

LANGUAGE ISSUES IN TESTING IN A PLURAL SOCIETY

John.N. Odili

Testing is the process of conveying stimulus task to testees in order to elicit response which is graded as an index of level of achievement of learning outcomes. The validity of such stimulus task is the extent to which it is able to convey the intended task in the subject matter and nothing more. Language of test item is the medium of communication in which the stimulus task is conveyed. In Nigeria English is the medium of testing in all the subjects of the curriculum. Testing in the medium of English language in Nigeria poses dual problems. The first is the fact that English is used as a second language. Nigerians have their mother tongue. The second is that Nigeria as a plural society has many cultures and languages. Among the English literates there are subgroups differentiated by their level of competence in English language. Research has demonstrated that poor management of English in testing could result to under achievement of learners as well as the problem of differential item functioning. Testing in the medium of English in a plural society like Nigeria calls for caution in order to avoid scapegoating among the population of learners. Given the increasing use of testing for selection, placement and certification of learners in our society there is dire need for educators to be guided on the proper use of English language in testing subjects of the curriculum in our school system.

Theoretical Framework

Two theories have sustained the practice of testing. These are classical test theory and item response theory. Understanding of these theories is essential in knowing how language of test items can influence valid testing.

Classical Test Theory

Classical test theory (C.T.T) sustained measurement practice and testing up to the mid twentieth century. Nenty, (1998) reported that Charles Spearman is credited with the fundamental ideas of C.T.I. The theory postulates that the raw score of a testee (that is the count of the number of correct response) in a test or the observed score (X) can be decomposed into two orthogonal quantities- the true score component (T) and the error score component (E). Represented in a mathematical equation:

$$X = T + E \quad \text{(Equation 1).}$$

The true score represents the quantity of the variable or attribute under measurement, while the error score component represent contributions that arose from factors which have no bearing with the attribute under measurement. Sources of error variables include anxiety of the testee, announcement by invigilators and distraction from components are random. Random errors are those which can be eliminated by sampling in the testing process. C.T.T. proponents posit that to eliminate such errors testees should be given a long test length so that the errors can cancel out. The value of the true score component is dependent on prudent and specialized items writers.

Assumptions of C.T.T.

As a theory of measurement, C.T.T. has some basic assumptions. According to Mordi UU13), C.T.T. assumes that differences between total score in the test is a result of difference in their standing

in the ability under measurement. For instance, if Ola, the son of principal or lecturer has 75% and Jide the son of a farmer in a village school has 45%, in a subject like biology such difference represents their standing in biology ability. Nenty (1998) also stated that CTT assumes that the correlation between the true score component and the error score component of testees is

Criticism of C.T.T

C.T.T has been criticized on the following grounds. The first is the way it arrives at the performance of the testees. This is by comparing the testees score with the score of other testees *who took* the test. For example a testee with percentile score of 60 and a percentile rank of 99 in his/her group is described as being on top of the class. When the same student is in another group where a score of 60 is at the percentile rank of 75 he/she seems to be on top of the class. Secondly, item parameters such as difficulty and discrimination indexes vary according to the group that is used in establishing them. C.T.T is also criticized on the ground that it makes every error in testing under error score component.

Item Response Theory (IRT)

Item response theory, also known as item characteristic curve or latent trait theory⁷ is a more recent theory of measurement which addressed the errors in C.T.T. The theory postulated by Ferguson (1942) and popularized by Rasch (1960) and Lord (1980) postulates that latent trait is responsible for an individual's performance in a given test. Testing is a process of ascertaining the amount of latent trait that is possessed by an individual. During testing, there is an encounter between the latent trait demanded by the stimulus task and the latent trait possessed by the testee. The individual can only get the item correctly if the latent trait he possessed of the subject is well above that of the stimulus task. According to IRT an individual score (X) in a test can be decomposed into three orthogonal quantities viz: true score component, extraneous error component and random error component. This is represented in a mathematical equation below:

$$X_i = \theta_i + \lambda_i + e_i \quad \text{Equation 2}$$

Where:

X_i = observed score for the individual i , θ_i = true ability score component for the individual i , λ_i = systematic error score variable component e_i = Epsilon is the random error component of the score

Comparison of equation 1 and 2 will reveal that equation 2 which is the equation of IRT introduced the extraneous error component of the test score. This represents a major deviation from CTT. The extraneous error component is systematic and cannot be eliminated by sampling. Sources of extraneous error component are language of test items (in terms of their complexity) wisdom, and communication skills (Nenty, 1996). These traits are not normally distributed among population of testees. In testing, efforts should be made to correct for them during item development stage.

Assumptions of IRT

Item response theory has some basic assumptions among which two have been presented here. These are assumption of local independence, and the assumption of unidimensionality. The former states that test item should be constructed such that in a test length no item should provide a clue to another item. The latter states that test items should be constructed to measure a single trait. For instance, in testing learning outcomes in biology, test items

should be constructed to measure biology trait not English, if mathematics, it should be mathematics as learnt in biology. Violation of the unidimensionality assumption results in test items that differentially function for individuals of the same ability in the subject matter. Odili (2012) demonstrated that English language of test items is a source of violation of unidimensionality assumption when it is not managed during test construction.

Language of Test Items and Students' Performance

Research has implicated medium of instruction and testing as factor of students understanding and achievement. Igwebuikwe, Odili and Ajuar (2012) reported that senior secondary school students in Warri South Local Government of Delta State, Nigeria taught biology in pidgin English performed significantly better than their counterparts who were taught the same subject in formal English language. Kokakulah, Ustunluoglu and Kocakulah (2005) also demonstrated that Turkish children taught science using native language had better conceptual understanding compared with their counterpart taught in the medium of foreign language. Johnstone and Selepeng (2001) also reported that the level of complexity of English language is capable of affecting the performance of students in chemistry. They observed that the problem of complexity in English language in testing is more for students whom English is their second language. They provided an information processing model to explain the source of this complexity. The model is presented in figure I

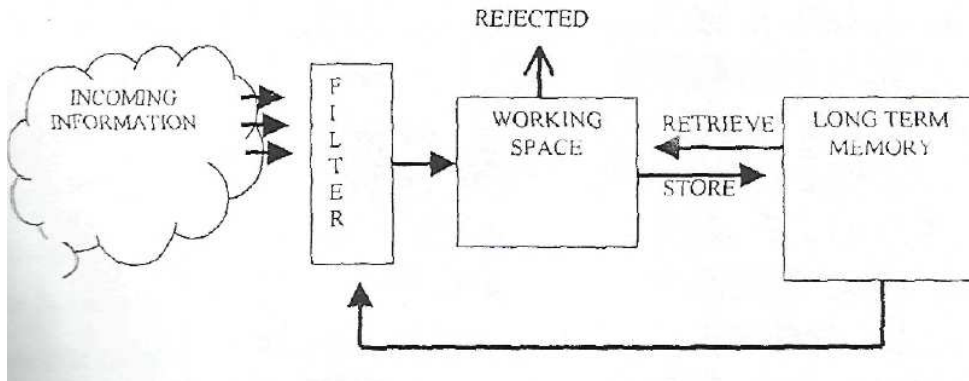


Fig 1: *Information Processing Model Source: Johnstone and Selepeng (2001): 23*

From figure I. incoming information represent stimulus task in the test. They are in form of questions, or statements. The filter represents the sense organs which are the gate way to the intellect. It includes the ear, eye, nostril, skin and tongue. Incoming information which are complex, unfamiliar and ambiguous would be considered "heavy" and may have difficulty in filtering through the senses organs and this has implication for what gets into the intellect. Psychology of learning taught us that familiarity of concepts is an important factor in learning, retention and transfer (Gagne, 1970). The working memory represents the short term memory where information processing takes place. How efficient it is in functioning is a function of how much information that gets to it from the filter. Like it is in modern computer parlance, garbage in, garbage out. Inefficient filtration will result in poor information getting through it, consequently poor processing and thus poor

discrimination among distracters of a multiple choice test. The result is poor test performance. For individuals whom English is their second language, the working space is divided into two sections. One half is used for thinking and translating information from the native language to English language before effective processing of information. Thus, standing in English language ability is determinant of the size of working space that is devoted to information processing in the medium of English language. For a learner that is highly proficient in English language his/her working space is comparatively higher than that with low proficiency in the English language. When test items are phrased in complex English language, a testee with high proficiency in English language will have advantage over his counterparts with low proficiency. Bernstein (1961) stated that children from high socioeconomic status (SES) have higher proficiency in English language compared with their counterparts from low SES. Obayan (1982) also stated that children in urban centres have higher proficiency in English language than their counterparts from the rural centres. Test items phrased in complex English language will present differential difficulty to testees depending on their SES and location. This is capable of defeating the goal of using education for achievement of egalitarian society.

The long term memory is a storage section in information processing. It retains processed information. During thinking or reasoning relevant stored information are retrieved from it into the working space for information processing. Efficiency of retrieval and storage is also a function of familiarity of information needed. Standing in English language ability influences the ease of storage and retrieval in information processing.

Effect of Language Complexity of Test Items on Performance of Testees

Cassel and Johnstone (1984), Odili (2012) have presented research findings on the effect of Complexity of English language of test items on performance. These are presented below:

1. Long sentence

The use of long sentence in stem and options of a multiple choice test reduced the number of testees who passed an item compared with when the item is phrased in short sentence. Cassel and Johnstone (1984) gave this example in chemistry.

A. When a metal Z was added to the sulphate of a metal X, the metal X was precipitated and there was no effervescence. When the test was repeated using the metal T in place of Z no reaction occurred. When a metal Y was added to a solution of the sulfate of X, a brisk effervescence occurred. Which one of the following is the correct order of decreasing activity (i.e the most reactive first) of the four metals? (47%).

The above question was rewarded as follows:

B. Three metal Z, T and Y were added to separate solutions containing metal X ions. Z precipitates X, T had no effect; and Y a gas was given off from the solution. Which one shows the metals in the correct order of activity (the most reactive first)? (67%).

Another example of effect of reducing the number of words in a test item in biology was reported by Odili (2012) below:

A. Which of the following diseases results from the deficiency of insulin (61%).

B. Deficiency of insulin causes (67%).

2. Negative Phrase

This is when stimulus tasks are presented in negative phrase using the words *except, does not, is not* etc. an example of effect of such phrase in performance of testees in chemistry is given by Cassel and Johnstone (1984) as follows:

- A. Which one of the following particles does not have the same number of electrons as a calcium ion? (24%).
- B. Which one of the following particles has the same number of electrons as a calcium ion? (80%).

The example in biology was given by Odili (2012) as follows:

- A. Terrestrial plants exchange gases through the following except (22%).
- B. Which of the following is used in gaseous exchange in terrestrial plants? (63%).
3. Use of unfamiliar word in key position

A key word in a test is that which gives direction to the kind of answer that is expected in a test. Unfamiliar key word will inhibit information processing compared with a familiar word. Below is an example given by Cassel and Johnstone (1984),

- A. Which of the following is a pungent gas? (56%).
- B. Which of the following is a choking gas? (63%).

Odili (2012) gave the following example in biology.

- A. Which of the following indicates that Euglena is a plant? (13%).
- B. Which of the following shows that Euglena is a plant (93%).

In the above examples, the numbers in parentheses represent the % of testees that passed the items.

Test items phrased in complex English Language has also been found to account for differential item functioning. Differential item functioning (DIF) is a condition in which individuals with the same ability in the subject matter do not have equal probability of getting an item correctly (Dogan, Guerrero, & Tasuoka, 2005). According to Odili (2010) biology multiple choice items used by West African Examination Council in Senior School Certificate Examination differentially functioned for students from high and low SES, urban and rural location, and gender. Testing Nigerian learners with test items which are not checked for complexity in English Language will impact negatively on the performance of testees, and as well as create differential item function. Given the high premium that is placed in paper qualification in Nigeria for selection, classification and placement there is absolute need for practitioners in the field of education to check test items for language complexity. This is essential in achieving test fairness in a plural society like ours.

Conclusion

The use of English Language as a medium of instruction and testing, in a society where English is used as a second language creates the dual problem of comprehension and achievement in testing. This problem is accentuated when efforts are not made to check complexity in English language used in instruction and testing. In a plural society like ours where English language ability is not uniformly distributed this will hinder the objective of using education as a tool for achieving egalitarian society. Difficulty of test items should arise from the level of cognitive domain that is tested not the complexity of the language for phrasing test items.

Recommendations

From the foregoing the following recommendations are made to guide testing in our school system.

1. Test item writers should avoid phrasing test items in long sentences with redundant non technical words in stem and distracters of a multiple choice test and as well as essay questions.

2. Item writers in testing should avoid the use of negative phrase such as does not, except do not etc. They introduce additional difficulty which is not of the subject matter.
3. It is necessary to avoid the use of unfamiliar words in key position so as to avoid inhibition of reasoning in the testees.
4. Examination bodies like West African Examination Council, National Examinations Council etc which engage in regional and nationwide testing should edit test items so as to eliminate such items that present unfairness on the ground of complexity of English language.
5. Given the emphasis in the use of mother tongue as a medium of instruction in our school system, government through the ministries of education should legislate against the use of complex English language in teaching and testing so as to use education for national integration and equal opportunity.

References

- Bernstein, B. (1961). Social class and linguistic development. In J. Floud, A. A. Halsey & C. A. Anderson (Eds), *Education and Society*. Glencoe, Illinois: Free Press.
- Cassel, J. R. T. & Johnston, A. H. (1984). The effect of language on student performance in multiple choice in Chemistry *Journal of chemistry education*, 61(7), 613-615.
- Gagne, r. M. (1970). *The conditions of learning* (2nded.) New York: Molt, Kmehart &, Winston, Inc.
- Igwebuike, T. B., Odili, J. N. & Ajuar, H. N. (2012). Medium of Instruction controversy in Nigeria: A test of the Efficacy of Pidgin for Teaching Biology. *American Journal of Scientific research*, 62-71.
- Johnston, A. H. & Selepeng, D. (2001). A language problem revisited. *Chemistry Education: Research and Practice in Europe* 3(1), 19-29.
- Lord, F. M. (1980). Applications of item response theory to practical testing problems. *Vocational Measurement*, 6, 39-48.
- Mordi, C. E. (2013). Understanding students' academic achievement (30 in the series of inaugural lectures of the Delta State University, Abraka, Nigeria, February 21).
- Nenty, H. J. (1996). Advances in test validation. In A. B. Gani & I. P. Odor (Eds.), *challenges of managing educational assessment in Nigeria*. Abuja, Nigeria: JAMB, NABTEB, NBEM.
- Nenly, H. J. (1998). Introduction to item response theory. *Global Journal of Pure anJ Applied Sciences*, 4(1), 93-100.
- Obayan, P. A. I. (1982). *Secondary English teaching*. Lagos: Macmillan.

Odili, J. N. (2012). Effect of Simplifying English Language of biology test items on test fairness. *Journal of Educational Assessment in African*, 7.

Odili, J. N. (2010). Effect of language manipulation on differential item functioning of test items in biology in a multicultural setting. *Journal of Educational Assessment in African*, 4.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: The Danish Institute for Educational Research.

Kocakulah, S., Ustunluoglu, E., & Kocakulah, A. (2005). The effects of teaching in native and foreign language on students conceptual understanding in science courses. *Asia-Pacific forum on science learning and Teaching* 6, Article 2.