EFFECT OF LANGUAGE MANIPULATION ON DIFFERENTIAL ITEM FUNCTIONING OF AGRICULTURAL SCIENCE MULTIPLE CHOICE TEST ITEMS

Ursula Ngozi Akanwa Prof[1] & **Kelechi Joshua Princewill Ihechu**[2]
Michael Okpara University of Agriculture,
Umudike, Abia State, Nigeria
&
P.C. Nkwocha *PhD*[3]
College of Education, Owerri , Imo state.

Abstract
This study assessed effect of language manipulation on differential item functioning of Agricultural Science multiple choice test items used by National Examination Council in the Senior School Certificate Examination (SSCE) 2013-2015. A post-test only control group of quasi experimental design was employed. The population consisted of 3,238 senior secondary school students in class three (3) in public secondary schools in Imo State who studied Agricultural Science as a certificate subject. A sample size of 100 students was used for the study got by three stage proportionate stratified random sampling techniques. Three research questions and three hypotheses were asked and formulated respectively. The instruments of data collection were: Agricultural Science paper 3 objective test used by NECO in 2013-2015 examinations; Agricultural Science form A test made up of 30 test items that were identified as differentially functioning; Agricultural Science form B test which was made up of the 30 items in form B, but presented in simplified language format. The fourth instrument was student's socio-economic status questionnaire that was used to collect students' biographical information. Scheuneman modified chi-square statistic was used to analyze the research questions. Dependent t-test statistic was used to test the null hypotheses at 0.05 level of significance. The study revealed that Agricultural Science multiple choice questions used by NECO in the SSCE 2013-2015 contain test items that significantly function differently for students based on socio-economic status, geographical locations and gender of testees. It was also found that simplifying the language of Agricultural Science test items brought about significant reduction in differential item functioning. Based on the findings it was recommended that test item writers for public examinations should be encouraged to keep simple English language used to reduce the tendency of introducing additional tasks into the measurement of Agricultural Science ability.

Introduction

Education is an instrument for development of human potentials and self -actualization. According to Eggleston (2012), education helps to reduce class, race, and gender differences. Education is a tool for the achievement of stable democracy and the building of a society where everyone has equal opportunities. Education could help to achieve the purpose of development of individuals and creation of equal opportunity by two means of the process of instruction, and measurement of learning outcomes for the purpose of classification of learners and certification. In carrying out these processes, education practice must recognize individual differences in learners. Students come to school from different backgrounds, with corresponding difference in knowledge, attitude, values and language ability. Despite the wide spectrum of differences, it is the task of the school to present common knowledge, attitudes and values that are based on the objectives of the curriculum. It is also the task of the school to administer tests that measure the knowledge, attitude and language abilities taught in schools. Thus, a major task of test developers in this direction is to construct tests that measure these common skills and bodies of knowledge, without introducing any extraneous elements into the performances on which the measurement is based. Studies exist that provide evidence to show that test used in some public examinations in Nigeria contain items that differentially function. The issue of educational measurement in research pointing towards enhancing the fairness of test or examination across sub groups of examinees is very essential because important decisions are made based on scores of the examinees. Test consists of a set of uniform questions or task to which a student or testees is to respond independently and the result of which can be treated in such a way as to provide a quantitative comparison of the performance in different students (Nworgu, 2016). The testees or examinees are group of individuals who are examined by a standardized or teacher made examination. Ogbebor, (2012) opined that, testees or test takers of the same latent trait should respond to test items correctly irrespective of their gender, school location and school type.

An examination item is said to be biased if it functions differently for a specified subgroup of test takers. Ogbebor, (2012) states that biased test measure characteristics that are not necessary or items that are irrelevant to the test. Frequently, examination items are considered biased because they contain sources of difficulty that are not relevant to the construct being measured and these extraneous sources affect test-takers' performance (Zumbo, 2009). Any item in which a subgroup performs significantly better than other subgroups, is said to have differential item functioning (DIF) with respect to the two groups. Ogbebor (2012), noted that DIF occurs when a test item measures an ability

which is unfamiliar to the subject matter, such that students' scores on the item is now sustained by abilities which are unfamiliar to the subject matter. Moreover, students score in a test is made up of three components that sustained by the subject matter ability, random error in measurement and extraneous factors.

According to Lord and Novick cited in Odili (2014) noted that sources of extraneous variables include communication skills such as reading comprehension and writing abilities (language skills), which may differ among individuals or groups. An item writing process that fails to check for the influence of the sources of extraneous variable error will give rise to test items that will differentially function for different subgroups of test takers. According to Novick and Nenty cited in Odili (2014) language of test item is an important source of extraneous variable in test performance. Research studies show that students' performance in a test is influenced by level of sophistication of language in which test items are couched. Language manipulation is a procedure by which the non-technical English words in Agricultural Science multiple choice question are modified to reduce their contribution to task demanded in such questions. Reporting, Cassels and Johnstone (2014), revealed that manipulating the language of test items through restructuring of the nontechnical words in multiple choice test increased facility of test items. Johnstone and Selepeng (2016) repeated the same study and found that the influence of language in test performance is still a major problem of test takers. Studies have revealed that English language competence is influenced by variables like socio-economic status, geographical location, and gender of students. Like in other science subjects, Agricultural Science questions used by National Examination Council (NECO) in the Senior School Certificate Examination (SSCE) are couched in English language. According to Chinoy (2007), Socio-Economic Status (SES) refers to the ranking of a person on the bases of values like education, wealth and occupation.

According to Obanya (2012), Johnstone and Selepeng (2016), stated that students location affects their English language competence. They posit, respectively, that students in urban cities are more competent in English language than their counterparts in the rural villages. According to Pickersgill and Lock (2011) and Egenege (2012), there is a relationship between English language competence and students gender. For instance Egenege reported that male students achieved higher in reading comprehension than their female counterparts. Given the above circumstance, one can argue that if the English language of Agricultural Science multiple choice questions are complex, they will pose differential difficulty to students of the same subject matter ability

from high, middle and low SES, urban and rural geographical locations, and male and female students.

Agricultural Science is one of the subjects that are examined by NECO in the SSCE. However, students' academic achievement in Agricultural Science has been discouraging in Nigeria. In 2013/2014, out of 19,302 Agricultural Science, only 19% of this number had credit level pass. In 2014/2015, a total of 41,080 candidates sat for Agricultural Science, only 16% of this number had credit level pass. In 2015/2016, a total of 41,080 candidates sat for Agricultural Science, only 13% of this number had credit level pass (State Statistics Unit of National Examination Council, Umuahia, 2017). Tests with differential functioning items cannot be used for decision making as they do not capture exactly the actual performance of the learner. There is the need, more than ever before for education and indeed measurement specialists to find out ways by which differential item functioning can be minimized in tests used in public examinations being an instrument for decision making. The problem of this study is stated in the question form as will manipulation, in terms of simplifying the English language of non-technical words in differentially functioning test items reduce the incidence of differential item functioning for candidates from different socioeconomic status, schools located in urban and rural areas, male and female students?

The following research questions and hypotheses guided the study:

- How well does differentially functioning test items manipulated in terms of language reduce the index of differential item functioning for students from high and low SES?
- How well does differentially functioning test items manipulated in terms of language reduce the index of differential item functioning for students form urban and rural areas?
- How well does differentially functioning test items manipulated in terms of language reduce the index of differential item functioning for female students?
- There is no significant difference in the mean value of index of differential item functioning of Agricultural Science multiple choice test items manipulated in terms of language compared with those presented in the original language formats for students from high and low SES.
- There is no significant difference in the mean value of index of differential item functioning of Agricultural Science multiple choice test items manipulated in terms

of language compared with those presented in the original language formats for students from urban and rural schools

- There is no significant difference in the mean value of index of differential item functioning of Agricultural Science multiple choice test items manipulated in terms of language compared with those presented in the original language formats for male and female students.

Method

Post-test only control group design of quasi experiment is use for the study. The population of the study consisted of 3,238 senior secondary school students in class three (3) in public secondary schools in Imo State who studied Agricultural Science as a certificate subject. Senior Secondary School students in class 3 in public schools were used because they run the same academic calendar. They are also expected to be at the same level in coverage of NECO Senior Secondary School Class 3 Agricultural Science syllabus upon which NECO SSCE is based. The sample size of 100 students was used for the study. This number was representative as well as make for effective management of subjects during the experimental treatment. The three stage proportionate stratified random sampling approach was used in sample selection. The schools were first stratified according to the three senatorial districts in Imo State at the time of this study. At the second level, the schools in each senatorial district were stratified into urban and rural. The population was sampled according to the names of the schools. The population of each school was written on a piece of paper and thrown into a box. Through balloting, schools were randomly sampled. Every student in SS3 in the sampled school was used. Secondly, the socio-economic status questionnaire was. Administered before the test to enable the investigator group students according to socioeconomic status and sex

The first step in the experimental procedure was the distribution of the students in the sampled schools into the experimental and control groups. In order to achieve this, the students in the schools were administered the socio-economic status questionnaire. On this basis, they were separated into high and low socio-economic status for the purpose of this study. Equal numbers of students in high and low socio-economic status, male and female in each of urban and rural schools were assigned into the experimental and control groups, except where the number of students were odd.

Four instruments were used for data collection. They were NECO/SSCE Agricultural Science paper 3 questions, used in the 2013, 2014 and 2016 examinations. Each contains 60 multiple choice type questions

constructed by subject experts and developed by test experts in NECO office into a test form. They were used to collect data on students responses in order to identify differentially functioning test items. The second instrument was Agricultural Science test Form B, which was developed by the investigator. It was a set of differential functioning NECO/SSCE Agricultural Science questions manipulated by simplifying the language of the nontechnical words. It contains 30 items. These items were scattered in the test to avoid creating an order that could influence the response pattern of students. Another reason why the items were scattered was to get a test which adopts item presentation format used by NECO. Test form B also served as the treatment variable. This procedure was in line with the method employed separately by Cassel and Johnstone (2014); Akpabio and Nenty (2010). These studies illustrated an experimental design in which the measurement instrument also served as treatment variable.

The third instrument was Agricultural Science test form A. It contains 30 multiple choice Agricultural Science questions which were identified as differential functioning. They were presented in the original form used in NECO/SSCE Agricultural Science paper 3 test in 2013, 2014 and 2015 examinations. This instrument was used to collect base line data upon which the effect of language manipulation of the non-technical words on differential item functioning was measured. The fourth instrument was a socio-economic status questionnaire constructed by Adelusi (2012) adapted for the study. It was used to collect data on students' biography.

Content validity was established for the Agricultural Science Test Form B and Form B through expert judgment. Since the questions were manipulated by the investigator, Agricultural Science teachers who were also experienced in test item writing were used to judge the extent to which Agricultural Science Test Form B and Form A measured the same content and behaviour process with Agricultural Science. A trial test method of reliability was used to establish reliability for both form A and B of the Agricultural Science test. Data obtained were analyzed using Kuder-Richardson 21. A value of 0.56 was obtained for form A, while 0.65 was obtained for form B. Face validity for the socio-economic status questionnaire was established by ensuring that the instrument contain items that measure the yardstick for classification of individuals into high, middle and low socio-economic status. Test-retest method was used to establish reliability for the socio-economic status questionnaire. A value of 0.87 was obtained using Pearson Product Moment Correlation Coefficient. The interval between the test and retest was two weeks

Control of Threats to Internal and External Validity were identified in the present study, using testing and experimental mortality. Experimental mortality was handled by administering the treatment on a date that coincides with the school continuous assessment test. The threats to external validity were artificiality of experimental treatment (Hawthorne effect) and measuring instrument. Scheuneman modified chi-square statistic was used to analyzed the research questions while dependent t-test statistic was used to test the null hypotheses. Dependent t-test was used because the two test forms (A and B) whose pattern of response is being considered are measuring the same thing. They only differ in terms of language.

Results

Table 1: 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	3.11	.52
2	1.42	.82
3	2.43	.04
4	4.60	1.02
5	2.79	.35
6	3.34	2.63
7	3.38	4.18
8	0.69	1.51
9	3.58	1.98
10	4.20	2.73
11	4.44	1.70
12	1.64	1.23
13	4.11	2.26
14	7.56	2.33
15	2.56	0.23
16	0.51	2.53
17	6.71	5.73
18	3.16	1.44
19	5.61	1.36
20	2.16	0.97
21	1.95	0.67
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	1.040	.19
28	2.97	1.06
29	7.73	4.42
30	1.23	2.84

The above Table shows that 25 test items representing 83% of 30 items in test form B have reduced differential item functioning (DIF) index for students from high and low SES when compared with test form A. Only 5 test items representing 17% did not show evidence of reduction.

Table 2: 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	3.18	0.43
2	1.64	1.13
3	3.97	2.62
4	1.51	1.07
5	1.02	1.42
6	2.37	2.75
7	1.57	1.75
8	4.46	3.30
9	3.52	2.49
10	4.52	3.60
11	6.25	1.64
12	1.18	2.31
13	0.95	1.64
14	4.31	2.37
15	4.29	2.87
16	0.83	1.02
17	2.08	3.29
18	2.30	0.87
19	5.82	0.76
20	.90	.48
21	2.85	1.56

22	7.35	2.91
23	4.44	2.37
24	1.32	.50
25	3.03	.42
26	2.56	0.79
27	1.13	2.85
28	4.75	2.21
29	2.96	1.33
30	0.76	2.92

The table shows that differential item functioning index in form B reduced in 20 test items (represent 67%) when compared with form A for students from urban and rural schools.

Table 3: 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	5.42	3.48
2	4.23	3.69
3	1.05	1.49
4	3.57	3.38
5	8.85	2.90
6	4.55	0.87
7	1.48	2.26
8	184	2.36
9	7.50	2.90
10	5.22	0.76
11	9.62	2.30
12	.61	1.92
13	3.19	1.91
14	6.00	4.26
15	3.01	4.53
16	0.32	3.48
17	1.70	1.08
18	5.62	2.00
19	1.93	0.87
20	5.82	1.18
21	5.11	3.58

22	3.62	3.70
23	2.51	2.82
24	7.84	1.45
25	2.14	1.50
26	1.72	0.33
27	2.93	1.87
28	5.46	3.19
29	5.37	4.12
30	2.65	2.12

The Table shows that differential item functioning index in the form B reduced in 24 test items (representing 80%) when compared with similar values in test form A for male and female students. The results revealed that manipulation of the language of Agricultural Science multiple choice test items is capable of reducing the index of DIF for students from high and low SES, urban and rural areas, male and female students.

Table 4. Dependent t-test Analysis of the Effect of language Manipulation on the Value of SES Differential Item Functioning of Agricultural Science

Variables	**X**	**SD**	**Df**	**X Dif**	**SD**	**SE**	**T**	**T Critical**
Form A (Original Language Format)	3.57	2.14						
			29	1.68	1.92	0.35	4.79	2.045
Form B (Simplify Language Format)	1.89	1.35						

Multiple Choice Test (n=30 items)

The Table shows that the calculated t-value of 4.79 is above 2.045, which is required for statistical significance at the .05 level of significance. As a result, we reject the null hypothesis and accept the alternative hypothesis. Thus, the result shows that there is a statistical significant effect of language manipulation on reduction of differential item functioning in Agricultural Science multiple choice test for students from high and low SES.

Table 5. Dependent t-test analysis of the effect of language manipulation on the value of location differential item functioning index of Agricultural Science test items (n =30 items).

Variables	X	SD	Df	X Dif	SD	SE	T	T critical
Form A (Original Language Format)	2.92	1.76						
			29	1.08	1.76	0.32	3.35	2.045
Form B (Simplify Language Format)	1.84	0.97						

From the above, it is evident that the calculated t value of 3.35 is above 2.045 at the 0.05 level of significance. As a result, we reject the null hypothesis and state that there is a statistical significant effect of language manipulation on reduction of differential item functioning in Agricultural Science multiple choice test from Urban and Rural schools.

Table 6. Dependent t - test analysis of the effect of item modification on the value of sex differential item functioning index of Agricultural Science test items (n=30 items).

Variables	X	SD	Df	Xdif	SD	SE	t-cal	t-critical
Form A (Original Language Format)	4.23	2.60						
			29	1.87	2.63	0.48	3.90	2.045
Form B (Simplify Language Format)	2.39	1.17						

From the above Table, it was evident that the calculated t-value of 3.90 is above 2.045 at the 0.05 level of significance. As a result, we reject the null hypothesis, and state that there is statistical significant effect of language manipulation on reduction of sex differential item functioning in Agricultural Science multiple choice questions.

Discussions

The result of the present study showed that there is statistical significant difference in mean value of differential item functioning index of Agricultural Science test items with simplified language compared with those presented in the original language format. In other words when the language of Agricultural Science test items with significant differential item functioning are simplified, they tend to measure the same thing for testees of the same

ability from high and low SES, urban and rural schools and male and female. This finding was in agreement with the finding of Johnstone and Selepeng, (2016) who found out that test items couched in complex language presented additional task to the testees. Ability to overcome this additional task does not depend on the subject matter ability of the testees but his language ability. According to Bernstein (2011), children whose parents have high socio-economic status are brought up with elaborated language code. They have wider vocabulary compared with their counterparts from low socio-economic status who were brought up in restricted language code. Ability to handle test stimulus couched in complex words is easier for children with high SES parents. This no doubt will lead to differential functioning of test items for these groups of students. Simplifying the language of a test stimulus task will tend to eliminate this additional task, and thus present only Agricultural Science task to the testees. This will tend to reduce differential item functioning. The same explanation seems to hold for testees from urban and rural schools.

The result of this study however showed that when language of Agricultural Science multiple choice questions are simplified, they tend to measure the same thing for boys and girls who are of the same subject matter ability from high and low SES, and location. This result seems to suggest that the achievement of students will improve if efforts are made to reduce the complexity of language of test items. This finding is in agreement with the finding of Cassels and Johnstone (2014), who revealed that manipulating the language of test items through restructuring of the nontechnical words in multiple choice test increased facility of test items.

Conclusion

This study focused on effect of language manipulation on differential item functioning of Agricultural Science multiple choice test items used by National Examination Council in the Senior School Certificate Examination (SSCE) 2013-2015. From literature and analysis of data collected it could be concluded that Agricultural Science multiple choice test used by NECO in the SSCE contain test items with significant differential functioning. Such items measured different things for testees of the same subject matter ability from high and low SES, location and gender of testees. It was also concluded from data analysis that location was the greatest influence on differential item functioning. Modification of the language of test items that were significantly functioning differently brought about significant reduction in differential item functioning. Reduction in the number of non technical words in the stem of a

question could significantly reduce differential item functioning for testees from high and low SES, urban and rural schools and male and female testees.

Recommendations

The following recommendations are made.

1. Test item writers for public examinations should be encouraged to use simple English language. This will reduce the tendency of introducing additional tasks into the measurement of Agricultural Science ability.
2. Item writers are encouraged to use few words as much as possible in the stem of multiple choice questions. Unnecessary clauses should be avoided.
3. NECO and other public examination bodies should analyze students' response to test items for differential functioning before use.

References

Adelusi, 0.1. (2012). Relationship Between Selected Variables on Achievement in English as a Second Language, "Unpublished Ph.D Thesis" University of Ibadan, Ibadan.

Akpabio C.U. & Nenty H.J. (2000) Item Arrangement as a Strategy of Enhancing Quality Assessment in Science Education. Journal of Southern Africa Association for Research in Mathematics, Technology and Science Education. 4 (1); 45-56.

Bernstein, B. (2011). Social class and linguistic Developments. In J. Floud; A.A. Halsey, & C.A. Anderson (eds.) Education and Society. Glencoe, Illinois: Free Press.

Cassels, J.R.T & Johnstone, A.H. (2014). The Effect of Language on Student Performance on Multiple Choice tests in Chemistry. *Journal of Chemistry Education. 61 (7); 613-615.*

Chinoy, E. (2007). *Society* 2nd Edition. New York: Random House.

Egenege, C .C. (2012). 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.

Egglestone, J. (2012). The Challenge for Teachers. London: Cassel.

Johnstone, A. H. & Selepeng, D. (2016). A Language Problem Revisited. Chemistry Education: Research and Practice in Europe. 3 (1); 19-29.

Nworgu, B.G. (2016). Educational Research: Basic Issues and Methodology. Ibadan: Wisdom.

Obanya, P.A.I. (2012). Secondary English Teaching. Lagos: Macmillan

Odili, J. N. (2014) Effect Of Language Manipulation On Differential Item Functioning Of Biology Multiple Choice Test "Unpublished Ph.D Thesis" Submitted to the Department Of Science Education, Faculty Of Education, University Of Nigeria, Nsukka.

Ogbebor, U. (2012). Differential item functioning method as an item bias indicator. *International Research Journal. 4(4) 367 373.*

Pickersgill, S.& Locks R. (2011). Students understanding of selected non-technical words in science. Research in Science and Technological Education 9 (1); 71-79

Statistics Unit of National Examination Council, Umuahia, 2017;

Zumbo, B. D. (2009). A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.