8. Calculate the M-H chi-square using the formula $(|\Sigma_k A_k - \Sigma_k E(A_k)| - 0.5)^2/\Sigma_k var(A_k)$.Where $Var(Ak) = N_{rk}N_{fk}N_{1k}N_{0k}/N_k^2(N_k-1)$. The total chi-square will be the sum of the chi-square for each matching group.
9. Find the table value of df=1 at 0.05, if X^2-calculated $\geq X^2$-table value the null hypothesis is rejected but if X^2-calculated$< X^2$-table value the null hypothesis is accepted.

10.The common odds-ratio is calculated by $\alpha_{MH} = (\Sigma_k A_k D_k/N_k)/ (\Sigma_k B_k C_k/N_k)$. When it is less than 1 it indicates a possible bias against the focal group but when it is more than 1 it indicate a possible bias against the reference group, while a value of 1 signifies no bias.

11.The odds-ratio estimator is transformed into the Educational testing service "delta metric", which indicate whether DIF is negligible, moderate or large. $\Delta\alpha_{MH} = -2.35\ln(\alpha_{MH})$

A. Negligible DIF, where X^2 is non-significant or the absolute value of Δ is less than 1.0
B. Intermediate (moderate) DIF, when X^2 is significant and Δ ranges from 1.0 to 1.5 in absolute value
C. Large DIF, where X^2 is significant and the absolute value of Δ is more than 1.5

C items are biased items, A items are not biased items and B items are slightly biased.

5. Conclusion

Bias items in test as discuss in this paper can affect the evaluation process of our educational system. Education the major vehicle towards achieving national development should be properly implemented at all stages. It is hoped that if biased items are removed from our tests it will greatly improve the evaluation aspect of our educational system. It is recommended that: Before any test is put into public use, every effort should be made to detect the biased item and do away with them; Item writers should be trained on how to identify bias items and on how to write bias-free items; A review panel should be set up for national/state wide examinations to review biased items; Test writers should construct items that are free from writing errors such as, offensiveness, inflammatory, irrelevancy, stereotyping, controversy, complex vocabulary, ethnocentrism, demeaning, elitism, specialized legal or business terms, regionalism, specialized tools, sports, transportation terms. General terms are better. Test writers should be sensitive to cross-cultural issues and religious issues; Test writers should try to represent the heterogeneous nature of Nigeria by showing various ethnic groups in their test items.

References

Alordiah, C. O. (2013). A Theoretical comparison of selected Differential Item Functioning detection procedures: A challenge for psychometricians. *African Journal of Studies in Education*.**9**(1), 99-104.

Alordiah, C. O. (2012). Item Response Theory (IRT) – A more desirable choice for quality assurance in Education. *Research in Education*.**18**(1), 255-258.

Atar, B. (2006). Differential Item Functioning Analysis for mixed response data using IRT likelihood-Ratio Test, Logistic Regression, and GLLAMN procedures. Electronic.Treatises and Dissertations (ETDs).Paper 248. http//diginole.lib.fsu.edu/etd/248.

Berk.(2007). Item Bias detection methods for small samples.Dissertation abstract international.

Emaikwu, S. U. (2012). Issues in Test Item Bias in Public Examinations in Nigeria and implications for Testing.*International Journal of Academic Research in Progressive Education and Development.***1**(1), 175-187.

Federal Republic of Nigeria.(2004). *National Policy on Education*. Lagos: NERDC press.

International Labour Organization (ILO, 2004). R195: Human resources development recommendations, 2004. Article 19(f) and (g). Retrieved from : http://www.ilo.org/ilolex/cgi-lex/convde.p/?R195.

Joshua, M. T. (2005). Test/Item bias in psychological testing: Evidence in Nigeria system. A paper presented at the annual conference of Nigerian Association of Educational Psychologist held at Ahmadu Bello University Zaria from 24th-28th march.

Ndiribe, O. (2013). Reps summon JAMB Registrar over mass failure. Vanguard forum. Retrieved from http://www.vanguardngr.com/2013/05/repssummon-jamb-registrar-over-mass-failure/

Nenty, H. J. (2010). Gender-Bias and Human Resources Development: Some measurement considerations. *Ilorin Journal of Education*, **29**,13-26.

Popham, W. J. (2012). Assessment Bias: How to Banish It. Boston: Allyn& Bacon.