Development of a Scale for Measuring Students' Testwiseness in Senior Secondary School Examination in Nigerian

Opesemowo, O. A.[1], Afolabi, E. R. I.[2], Awopeju, O. A.[3] & Oluwatimilehin, T. R.

1 Department of Educational Foundations and Counselling, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Nigeria. Tel: 07033403279, E-mail: ope4biz@gmail.com

2 Department of Educational Foundations and Counselling, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Nigeria. Tel: 08063218584, E-mail: eriafolabi@gmail.com

3 Department of Educational Foundations and Counselling, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Nigeria. Email: josyemus66@gmail.com

4 Department of Educational Foundations and Counselling, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Nigeria. Tel: 07034584009, Email: oluwatimilehint@gmail.com

Correspondence: Opesemowo, O. A., Department of Educational Foundations and Counselling, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Nigeria. Tel: 07033403279, E-mail: ope4biz@gmail.com

Abstract

The study developed a scale and generated items capable of measuring students' testwiseness, determined the construct validity and reliability of the scale. The study adopted survey research design. A total of 750 senior secondary students (SSS) participated in the study. They comprised 380 male and 370 female students ranging from 10-23 years old. Furthermore, 26.1%, 50.9%, and 22.9% were SSS 1, 2, and 3 students respectively. Additionally, 46.3% and 53.7% were students from private and public schools. Data were analyzed using reliability analysis and exploratory factor analysis. Cronbach's coefficient Alpha, Spearman Brown's and Guttmann's split–half coefficients were used to determine reliability of the scale. Inter–item correlation of the scale was carried out both on the initial and final items. Govaerts and Gregoire (2008) item reduction criteria were based on the use of corrected item-total statistics and Cronbach-Alpha-if-item-deleted approaches. Results showed that after item deletion and retention, 24-items emerged. Speed and eschew mistakes were the two factors found to underlie the construct of testwiseness. Reliability coefficients obtained from the 24–items were higher than the initial 32-items, each being significant at p<0.05. Cronbach alpha, Spearman Brown split-half and Guttman coefficient for the 32-items scale were 0.70, 0.68, and 0.74 while the final scale were 0.78, 0.77 and 0.76 respectively. These coefficients indicated there was an considerable improvement in the final scale over the initial. The study concluded that the scale was reliable and valid for measuring testwiseness of secondary school students in Nigeria.

Key Words: Tests, Testwiseness, Scale, Validity, Reliability

1. Introduction

Education is the process of learning, the bedrock to a nation's development and it is in continuum. It is the best legacy a nation can give her citizenry, an instrument use in awakening the child's professional training and

International Journal of Research

Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 05 Issue 19
August 2018

assisting him/her adjust adequately to his/her environment. However, the cost of it is an investment made by any nation in her children for harvesting a future that is responsible and productive. It can give rise to a well-functioning society. The importance of education cannot be over emphasized, as it is an impeccable instrument for the acquisition of knowledge and technical skills for development. In view of this, there must be teaching and learning and the teacher teaches the learners what they don't know but are expected to know. Furthermore, this process will be fragmental without getting a feedback that enables the teacher to appraise, judge and determine to what extent the learner have learnt over a period. This will then be used to make informed decisions on areas where learners need further attention or assistance. On the other hand, it is also in other to evaluate the teachers' method of teaching, one way by which this can be done is through testing the learners.

Tests are tasks presented to individuals that require them to respond to a set of items or questions within a specific period of time. It therefore appears to be a mechanism commonly used to measure learners' cognitive and non-cognitive ability of students. Also, it is a set of questions, statements or problems that can be used to measure what the learner have learnt over a period of time. There are different types of tests that serve different purposes, They includes intelligent tests, personality tests, aptitude tests and achievement tests e.t.c. Intelligent tests are structured to judge what an individual can do under favourable conditions, while trying to do his/her best. Personality tests are organized to measure the non-intellectual aspects or the characters of an individual psychological makeup. Aptitude tests are tests used to measure the ability of the learner. Achievement tests are structured to measure the level of skills and knowledge of content materials in a particular grade level via planned instruction such as classroom instruction. They are also used to measure the proficiency of what students have learnt over a period of time.

The score of these tests are used in educational system to ascertain the level of instructions students have acquired. A high level of this score (achievement test score) indicates mastery of a particular material while low level of achievement test score means inability or lack of adequate preparation of study materials. In classroom assessment where achievement tests are frequently used, students who do not have mastery of the content material use testwiseness while taking test/examination to get high score.

Testwiseness (TW) emanates from testees regular taking of examinations and can confer a significant advantage to the experienced testees over those that are not experienced (Mustapha, 2014). However, it is that quality possessed by a student at a particular period which enhances his/her performance in a test. Similarly, testwiseness is the ability of a test taker to answer correctly a particular test item or question without having adequate knowledge of the subject matter being tested. TW has been seen as the various sources of test variance. It was suggested as one of the sources of variance in educational test scores other than item content and random error (Thorndike, 1949; Millman, Bishop & Ebel, 1965). Specifically, due to testwiseness, the response set is known to influence the performance of students in objective tests (Ajeigbe & Afolabi, 2013).

It is important to distinguish testwiseness and informed guessing. Testwiseness is based on little or no content knowledge and is merely an attempt to select the answer correctly based on errors in the test construction. In contrast, making informed guessing requires the student to have some measure of content knowledge, enough at least to rule out some plausible distracters and reducing the number of possible answers from which a guess may be made (Mahamed, Gregory & Austin, 2006).

Particularly, several studies have recognized multiple choice items to be more susceptible to testwiseness clues (Katalin, 2000; Edwards, 2003). Multiple choice items contain numerous components (such as stem and alternatives) and hence clues which are more outstandingly seen by test takers. Basically, they are usually difficult to construct, easy to mark and is likely to suffer from shortcomings including testwiseness. Furthermore, teachers are not necessarily required to pilot their test in other to increase the reliability, validity and item characteristics of the test. Also, it is a subject capable of utilizing the characteristics and formats of the test and/or the test taking situation to receive a high score. Testwiseness is logically independent of the testee's knowledge of the subject matter for which items are supposedly measured (Millman et al., 1965). It is on this note that makes it imperative to develop a scale to measure students' testwiseness skills.

Scale is an instrument used by researchers to quantify human psychological responses to stimuli and it is developed to capture feelings, judgments, opinions and perceptions of an individual towards a particular construct such as testwiseness, self-concept e.t.c. Also, scale is a kind of response format presented to respondents in other to ascertain their choice of satisfaction to a particular item or statement. Additionally, it is a set of categories or

numeric values assigned to individuals, objects, or behaviour for the purpose of measuring variables. However, scale is develop to measure psychological constructs such as testwiseness, academic motivation, achievement motivation, test anxiety, self-actualization, inferiority complex, self-discipline, pride , introversion, extroversion, self-concept, study habit and humility e.t.c.

Scales differ from tests or examinations in the sense that the results of the scale unlike tests or examinations do not depict success or failure, strength or weakness of respondents but rather, they are used to measure the degree to which an individual demonstrate the characteristics of interest or opinion. For instance, scale can be developed to measure testwiseness for secondary school students or on any topics.

There are certain assumptions that should guide the development of an acceptable scale such as testwiseness. These are: validity, reliability, unidimentionality, and linearity. A scale is said to be valid when it measures what it is supposed to measure. Reliability of a scale is when it exhibits the consistency of what is purport to measure. Unidimensionality of a scale is when all the items in the scale are pointing towards the same construct. They must not have multiple interpretations and seen as measuring the same trait. Linearity is when the response set of the scale is increasing in sequence. In Nigeria, some researches have been conducted on testwiseness but none of these studies to the best knowledge of the researcher delved into development of testwiseness scale for secondary school students.

1.1 Statement of the Problem

Testwiseness is the ability of test takers to respond to test items correctly without having prior knowledge of the subject matter being tested. Although, it is possible to increase the overall test performance, which in turn makes it difficult to ascertain the true score and ability of the test takers since the scores will only reflect the testwiseness skills of the students . It is also a source of test variance among testees performance in tests or examinations. Furthermore, there is the tendency for students who possess testwiseness skills to have a meaningful advantage over their counterparts. Testwiseness can also be a positive ingredient in personal self-development. Knowledge of secondary school students' testwiseness skills in science, arts and social science subjects will facilitate its use.There are inadequacies in the existence of standard testwiseness scales that reflect the educational and cultural milieu of Nigerian students. Based on the above reasons, the specific purposes below were raised for the study.

1.2 Objectives of the Study

The specific objectives are to:
  i.   develop a testwiseness scale for public and private secondary school students in Osun State;
  ii.   determine the construct validity of the scale; and
  iii.   examine the reliability of the scale.

1.3 Research Questions

From the objectives this study, the following research questions were raised.
  i.   What items will be adjudged to measure students' testwiseness?
  ii.   What is the construct validity of the scale?
  iii.   What is the reliability of the scale?

2. Methodology

This study employed a descriptive survey research design. The population comprised Secondary School Students in Osun State, Nigeria. A total of 750 students participated in the study using multi-stage sampling technique. They comprised 380 males and 370 females with age range between 10-23 years. Furthermore, 26.1% of the samples were SSS 1, 50.9% were SSS 2 students, and 22.9% were SSS 3 students. In addition, 46.3% of the respondents were from private schools, while the remaining 53.7% were from public schools. From the three senatorial districts in

Osun State, three Local Government Areas (LGAs) were randomly selected. Five secondary schools were selected from each of the LGAs using stratified sampling techniques using school type as indices for stratification. The students were selected using stratified sampling techniques with age, class and gender as stata. The instrument for the study was Testwiseness Scale (TWS), whose items were generated from the literatures. This scale was developed with four dimensions sub-scales. The scale had 32 items comprising of 26 positive and six negative worded items. The items were subjected to expert review in Tests and Measurement for professional judgment on ambiguity, relevance and sentence structure. Thereafter, the TWS was administered on the students. The data were collected over two weeks. After scoring, the data were subjected to reliability analysis and exploratory factor analysis. Cronbach coefficient Alpha, Spearman Brown's and Guttman's split–half coefficients were used to determine the reliability of the scale. The inter–item correlation of the scale was carried out both on the initial 32 items and the final 24–item scales. Item retention decisions were based on the use of corrected item-total statistics and Cronbach Alpha if-item-deleted approaches.

3. Results

Table 1 shows the item statistics of the initial 32 items and final 24 items.

Table 1. Summary item statistics of initial and final scale items

	Initial Scale				Final Scale			
	Mean	Min.	Max.	Range	Mean	Min.	Max.	Range
Inter–Item Correlation	0.08	-0.22	0.41	0.63	1.30	-0.73	0.41	0.48

The table 1 showed that there were low but positive correlations between the items, indicating that the items were measuring different dimensions of testwiseness. The deletion of items having negative and low correlated item–total statistics in the initial scale items resulted to an improvement in the inter–item correlation of the final 24–items scale.

Table 2. Eigen values and total scale variance explained by the factors on the testwiseness scale

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.130	17.209	17.209	2.580	10.751	10.751
2	1.715	7.145	24.354	1.871	7.796	18.547
3	1.420	5.915	30.269	1.680	7.000	25.547
4	1.257	5.237	35.506	1.577	6.571	32.118
5	1.066	4.443	39.949	1.475	6.145	38.263
6	1.033	4.303	44.252	1.437	5.989	44.252
7	.977	4.070	48.322			
8	.939	3.911	52.232			
9	.930	3.877	56.109			
10	.912	3.799	59.908			
11	.880	3.666	63.574			
12	.845	3.522	67.096			
13	.798	3.324	70.419			
14	.784	3.268	73.687			
15	.751	3.128	76.815			
16	.723	3.014	79.828			
17	.710	2.960	82.788			

18	.693	2.888	85.676			
19	.663	2.761	88.438			
20	.621	2.588	91.026			
21	.607	2.528	93.554			
22	.574	2.390	95.944			
23	.497	2.069	98.014			
24	.477	1.986	100.000			

Extraction method: Principal Component Analysis.

From table 2, initial factor analysis using eigen-value greater than one suggested the presence of six factors in the scale, accounting for 44.25% of the total scale variance.

Figure 1. Scree plot of testwiseness scale
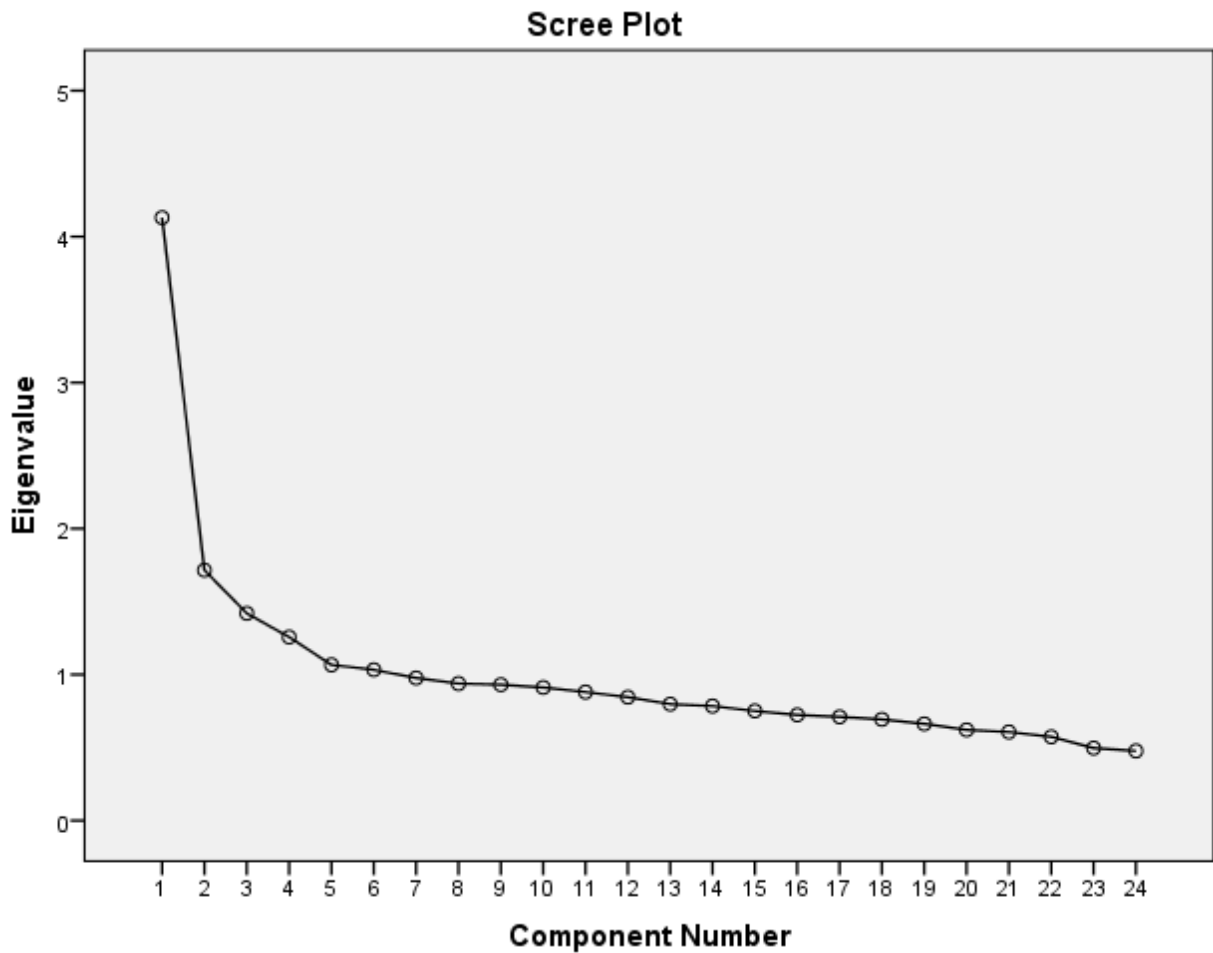
From the result in Figure 1 above, the plot showed eigenvalue generated for the factor loading of the items of testwiseness scale. This showed six factors with eigenvalue greater-than-1 stopping at eigenvalue of 4.13 in the upper left and descending part of the plot before the eigenvalues start to level off parallel to the axis. Also, a total of 24 items emerged from the scale. Hence, the scree plot suggested two factors.

Table 3. Coefficients of reliability of the TWS

Scale Items	Guttman Coefficient	Cronbach Alpha	Spearmna Brown Split Half	Guttmann Split Half Coefficient	Common inter-item Correlation	Item Variances
N=32	0.738	0.704	0.675	0.668	0.076	1.513
N=24	0.764	0.777	0.766	0.765	0.130	1.390

Table 3 showed that the reliability of the final 24-item TWS was consistently greater than that of the initial 32 item scale in each of the three reliability measures, namely Cronbach alpha ($r = 0.777$), Guttmann's split-half ($r = 0.765$) and Spearman Brown's ($r = 0.766$) were all greater than the generally acceptable rule of 0.7 reliability. Hence, the Cronbach alpha, Guttmann's split-half and Spearman Brown's were all reliable. Invariably, the item variances of 1.513 of the initial items reveal the homogeneity of the items that had relatively lower mean scores. This was also corroborated by the inter-item correlation values of 0.076 and 0.130 for the initial and final scale items respectively. Therefore, the TWS is considered reliable in terms of the internal consistency of its items.

Table 4 shows the final scale items of the TWS.

Table 4. Final scale items of TWS

Old s/no	New s/no	Items
4	1	I check all available options carefully before picking the correct answer.
5	2	I review the task requirement of an item before providing a response.
6	3	I guess when I don't have an idea of the question.
7	4	I work as rapidly as possible when taking tests.
8	5	I answer easy questions first in a test or examination.
9	6	Before marking the correct option, I eliminate options I know can't be the correct answer.
11	7	I choose between options that could possibly be the answer.
14	8	I answer all questions no matter how difficult they are.
15	9	I allocate time to each question based on the time allowed in the test.
16	10	I revise answers immediately after answering each question.
17	11	I usually use all available testing time.
18	12	I answer all questions even when I have no knowledge of the question.
19	13	I revise all answers before submitting my script.
20	14	I use prior questions to get help in answering subsequent questions.
21	15	I avoid last minute changes while taking tests or examination.
22	16	I use any time remaining after completion of the test to recheck my answers.
23	17	I use time remaining to look out for omitted questions.
24	18	I move quickly to the next item when I have difficulty with an item.
25	19	I utilize relevant content information in other items and options.
26	20	I normally establish an answer protocol for a test before responding to any item.
27	21	I mark omitted items to ensure that I do not forget to get back to them.
28	22	Even when an initial option is plausible, it is still desirable to check all the options.

| 29 | 23 | I'd rather guess if there is no penalty for guessing rather than omit a difficult item. |
| 30 | 24 | I do not start to answer the items in a test without quickly going through to have a proper feel of the test items. |

4. Discussion

The initial items generated for the testwisenes scale was 39, 7 items were deleted by experts in the field of Tests and Measurement due to inappropriate wordings, ambiguity, multiple interpretation and item length. The reduction gave rise to 32 items on the scale; the items generated were subjected to Goveart and Gregoire (2008) item reduction criteria in other to ascertain items adjudged to have measured students' testwiseness, this led to deletion of 8 items. The constructed and validated scale in the study gave rise to 24 final items on the scale. Two factors were found to underlie students' testwiseness, these are speed and eschew mistake. A student with great speed will manage his/her testing time wisely by consciously pacing the time allotted for the test and will even maximize their performance in the test. This was supported by Mustapha (2009) that time management is a technical skill which very students need to possess in order to improve their performance. The needs for testees to be conversant with the exigency of time thus, helping them maximize their performance in tests/examinations. It thus implies that students should be conscious of time in executing academic activities so that they can perform brilliantly.

Research on testwiseness by Watter and Siebert (1990) showed that testwise students with appropriate time management skills receive more points than they would have predicted, feel more relaxed and better organized while taking tests/examination and are able to complete tests in the allotted time, manage their test time appropriately, and seldom leave out important information from answers. Since testwiseness is thought to comprise various skills which includes time management skills, answering all questions and making sure that one fully understands the directions of the tests/examinations (Millman et al., 1965). Some students have good understanding of the subject-matter, good reading ability but are unable to manage their testing time wisely (Mustapha, 2012).

The second factor that found to underlie students testwiseness was eschew mistake. This concerns the degree to which students avoid mistakes during tests or examination. The main concern here is that student who read instructions carefully, who ask questions from examiner for clarification when necessary and those review task requirement of an item before providing response are of a better performance than students that doesn't possess such skills.

Benson (1985) employed first exploratory factor analysis and then confirmatory factor analysis to assess the dimensionality of testwiseness. He obtained four factors which he temporarily labelled as: (i) thoroughness; (ii) preparation; (iii) achievement motivation; and (iv) perseverance.

The convergent validity of the Testwiseness Scale (TWS) with Achievement Motivation Inventory (AMI) was correlated using mean score and standard deviation score. The result suggested that the correlation coefficient between the two scales, TWS and AMI, was 0.545, which was significant at 0.01 level of significance. Since the AMI is a widely used scale with a Cronbach's alpha (α) reliability of r = 0.749, a high and positive correlation with it by the TWS thus establishes the validity of the latter. That is, the TWS does measure testwiseness of secondary school students in Osun State, Nigeria. Likewise, examining the correlations among five sub-scales (longer correct alternatives, stem-option clues, specific determiners, grammatical clues, and similar options), Diamond and Evans (1972) reported low correlations ranging from 0.02 to 0.33 among five sub-scales. Their findings suggested that the use of testwiseness strategies was actually a set of several skills rather than a general skill. While validating his instrument to measure testwiseness through the multitrait-multimethod approach (Campbell & Fiske, 1959), Millman (1966) incidentally found that the instrument consisting of four separate sub-scales (guessing, deductive reasoning, stem-option clue, and inconsistencies in the response options) did not quite hold up. The mean correlation of these sub-scales for monotrait-monomethod was 0.53 for the high school samples and 0.38 for the college students.

The results of the study also showed that the 24 items on the TWS were reliable to measure students' testwiseness. The reliability of an instrument is the consistency with which it could obtain responses when investigated via internal consistency method or when tested for stability.

The estimated reliability coefficients of TWS (Cronbach Alpha = 0.777, Guttman Coefficient = 0.764, Spearman Brown Split Half = 0.766 and Guttmann Split Half Coefficient = 0.765) were acceptable. Previous

research on testwiseness by Allan (1992) the internal consistency was low for each of the subtests, ranging from 0.04 (opposite options) to 0.24 (stem option cues), and also low (r = 0.37) for the entire test of testwiseness. However, the point-biserial coefficients for the majority (84.2%) of the items were acceptable, ranging from 0.20 (item 21) to 0.67 (item 4). As suggested by Rogers and Bateson (1991a), the low internal consistency may be attributed partly to the small number of items involved and partly to the student's "educated guessing" applied in response to a question which the students, will able to eliminate one or two options as incorrect, did not know the answer from among the remaining options.

Conclusion

The study concluded that the testwiseness scale was considered to be reliable and valid for secondary school students in Nigeria. Also, from this study, it was found out that speed and eschew mistake were the most prominent components of testwiseness. So, teachers, school counsellors and other stakeholders in education can make use the scale in helping students to be better off in their academic performance.

References

Ajeigbe, T. O., & Afolabi, E. R. I. (2013). A study of item response and anchor bias in economics objective tests among senior secondary school students in Osun State. *Academic Journal of Interdisciplinary Studies, 2*(3), 173-179.

Benson, J. (1985). *Distinguishing test-wiseness strategies among, ethnically diverse elementary students.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Cambell, D. T., & Fiske, D. W. (1959). *Convergent and discriminant validation by the multirait-multimethodmatrix.* Psychological Bulletin.

Diamond, J. J., & Evans, W. J. (1972). An Investigation of the cognitive correlates of test-wiseness. *Journal of Educational Measurement*, *9,*145-150.
Edwards, B. (2003). *An investigation of factors contributing to a reduction in race-based subgroup differences on a constructed response paper-and-pencil test of achievement.* Unpublished PhD. dissertation. Office of Graduate Studies of Texas A & M University.

Govaerts, S. O., & Gregoire, I. (2008). Debt and construct validation of academic emotions scale. *International Journal of Testing, 8*(1), 34-54.

Katalin, B. (2000). Reflections on the test-taking strategies of 7th and 11th grade Hungarian students of English. *ELT Journal, 7*,111-118.

Mahamed, A., Gregory P. A., & Austin Z. (2006). Test-wiseness among international pharmacy students. *Am J Pharm. Educ, 70*(6), 131-140.

Millman, J. (1966). *Test-wiseness in taking objective achievement and aptitude examinations* (Final Report). Princeton, NJ: College Entrance Examination Board.

Millman, J., Bishop, H., & Ebel, R. (1965). An analysis of test-wiseness.*Educational and Psychological Measurement*, *25*, 707-726.

Mustapha, A. Y. (2014). Relationship between secondary school students' level of test-wiseness and their performance in mathematics in Bauchi metropolis of Bauchi State, Nigeria: *IOSR Journal of Humanities and Social Science* (IOSR-JHSS), *19*(3), 41-44.

Mustapha, A. Y. (2012). *Analysis of psychometric properties of junior secondary school certificate mathematics examinations in Plateau State*. Unpublished PhD. Thesis, University of Jos, Nigeria.

Mustapha, A. Y. (2009). A survey of secondary school teachers' level of possession of test-taking skills in Jos metropolis of Plateau State. *Journal of Educational Technology and Instruction* (JETI), 67-70.

Muthee, J. M., & Thomas, I. (2009). Predictors of achievement motivation among Kenyan adolescents. *The Psychespace, 3*, 39-44.

Rogers, W. T., & Bateson, D. J. (1991a). The influence of testwiseness on performance of high school seniors on school leaving examinations. *Applied Measurement in Education*, *4*, 159-183.

Thorndike, E. L. (1951). Reliability in E. F. Lindguist (Ed.), *Educational Measurement* 560-620. Washington, DC: American Council on Education.

Watter, T., & Siebert, A. (1990). *Students' success: how to succeed in college and still have time for your friends.* New York: Holt, Rinehart and Winston, Inc.