EFFECT OF LANGUAGE MANIPULATION ON DIFFERENTIAL ITEM FUNCTIONING IN A MULTICULTURAL SETTING.

BY

JOHN NWANIBEZE ODILI (PH.D)

odilijohn@yahoo.com

DELTA STATE UNIVERSITY, ABRAKA, NIGERIA.

PAPER PRESENTED AT 27TH ANNUAL CONFERENCE OF THE ASSOCIATION FOR EDUCATIONAL ASSESSMENT IN AFRICA, YAOUNDE CAMEROON, AUGUST 24-28, 2009

ABSTRACT

The study investigated the effect of manipulating the language of test items in order to simplify them on differential item functioning. It employed the post-test only control group experimental design to investigate the index of differential item functioning between testees who are exposed to original test form and those exposed to a language manipulated test items. The result showed that manipulation of test items to simplify their language reduced the index of differential item functioning among testees in a multicultural setting. It recommended that for test validity in multicultural setting the language of test items should be simplified to reduce differential item functioning.

Key Words: language manipulation, differential item functioning, test validity, multicultural setting.

Introduction

Recognition of extraneous error variable in test performance is one of the major shifts in Item Response Theory (IRT) of measurement from Classical Test Theory (CTT). Extraneous error variables are those which can bring about difference in performance of testees in a test item other than their ability in the trait that is being measured.

According to Nenty (1996) sources of extraneous error variable in testing include communication skills, such as reading comprehension and writing abilities, illustrations and examples which may be more familiar to examinees from specific subgroup of population of test takers. An item writing process that fails to check for the influence of the sources of extraneous error variables will give rise to test items that will differentially function for different subgroups of test takers.

English language is the medium of testing achievement in biology and many other subjects in West African Senior School Certificate Examination. This examination is taken by testees from different cultural settings. English language competence differ among students from different socio-economic status (Bernstein, 1961; Egglestone 1992). When an examiner fails to edit test items to reduce the influence of language, examinees will perform differentially as a result of their standing in English language competence. This raises the question of validity in test measuring achievement in subjects other than English language.

Background of the Study

Statistics from West African Examinations Council (WAEC) on students performance in Senior School Certificate Examination (SSCE) reveal that a high percentage of the candidates fail biology examination. Only a few percent (less than 20%) pass at credit level.

This scenero has existed for more than two decades inspite of several research on curriculum improvement and instructional methods.

Research in the field of testing reveals that apart from an individuals knowledge in a subject like chemistry the complexity of the English language in which the test item is couched can influence performance in the subject. Cassel and Johnstone (1984), Johnstone and Selepeng (2001) demonstrated that test items with long sentence, those with complex word in key position, questions that are phrased in the negative, questions with contrasted terms of quantity, introduced additional task in testing. Performance in the items is influenced by the examinees competence in English language.

Analysis of questions used in some public examinations in Nigeria showed that some of the items have the characteristics that are mentioned above. For instance, in 2008, the multiple choice questions used by West African Examinations Council in the Senior School Certificate Examination has 19 questions that are phrased in the negative, 6 questions with long stem, and one question that is complex or ambiguous. In similar examination conducted by National Examinations Council (NECO) in Nigeria, 16 questions were phrased in the negative, 2 with long stem. Thus, about 43% of the questions used by WAEC can be further managed to reduce the effect of language. For NECO such items constituted 30%. Nworgu and Odili (2005) have found similar trend in biology multiple choice

questions used by WAEC in previous examinations. This situation may have contributed to the phenomenon of poor performance of examinees in biology in the senior school certificate examination.

This will negatively impact on the use of education for achieving the concept of egalitananism.

Problem

The problem of this study is: What will be the effect of manipulating the language of biology multiple choice questions in order to reduce their complexity on the index of differential item functioning among testees from different socio-economic status?

Purpose

The purpose of the present study are first to manipulate by simplifying the language of biology multiple choice test items which were earlier identified as having significant index of DIF. Secondly, to investigate the effect of such manipulation on index of DIF for testees form high and low socio-economic status (SES)

Significance of Study

Firstly, the study will present information on psychometric literature on possible intervention to reduce DIF in tests. Secondly, findings from this

study will be useful to public examination bodies like WAEC, NECO, National Business and Technical Board (NABTEB) etc, in presenting guidelines that will enable item writers to write test items that are unidimensional in measurement of learning outcomes.

Thirdly, teachers and textbook authors will be guided by the result to recognize the need to pay attention in the explanation of non-technical words in biology teaching.

Scope of the Study

Only biology multiple choice questions were investigated. The students covered were those from high and low SES. Language manipulation covered only items with reducible long stem, those phrased in the negative and items with complex synonyms in key position.

Research Questions

The following research questions guided the study:

1 How well does differentially functioning test items manipulated in terms of language reduce the mean index of DIF for students from high and low SES compared with the original test form?

2 How well does differentially functioning test items manipulated by reducing the number of non-technical words reduce the mean index

of DIF for students from high and low SES compared with the original test form.

3 To what extent do differentially functioning test items manipulated by changing negative to positive phrase reduce the mean index of DIF for students from high and low SES compared with the original test form.

4 To what extent do differentially functioning test items manipulated by using simpler synonym in key position reduce the index of DIF for students form high and low SES compared with the original test form?

Hypotheses

In order to test further the answers to the above research questions, the following hypotheses were tested at 0.05 level of significance

HO1: There is no significant difference in the mean value of index of DIF of Biology multiple choice test items manipulated in terms of language compared with those presented in the original language formats for students form high and low SES.

HO2: There is no significant difference in the mean value of index of DIF of Biology multiple choice test items manipulated by reducing the number of non-technical words compared with the original test form for students from high and low SES.

HO3: There is no significant difference in the mean value of index of DIF of Biology multiple choice test items manipulated by changing negative phrase to positive phrase compared with the original test form for students from high and low SES.

HO4: There is no significant difference in mean value of DIF index of Biology multiple choice test items manipulated by using simple non-technical word in key position compared with the original test form for students from high and low SES.

Literature

Theoretical frame work

According to IRT of measurement test score can be decomposed into three orthogonal components-true score component, extraneous error component and random error component. The true score component is that which is sustained by the examinees standing in the latent trait that is measured. The more the contribution of this component in a test score the higher the reliability of the score and test that yields it. The extraneous error variable component is that portion which is contributed by abilities alien to the latent trait under measurement. When it is not checked in testing it increases invalidity and reduces reliability of the score and test that yield it. Sources of extraneous error component include

communication, skills, illustrations that have cultural bearing etc. The random error component can be minimized by sampling.

Differential item functioning (DIF) is the tendency of testees of the same standing in the latent trait to perform differently in a test item. According to Dogan, Guerrero and Tatsuoka (2005), “DIF occurs when test-takers form different subgroups with identical overal test score or ability level, differ systematically with regard to the probability of solving particular test item(s)”p4 .A test item whose sources of extraneous error are not checked will likely function differently for subgroups with identical overall test scores.

Cassel and Johnstone (1984), Johnstone and Selepeng (2001) implicated language of test item as a possible source of extraneous error. They found that test items in chemistry multiple choice which have complex word in key position, items with contrasting terms of quantity, items with negative phrase, items with large number of words presented additional task of language dimension and were more difficult for testees whom English is their second language. Such items are likely to function differently for testees from different English language competence background. Thus English language presents a problem in testing in multicultural setting.

Research in DIF

Interest in analysis of differential item functioning in test derives from the consideration that nations all over the world perceive education as instrument for achieving egalitarianism among persons. To achieve this demand that test items should measure traits which are taught in school subjects and not those traits that are alien to it. The violation of this reasoning was responsible for criticism of use of tests in United States of America. The argument was that tests items discriminated unfairly against minority groups. The result of such criticism gave rise to legislations that sought to protect the minority groups in the use of the test results as well as taking steps to detect and reduce differential item functioning.

Several techniques were developed to detect DIF in test (Angoff & Sharon, 1974; Scheumeman, 1979, Holand & Thayer, 1986, Mellenberg, 1982). Literature also showed that DIF has been detected in test items used in public examination in United States and Nigeria. For instance, Roggers and Kulick (1987) detected DIF in Sholastic Aptitude Test used in America. Mcpeek and Wild (1987) detected DIF in the Graduate Record Examination General Test and Graduate Management Admission test respectively. Dogan, Guerrero and Tatsuoka (2005) detected DIF in Third International Mathematics and Science Study-Repeat (TIMSS-R). In Nigeria, tests used in public examinations have been analysed for DIF. Umoinyang (1991) detected DIF in mathematics multiple choice test used

by WAEC in the 1990 General Certificate Examination. Nworgu and Odili (2005) detected DIF in Biology multiple choice test items, used by WAEC in 1999 Senior School Certificate Examination (SSCE). Similar research by Odili (2005a) and (2005b) isolated items that differentially functioned in WAEC SSCE Biology multiple choice test in 2000 and 2001 respectively.

According to Scheuneman (1987), research in the field of DIF should move beyond detection to procedures for reducing or eliminating DIF. Literature did not present methods of reducing DIF. Dogan et al (2005) was concerned with the use of DIF to detect strengths and weaknesses in Mathematics achievement across countries. The present study was concerned with language manipulation of items detected to have functioned differentially in order to examine the effect on reduction of DIF for testees from high and low SES.

Research studies have reported causal relationship between language of test items and their difficulty index (Cassel & Johnstone, 1984; Johnstone & Selepeng, 2001; Prophet & Badede, 2006). Hubbard (2006) also identified misunderstood word as one of the barriers to effective study and by the same vain performance in test.

Cassel and Johnstone (1984) reported that use of simpler synonymes in key position increased the number of candidates who passed the item. Similar observation held for items that were changed from negative to positive phrase, removal of contrasting terms of quantity as well as

reducing redundant words in stem of a multiple choice test item. Testees whom English is their second language were found to perform better in test items with simplified English language. The investigations did not specify SES of the subjects, and did not cover DIF.

English language competence is not evenly distributed among population of test takers. Bernstein (1961), Egglestone (1992) stated that competence in English language was higher for individuals who are from high socio-economic status. Bright and Megregor (1985) also stated that competence in English language was related to socio-economic status. According to Egglestone (1992) the presence of different levels of SES in the society represents micro-cultures within the larger cultural setting. Assessment carried out by WAEC cuts across several cultural settings, each with different competence in English language. The study investigated the effect of simplifying language of test items on DIF for students from high and low socio-economic status.

Methods and Procedure

The post-test only control group experimental design was adopted in the study. This design was capable of collecting data to compare the effect of language manipulation on DIF among students from high and low SES.

Population

Senior Secondary School three students in Delta State, Nigeria constituted the population of the study. All the students have enrolled Biology in West African School Certificate Examination. They are also distributed across the levels of SES.

Sample and Sampling Technique

The sample size comprised 1,025 students. It was composed using simple random sampling. At the first stage 10 Senior Secondary Schools were sampled. Using proportionate sampling technique 1,025 students were selected from the schools.

Instrument for data collation

The instrument of data collection was an SES questionnaire and Biology multiple choice achievement test which were in two forms. The SES questionnaire was adapted from an instrument that was earlier developed by Adelusi (1982). The Biology achievement test comprised 30 multiple choice test items which were identified in previous studies as differentially functioning (Nworgu and Odili 2005; Odili 2005a, 2005b).

It has two forms. Form B contains questions which have been manipulated to simplify the English language. It has 3 sub-tests:

Sub-test 1 has 12 items with reduced number of non-technical words e.g.

Original question in test administered by WAEC: (SSCE 1999).

Which of the following diseases results from deficiency of insulin?

A. Cretinism B. Goitre C. Beri-beri D. Diabetes.

Modified form of question

Deficiency of insulin causes

A. Cretinism B. Goitre C. Beri-beri D. Diabetes

Subtest 2 has 9 items which were changed from negative phrase to positive phrase, e.g.

Original question in test administered by WAEC: (SSCE 1999)

Which of the following organisms does not exist as a single cell?

A. Amoeba B. Euglena C. Volvox D. Chlamydomonas.

Modified form of question

Which of the following exist as a multicellular organism

A. Amoeba B. Euglena C. Volvox D. Chlamydomonas

Sub-test 3 has 9 items with more familiar synonyms in key position, e.g.

Original question in test administered by WAEC: (SSCE 1999)

The dentition of a herbivore is distinct from that of other animals due to the presence of

A. Diastema B. cups on molar teeth C. Canines D. Ridges on molar teeth.

Modified form of question:

The dentition of a herbivore is different from that of other animals due to the presence of

A. Diastema B. Cups on molar teeth C. Canines D. Ridges on molar teeth.

These items were scattered in the test in order to mirror the pattern in WAEC original instrument, and as well as reduce response pattern. Form A contains the questions as they were originally used by WAEC.

The reliability of the SES scale was established using test-retest method within an interval of two weeks. It yielded a coefficient of stability of 0.87. The content validity was ascertained by matching the items with criteria identified by Chinoy (1967) for classification of individuals into levels of SES. The reliability of the Biology achievement test was established using Kuder-Richardson method. A value of 0.56 was obtained for test form A, while 0.65 was obtained for form B as measures of internal consistency.

Experimental Procedure

The independent variable is language of Biology achievement test items used by WAEC in SSCE which were identified as DIF, while the dependent variable is pattern of response which can be DIF or non DIF. The independent variable was manipulated by presenting the Biology achievement test in the original language form (Form A) used by WAEC,

and a simplified language form (form B). They served as treatment as well as instrument of data collection. This procedure was in line with the method employed separately by Cassel Johnstone (1984); Newman, kundet, lane and Bull (1988); Akpabio and Nenty (2000).

Students were randomly assigned to the experimental and control groups. Students in experimental group were administered form B, while those in control group were administered form A. The SES questionnaire was administered to the students in the two groups. All the students in a class were used in the study.

Analysis of data

Index of DIF was established using Scheuneman modified chi-square statistic, while the hypotheses were tested using dependent t-test statistic. The SES scale has three responses options. The responses that indicate high SES has 3 points, while 1 point indicated low SES. Middle SES has 2 points. For the 20 items, students with score of 50 and above (upper real limit of 2 multiplied by total number of items) were classified as belonging to high SES, while values below 50 where classified as low SES. The analyses were performed with SPSS computer package, and the hypotheses were tested at the 0.05 level of significance.

Results

Results are presented in tables below according to the research questions (RQ) and hypotheses (HO).

RQ1. How well does differentially functioning test items manipulated in terms of language reduce the mean index of DIF for students from high and low SES compared with the original test form?

The data used to answer the above RQ was presented in table one.

Table one

Scheuneman chi-square SES differential item functioning indices for test items in form A (original language format) and form B (simplified language format).

Item	Scheuneman X^2 Form A	Scheuneman X^2 Form B
1	3.16	0.53
2	1.52	0.92
3	2.45	0.04
4	7.20	1.09
5	2.87	0.35
6	3.35	2.63
7	3.38	4.08
8	0.69	1.51
9	3.68	1.99
10	4.60	2.75
11	4.44	1.70
12	1.64	1.23
13	4.01	2.27
14	7.56	2.33
15	2.56	0.23

16	0.51	2.63
17	7.71	5.83
18	3.16	1.44
19	5.61	1.36
20	2.16	0.98
21	1.95	0.65
22	4.43	2.77
23	4.06	3.38
24	4.27	2.53
25	0.77	1.72
26	6.42	1.32
27	1.04	0.19
28	2.97	1.06
29	7.73	4.42
30	1.23	2.84
$\bar{x}$	3.57	1.89

The table shows that simplifying the language of test item reduced the index of DIF. This was evident in 27 out of the 30 items or 90% of the items. The mean DIF was also low for items with simplified language format.

RQ2: How well does differentially functioning test items manipulated by reducing the number of non-technical words reduce the mean index of DIF for students from high and low SES compared with the original test form?

The table below presents the data that was used to answer research question 2.

Table 2: Scheuneman chi-square SES differential item functioning indices for subtest I (reduced number of non-technical words)

Item	Scheuneman X^2 Form A	Scheuneman X^2 Form B
1	3.16	0.53
2	1.52	0.92
3	3.35	2.63
4	0.69	1.51
5	3.68	1.99
6	1.64	1.23
7	7.56	2.33
8	0.51	2.63
9	3.16	1.44
10	1.95	0.65
11	4.27	2.53
12	0.77	1.72
$\overline{x}$	2.68	1.68

Reduction in the number of non-technical words reduced the index of DIF in 11 of the 12 items or 98% of the items. The table also revealed reduction in mean value of DIF in test items manipulated by reducing the number of non-technical words compared with the original test form for students from high and low SES.

RQ3: To what extent do differentially functioning test items manipulated by changing negative to positive phrase reduce the mean index of DIF for students from high and low SES compared with the original test form?

The table below presents the data that was used to answer this research question.

Table 3: Scheuneman chi-square SES differential item functioning indices for subtest 2 (changing from negative to positive phrase)

Item	**Scheuneman X^2 Form A**	**Scheuneman X^2 Form B**
1	2.45	0.04
2	4.60	2.75
3	4.44	1.70
4	5.61	1.36
5	2.16	0.98
6	4.06	3.38
7	2.97	1.06
8	7.73	4.42
9	1.23	2.84
$\overline{x}$	3.92	2.06

Changing test items from negative to positive phrase reduced the index of DIF for all the items except one. The mean DIF was also lower for form B compared with that in form A. Thus changing multiple choice test items from negative to positive phrase reduces the mean index of DIF for high and low SES testees.

Research Question 4

To what extent do differentially functioning test items manipulated by using simpler synonym in key position reduce the mean index of DIF for students from high and low SES?

The data used to answer this research question is presented in table 4

Table 4: Scheuneman chi-square SES differential item functioning indices for subtest 3 (Using simpler non-technical word in key position).

Item	Scheuneman x^2 Form A	Scheuneman x^2 Form B
1	7.20	1.09
2	2.87	0.35
3	3.38	4.08
4	4.01	2.27
5	2.56	0.23
6	7.71	5.83
7	4.43	2.77
8	6.42	1.32
9	1.04	0.19
$\overline{x}$	4.40	2.01

The table shows that 8, representing 88% of the 9 items in this subtest have reduced DIF index in form B for testees from high and low SES. The mean DIF was also reduced in test form B. Thus, the use of simple synonyms or more familiar words in key position of biology multiple choice test item reduces the mean DIF for high and low SES testees.

In order to prove further the answers to the research questions, the following hypotheses were tested:

Ho_1: There is no significant difference in mean value of index of DIF of Biology multiple choice test items manipulated in terms of language formats for students from high and low SES.

Dependent-test analysis was used to test observed means in test form A and B for significant differences. The result is presented in table below.

Table 5: Dependent t-test analysis of the effect of item manipulation on the value of SES differential item functioning of Biology multiple choice test (n = 30 items).

	$\bar{x}$	**SD**	**DF**	Diff $\bar{x}$	**SD**	**SE**	**t**	**t Critical**
Form A	3.57	2.14	29	1.68	1.92	.35	4.79	2.045
Form B	1.89	1.35						

The table shows that the calculated t value of 4.79 is above 2.045, which is required for statistical significance at 0.05 level of significance. As a result, we reject the null hypothesis. Thus, the result shows that there is statistical significant effect of language manipulation on reduction of DIF in Biology multiple choice test for students from high and low SES.

HO_2: There is no significant difference in the mean value of index of DIF of Biology multiple choice test items manipulated by reducing the number of non-technical words compared with the original form for students for high and low SES

The data used to test the above hypothesis is presented in table 6.

Table 6: Dependent t-test analysis of the effect of reducing number of non-technical words on the value of SES differential item functioning index of Biology multiple choice test.

	$\bar{x}$	**S.D**	**df**	Diff		**SE**	**t**	**t**
				$\bar{x}$	**S.D**			**Critical**
Form A	2.68	1.99						
Form B	1.68	0.76	11	1.01	1.90	0.55	1.85	2.201

The calculated t-value of 1.85 is below 2.045 required for statistical significance at the 0.05 level of significance. Thus, we accept the null hypothesis that reducing the number of non-technical words in Biology multiple choice test items does not reduce DIF for testees from high and low SES.

HO3: There is no significant difference in mean value of DIF index of Biology multiple choice test items manipulated by changing negative phrase to positive phrase for students from high and low SES.

This hypothesis was tested with the data presented below.

Table 7: Dependent t-test analysis of effect of changing Biology multiple choice questions from negative to positive phrase on SES DIF.

	$\bar{x}$	S.D	DF	Diff		SE	t	t
				$\bar{x}$	S.D			Critical

Form A	3.92	1.99	11	1.86	1.69	0.56	3.302	2.201
Form B	2.06	1.38						

From the table it could be seen that the calculated t is larger than the critical value. Thus we reject the null hypothesis as stated. Therefore changing multiple choice questions from negative to positive phrase brings about significant reduction in DIF among students from high and low SES.

HO_4: There is no significant difference in mean value of DIF index of Biology multiple choice test items manipulated by using simple non-technical words in key position for students from high and low SES.

The table below presents the data used to test the above hypothesis.

Table 7: Dependent t-test analysis of effect of using simple non-technical word in key position on SES DIF.

	$\bar{x}$	SD	DF	Diff		SE	t	t
				$\bar{x}$	SD			Critical
Form A	4.40	2.27	8	2.39	2.07	0.69	3.46	2.306

Form B	2.01	1.94

The table shows that the calculated t-value of 3.46 is above 2.306 required for statistical significant difference. Thus, we reject the null hypothesis as stated .It shows that the use of simple synonyms in key position of Biology multiple choice questions will reduce the index of DIF for students from high and low SES.

Discussion

This study revealed that complexity in English language of Biology multiple choice questions is capable of increasing the index of DIF for students from high and low SES. Examples of such complexity include phrasing questions in negative, use of redundant words in the stem of questions, thus increasing the length, and use of complex (less familiar) words in key position. These kinds of questions features in multiple choice test items used by WAEC in the SSCE. Previous research reports showed that students performance have been influenced by the SES of the parents (Avoseh, 1983, Vamadevappa, 2002). Research have also revealed that competence in English language is high among students in SES (Bernstein, 1961, Egglestone,1992). Biology multiple choice questions phrased in complex English language will be biased to students who are from low SES because of their low English language competence. This presents a serious challenge to testing in a multicultural setting. In Nigeria

and many other nations education is expected to achieve the objective of creating an egalitarian society. Social mobilization among citizens ought to be independent of the position or status of parents. Egglestone noted that since school curriculum is presented to children from different backgrounds, what is measured or tested should be what is taught in school in a given subject. This is one of the ways of ensuring that public examinations are valid.

Test items presented in complex language present additional task which was greater for testees from low English language competence. Johnstone and Selepeng (2001) presented an information processing model which explains the source of this difficulty. This is presented in the diagram below

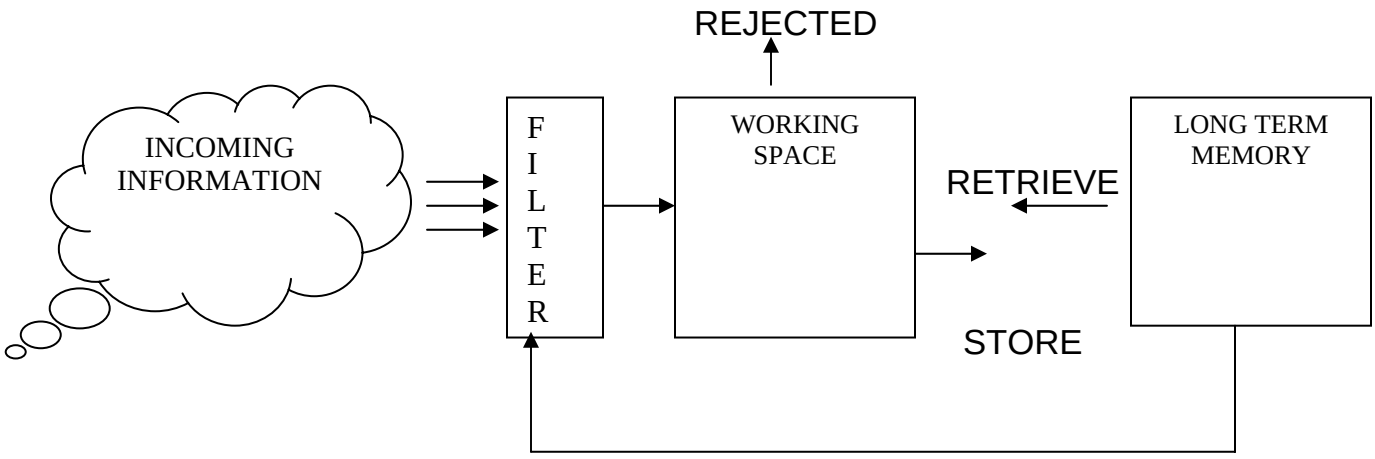

Fig. 1: Information processing model. Source: Johustone and Selepeng (2001) p.23

The model has three levels of information processing: filter, working space, and long term memory .Incoming information (stimulus task) undergoes filtration .What is admitted into the working space (short term

memory) depends on its familiarity, interest and meaningfulness to the testee. The working space performs two functions –holds the inputs material as well as process it for meaningful deduction. For the second language science learner, the working space performs a third function-translating information from English to second language and then back to English language. For science learning and testing, the testee is faced with two unfamiliar words-English and science. The more complex and unfamiliar the more difficult test items will be to testees. Since students from high SES are more competent in English language the problem of information processing is reduced. When the wordings of test item are reduced, there is ease of filtration and more information is available for processing.

According to Cassels and Johnstone (1984) negative phrases are more complex because they require more of the working memory capacity than those in the positive. They also require additional thinking level. The use of these in testing introduce additional problems to the testees. Johnstone and Selepeng also reported that a more familiar word at key position will present lesser task to the working memory especially for second language learner of science. The acceptance of hypothesis two not withstanding the model shows that reducing the number of non-technical words will reduce the problem of filtration. This will reduce noise inherent in managing large number of words.

Conclusion

From the study it can be concluded that simplifying the English language of non-technical words used in Biology multiple choice test reduces the index of DIF for students from high and low SES. Since education is expected to serve the function of egalitarianism, test item-writers should edit questions in order to reduce the English language complexity. It is the duty of public examination bodies to ensure that this is implemented at the item writing stage of test construction.

References:

Adelusi, O.I. (1982). Relationship between selected variables on achievement in English as a second language. *Unpublished Ph.D Thesis.* University of Ibadan.

Akpabio C.U and Nenty H.J. (2000). Item arrangement as a strategy of enhancing quality assessment in science Education. *Journal of Southen Africa Association for Research in Mathematics, Technology and Science Education. 4*(1) 45-56

Avoseh, O.O. (1983). Educational opportunity and academic performance: A study of secondary school students in urban area. *Journal of Educational Research Council 3* (1), 1-9

Bernstein, B. (1961). Social class and linguistic development in J. Floud, A. A. Halsey; and C.A. Anderson (eds.) *Education and Society*. Glencoe, Illinois: Free Press;

Cassels, J.R.T and Johnstone, A.H. (1984). The effect of language on student performance on multiple choice test in chemistry*. Journal of Chemistry Education. 61*(7); 613-615.

Chinoy, E. (1967). *Society 2nd Edition*. New York: Random House.

Dogan, E; Guerrero A; Tatsuoka, K (2005). Using DIF to investigate strengths and weaknesses in mathematics achievement profiles of 10 different countries. Paper presented at the Annual Meeting of

the National Council on Measurement in Education (NCME) Montreal, Canada, April 12-14.

Egenege, C.C. (1998). Gender issues and students achievement in reading comprehension. Paper presented at the national conference on sex equity and national development, college of Education, Warri, April.

Hubbard, L.R (2006). *Learning how to learn*. US: Effective Education.

Johnstone, A. H. and Selepeng, D. (2001). A language problem revisited. *Chemistry Education: Research and Practice in Europe. 3*(1) 19-29

Nenty, J. H. (1996). Advances in test Validation. In Gani, A. Badmus and Peter, I. Odor (eds.) *Challenges of Managing Educational Assessment in Nigeria.* JAMB, NABTEB; NBEM.

Newman, D.L; Kudert, D.K; Lane, D.S; and Bull, K.S. (1998). Effect of varying item order on multiple choice test scores*. Applied Measurement in Education, 1*(1); 89-97.

Nworgu, B.G. and Odili, J.N. (2005). Analysis of DIF in WAEC SSCE Biology Multiple Choice Test. *Review of Education 16,* (2) 140-152.

Odili, J.N. (2005a). DIF: A challenge to psychological measurement. *The Educational Psychologist 1,*(2) 204-214.

Odili, J.N. (2005b). Quality assurance in credentialing: Analysis of DIF in 2001 WAEC SSCE Biology multiple choice test. *Studies in Education 8*, 102-109

Pickersgill, S. and Locks, R. (1991). Students understanding of selected non-technical words in science. *Research in Science and Technological Education 9* (1); 71-79

Prophet, R. B. & Badede, N. B. (2006). Language and Student performance in Junior Secondary Science Examinations: The Case of Second Language Learners in Botswana. National Science Council, Taiwan.

Scheuneman, J.D. (1979). A method of assessing bias in test items. *Journal of Educational Measurement. 16*(3); 143-152.

Sheuneman, J.D. (1989). An experimental exploratory study of causes of bias in test items. *Journal of Educational Measurement. 24* (2); 97-118.

Umoinyang, I.E. (1991). Item bias in mathematics achievement test: *Unpublished M.ed Thesis*, University of Calabar.

Vamadevappa, H. V. (2002). Causative factors for underachievement in Biology. *Perspectives in Education. 18*(3); 171-178.