EFFECT OF TEST ITEM COMPROMISE AND TEST ITEM PRACTICE ON VALIDITY OF ECONOMICS ACHIEVEMENT TEST SCORES AMONG SECONDARY SCHOOL STUDENTS IN CROSS RIVER STATE, NIGERIA.

[1]**Idika, Delight. O.**
(delightoidika@yahoo.com); 08034635424
[1]Department of Educational Foundations, Faculty of Education,
University of Calabar, Calabar.

[2]**Shogbesan, Y. O. and** [3]**Ogunsakin, I. B.**
(yusufolayinka813@yahoo.com); 07032282769
(sakinbamikole@yahoo.com); 08039215009
[2&3]Department of Educational Foundations and Counseling, Faculty of Education,
Obafemi Awolowo University, Ile-Ife.

ABSTRACT

Every examiner or individual using test scores should be confident that the scores obtained by a test taker is a true indication of that person's level of knowledge or ability on the construct of interest with a proper guide against factors that can lead to score invalidity. However, this study determined the effect of test item compromise and test item practice on Economics Achievement Test Scores among secondary school students in Cross River State. It also examined whether test item compromise and test item practice affected the validity of test scores obtained in the Economics Achievement Test among secondary school students in the state. A quasi-experimental research design was adopted for the study. The population of the study consisted of all secondary schools in the 18 Local Government Areas (LGA) of Cross River State. The sample consisted of 90 SS2 Economics students randomly selected in the three Secondary Schools used for the study which was carried out in an intact classroom. There were three groups and 30 respondents were randomly assigned to each group. The three groups were compromise group (E1), practice group (E2) and control group (C1). The instrument used for the study was an Economics Achievement Test (EAT) developed by Shogbesan (2017) which consisted of 25 items of various formats with a liability index of 0.68. The EAT was administered to the three groups with 13 items exposed as treatments to E1 for the students to be familiar with some of the test items a few minutes before the test, while the practice group was given the 13 test items to practice with the help of the researchers and research assistants who are Economics teachers in the school. The results indicated among others that students' scores were inflated on compromised and practiced test items which contributed to the score invalidity. It was recommended among others that security of test items should be considered vital before and during test administration process.

Keywords: Test compromise, Test Item Practice, Score Validity and Economics Achievement Test

INTRODUCTION

Test refers to a method that is used to determine a student's ability to complete certain tasks or demonstrate mastery of a skill or knowledge and is meant to elicit information about a latent ability of an individual with respect to a particular variable of interest. Exploring the definition of test as given by Joshua (2005), it is evident that test is seen by some people only as an instrument that is useful in systematic measuring of a sample of behaviour whereas some other persons view it as both an instrument and a procedure for applying the instrument. However, be it as it may, valid test scores as an accurate reflection of test-takers' performances are becoming a more important currency throughout peoples' lifespan (Zara & Pearson, 2006).

Tests are meant to elicit information about a latent ability of an individual and to provide evidence so that educational decisions can be made about the individual. These decisions when made provide information about students on whether they have reached a particular level of skill and knowledge or not. It may help us evaluate a teaching programme or to make decisions about the next aspect of teaching for particular students. Although, several schools of thought have argued for and against the use of tests, yet, it has been the best available and mostly used instrument of measurement (Afolabi, 2012). The arguments against the use of tests may be as a result of various factors that affect test score reliability, validity and usability. This effect as it relates to the test itself can be as a result of different standard, motivational factors, familiarity with test items, bias, cheating and other related examination practice.

Specifically, when test takers are familiar with the test items, it may affect positively their scores on such test. The implicit hypothesis above being that prior knowledge of specific testing content may cause test-takers to receive inflated scores on the construct of interest as compared to their actual competence level, which according to Zara and Pearson (2006) will cause some amount of score invalidity. Also, according to Cannell (1988), lax security of test regardless of the stakes corrupts test scores. Then, it may be true that item exposure leading to compromise often presents a more tractable set of score invalidity issues. Cannell (1988) further cited educator dishonesty and lax security in test administrations as the primary culprits in the Lake Wobegon Effect, also known as "tests score inflation" or "artificial test score gains". Ojerinde (2015) also asserted that a breach of examination security is a major problem threatening the Joint Admission and Matriculation Board (JAMB) and other public examination bodies in Nigeria. He stressed further that examination insecurity manifests in various forms such as impersonation, swapping examination document, spying /coping from prepared answers, use of unauthorized calculator or similar electronic devices, smuggling out of question papers/ answer sheets among others. The menace of examination malpractice has impacted negatively on the integrity of public examinations over the years. Due to the increased concern on test security issues, the JAMB is now using compulsory CBT as a measure to combat examination insecurity in Nigeria. This is because most of the cases of examination malpractice are done with the paper - pencil testing mode (Idika, 2012; Ojerinde, 2015)

Obviously, the increased focus on test security issues by test developers presumes that the examinations have already been planned and developed to meet traditional validity concerns, since establishing test validity is an ongoing process and entails gathering different kinds of evidence from test content validity through construct validity to consequential validity evidence. Hence, to validate, interpret or use test scores is to evaluate the plausibility of the claims based on the scores. Kane (2013) suggested that the claims based on the test scores be outlined as an argument that specifies the inferences and support assumptions needed to get from test responses to score-based interpretations and uses. Therefore, these concerns of test security and their impact on validity as actualized in test takers achieving test results not actually reflective of their trait levels are well-placed with various suggested ways. These ways may be used to address this concern using psychometric research in the area of item exposure issues (see, e.g. Davey & Parshall, 1995; Stocking & Lewis, 1995; Sympson & Hetter, 1985; Linden & Veldkamp, 2005). This is because item exposure has been linked probabilistically to item compromise i:e the more items have been administered, the higher the likelihood that they may be compromised leading to the higher likelihood that test-takers may enter a testing opportunity with prior knowledge of specific test content (Zara & Pearson,2006).

Moreover, test-takers having prior knowledge of specific test content is an indication that the test items have been compromised either as a result of frequent use of test item in testing (repetition), item cloning, and item rotation, item over use or test item leakage or breaches. The test item compromise may be caused by educators themselves as humans as the rest of us; because some of them cheat and not all of them manage to keep test materials secure, even when they aren't cheating. Therefore, item compromise can be defined as being occurring when evidence exists that an item performance has changed during some defined time span and it is reasonable to believe that the performance changes are due to its content having been distributed beyond defined valid usage boundaries or due to over exposure to test takers (Zara & Pearson, 2006). From the above definition, it can be deduced that test compromise occurs when item compromise leads to improperly inflated scores. Compromised test items security refers specifically to a situation in which test-takers have access to test questions prior to completing the test (Drasgow, Nye, Guo, & Tay, 2009; Lievens & Burke, 2011; Naglieri et. al., 2004; Tippins, 2009).

Score inflation results when students' scores on tests increase but the increase does not reflect any genuine improvements in learning i.e. the instrument being used to measure learning acquisition and growth is providing a false reading because the testing design or processes are flawed or educators are inadvertently or intentionally inflating students' scores (Maynes, 2013). Educators are inadvertently or intentionally inflating students' scores when they provide the additional instruction and attention through coaching to students, teach to the test items, when test questions can be made easier, when students are given extra time to complete tests, practice of test item with the intention to improve

and succeed on the tests. These educational practices mentioned above can contribute to score inflation, and although some may be sanctioned, or even encouraged by school administrators, many are generally observed as cheating.

Educational practice such as teaching to the test might mean that teachers *re-allocate* time and *re-align* priorities in order to spend more time on content likely to be covered in the tests, thus, narrowing curriculum. They may also *coach* students in test-taking skills (Cizek, 1998; Koretz *et. al.*, 2001; Koretz, 2005; Popham, 2002; Smith & Rottenberg, 1991; Stecher, 2002). However, Coaching is never a desirable approach when it focuses on substantive aspects of tests, (e.g. if teachers provide students with advance access to tests, or provide inappropriate ASSISTANCE during the examination session). It is however desirable when teachers help students to develop test-taking tricks (*e.g.* how to recognize distracters in multiple choice tests, while not necessarily learning how to recognise the correct answers); they are engaging in "non-substantive" coaching (Koretz, 2005; Popham, 2002).

However, as suggested by Popham (2002), that for teachers to derive maximum instructional dividends from their classroom assessment, they must construct their assessment before instructions and continually ask themselves the question; can I teach what this assessment instrument measures? Hence, they will appropriately "teach to the test" without improper inflation of test scores. This will not only raise performance on a test, but also raises students' preparation-to-post preparation mastery of the assessment domain being tested. Also, the teacher should avoid the bad practices associated with "teaching to the test". Most importantly, the teacher should not direct instruction specifically towards the actual items on the test itself; rather he/she should endeavour to teach to the content represented by the test as outlined in the curriculum (Shogbesan & Faleye, 2016).

Moreover, when teachers often engage in practice of test items to provide intensive instruction and academic support to a group of students who are deemed most likely to improve their scores enough to meet expected benchmarks for improvement, technically this will have positive impact on the students' performance from just below to just above the cut-off score for "proficiency" on a test and avoid negative consequence which may raise fairness and bias issues as it relates to the test scores.

This is because the learning needs of other students in the class may be neglected. Consequently, tests should be fair, reliable, and valid for better decision making. To maintain these characteristics, tests need to be secured because decisions based on scores affected by cheating (compromised/practiced) are unacceptable. Conversely, with the daily access to the item pool, security becomes a real concern because they may be compromised or practiced because test takers may memorize blocks of test items and share these items with future test takers or practice them extensively. As a result, individuals with prior knowledge of some items may use that information to inflate their test scores (McLeod & Schnipke, 1999).

Test security is an important element of any examination programme, whether the examination is administered as Paper-Pencil Tests (PPT) or Computer Based Tests (CBT) (Ojerinde, 2016). An additional security risk of any administered test is item exposure and test content security with relative difference in the risk of occurrence from PPTs to CBTs. The various test navigation and presentation strategies made possible only through Computer Based Tests can either mitigate item exposure concerns as tests items are stored electronically and can either be pre-selected by the test sponsor for a given examination or randomly pulled on test day.

Therefore, in the Computer-Based test administration strategy, there is an improved tests security resulting from electronic transmission and encryption for total eradication of breaches of examination security (Ojerinde, 2014 as cited in Idika, 2015). Some researchers advocate frequently changing the item pools to lessen the security problem (McLeod, et. al., 1999). However, just with the addition of adaptive testing, the demand on item writers has already increased.

Item writers now introduce cloning of test items to boost item generation. Cloning as the name suggests entails having a set of test items that are similar in terms of psychometric properties and content area. That is to say that there will be an original question with three or four variations testing the same thing. Hence, options are also cloned to help guard against breaches of examination security (Ojerinde, 2015). This is because asking item writers and test developers to provide even more items may degrade some of the measurement properties of items.

However, there is a strong need for a quality control tool to measure the "freshness" of item pools so that compromised items may be removed more efficiently. As with access to outside resources, having access to test questions before officially taking the test could undermine test validity (Lievens & Burke, 2011) by allowing candidates to become more comfortable with test content, and to memorize answers in advance of the test (Tippins, 2009). Hence, rather than replacing an entire item pool, the flagged items may be removed and replaced with fresh (secure) items (McLeod, et. al., 1999). As more items on a tests are administered to more examinees for more reasons, the validity of individuals' test scores becomes more important, and it is incumbent on the examiner or individuals using test scores to be confident that the score proffered by a test taker is a true indication of that person's level of knowledge, skill, or ability on the construct of interest. Much of the current research designed to address concerns related to score invalidity issues does not address item compromise and test item practice per se, but majorly focuses on item exposure issues and cheating hence, this study.

The specific objectives of the study are to;

(a)   Determine the effect of test item compromise on economics achievement test scores among secondary school students in Cross River State.

(b)   Determine the effect of test item practice on economics achievement test scores among secondary school students in the state and;

(c)   Examine whether test item compromise and practice affect the validity of test scores obtained in the economics achievement test among secondary school students in the state

To achieve the stated objectives, two research questions were answered and three hypotheses were raised and tested respectively. They are;

Research Questions:

i.   Do students have better performance on compromised test items than uncompromised test items on Economics Achievements Test (EAT)?

ii.   Do students have better performance on practiced test items than unpractised test items on Economics Achievements Test (EAT)?

Hypotheses:

i.   There is no significant difference in the performance of students in the compromised and uncompromised test items on EAT.

ii.   There is no significant difference in the performance of students in the practiced and unpractised items on EAT.

iii. There is no significant difference between performance of students in the compromised and unpractised items on EAT.

iv.   There is no significant difference between the performance of students across the test item compromised, test item practiced experimental groups and control groups when moderating effect is controlled on the validity of test scores obtained on EAT.

Method

The research design adopted for this study was Quasi-experimental design. The population of the study consisted of all secondary schools in the 18 Local Government Areas (LGA) of Cross River State. The sample consisted of 90 SS2 Economics students randomly selected in the three Secondary Schools used for the study which was carried out in an intact classroom. Thirty (30) of the total respondents were randomly assigned to three (3) groups. The three groups were compromise group (E1), practice group (E2) and control group (C1).The three groups were not pre-tested but exposed to treatments with control group exposed to placebo treatment. The test items were exposed to the compromise group for the students to be familiar with some of the test items a few minutes before the test, while the practice group was given the test items to practice and work on, to ascertain understanding of each of the test items with the help of the researchers and research assistants who were Economics teachers in the schools. The control group was not given any treatment and served as total uncompromised/unpractised group for the study. The results of the three groups were

used to determine the effect of test item compromise, test item practice on score validity. The major instrument was an Economics Achievement Test (EAT) adopted from Shogbesan (2017), It consisted of 25 items with a reliability index of 0.68.The 25 test items of the EAT were divided into (5) sections with each section consisting of 5 items each of true /false, short answer, completion, multiple - choice and essay test item format. The EAT test blueprint is shown in the Table 1 below:

TABLE 1

Test Blueprint of test items format of the Economics Achievement Test

| | Contents | TEST-ITEMS FORMAT(Numbers) | | | | | |
		True-False	Short-Answer	Completion	MCQ	Essay	Total
1	Definition and scope of Economics	A1,A3		C3,C5	D1,D3, C5	E2	8
2	Basic tools for Economics analysis	-	B1,B2,B3,B4	C2,C4		E3	7
3	Basic economic problems of the society	A5				E1	2
4	Production	A2,A4		C1	D2	E5	5
5	Business organizations		B5		C4	E4	3
	TOTAL	5	5	5	5	5	25

NB: A- True/False Items, B- Short-Answer Items, C- Completion tests Items, D- Multiple-Choice Items and E- Essay items, MCQ – Multiple-Choice Question.

However, for the purpose of this study, 13 test items were compromised/ practiced for the 2 experimental groups (compromised and practiced group) respectively while 12 test items remained uncompromised/practiced. Also, the 25 items of the EAT were administered uncompromised/unpractised to the control group. The research questions were answered using Mean, Standard Deviation, and Coefficient of Variation while the hypotheses were tested using independent samples t-test, paired samples t-test and ANCOVA.

Results

Research Question 1: Do students have better performance on compromised test items than uncompromised test items on Economics Achievements Test (EAT)?

To answer this research question, the scores of the students on the compromised test item and the uncompromised test items on the EAT were obtained and were subjected to descriptive statistics using mean, Standard deviation and coefficient of variation. The result is shown in Table 2

TABLE 2

Mean and Standard Deviation of scores obtained in the compromised and uncompromised tests items on the EAT.

Test Items	Sample size		N	$\overline{X}$	S.D	C.V
Compromised test items	30		13	10.97	5.42	0.49
Uncompromised test items	30	40	12	7.73	3.64	0.47

NB: N=Number of items, $\overline{X}$ = Mean score, S.D= Standard Deviation of scores, C.V=Coefficient of Variation of scores

From Table 2 above, the mean score of students in the compromised test items is 10.97 (S.D=5.42) while the mean score of students in the uncompromised tests items is 7.73 (S.D=3.64). This reveals that performance of students in the compromised test items is greater than the performance of students in the uncompromised tests items of the EAT. However, the variation in performance of students in both the compromised test items (C.V=0.49) and uncompromised test items (0.47) is similar across the respondents with a minimal difference in C.V of 0.02.

Research Question 2: Do students have better performance on practiced test items than unpractised test items on Economics Achievements Test (EAT)?

To answer this research question, the scores of the students on the practiced test items and the unpractised test items on the EAT were obtained and were subjected to descriptive statistics using mean, standard deviation and coefficient of variation. The result is shown in Table 3

TABLE 3

Mean and Standard Deviation of scores obtained in the practiced and unpractised tests items on the EAT.

Test Items	Sample size	N	$\overline{X}$	S.D	C.V
Practiced test items	30	13	11.73	8.05	0.69
Unpractised test items	30	12	8.27	4.26	0.52

NB: N=Number of items, $\overline{X}$ = Mean score, S.D= Standard Deviation of scores, C.V=Coefficient of Variation of scores

From Table 3, the mean score of students in the practiced test items is 11.73 (S.D=8.05) while the mean score of students in the unpracticed tests items is 8.27 (S.D=4.26). This reveals that performance of students in the practiced test items is greater than the performance of students in the unpracticed tests items of the EAT. Furthermore, the variation in performance of students in both the practiced test items (C.V=0.69) and unpracticed test items (0.52) is not similar across the respondents with a 0.17 difference in C.V.

Research Hypothesis 1: There is no significant difference in the performance of students in the compromised and uncompromised test items on EAT

To test this hypothesis, the scores of the students on the compromised test items and the uncompromised test items on the EAT were obtained and were subjected to paired t-test. The result is shown in Table 4

TABLE 4
Paired t-test of scores on compromised and uncompromised test items

Tests items	Sample size	N	$\overline{X}$	df	t-value	p-value
Compromised test items	30					
		13	10.97	29	4.657	0.000*
Uncompromised test items	30					
		12	7.73			

NB: N=Number of Items, $\overline{X}$ = Mean score, p-value= Significant value
* Significant at 0.01.

From Table 4 given that $(t_{29=4.657}, p<0.01)$, it can be concluded that there exists a significant difference in the performance of students on the compromised and uncompromised tests items on the EAT.

Research Hypothesis 2: There is no significant difference in the performance of students in the practiced and unpractised items on EAT.

To test this hypothesis, the scores of the students on the practiced test items and the unpractised test items on the EAT were obtained and were subjected to paired t-test. The result is shown in Table 5

TABLE 5
Paired t-test of scores on practiced and unpractised test items

Tests items	Sample size	N	$\overline{X}$	df	t-value	p-value
Practiced test items	30			29	3.970	0.000*
		13	8.05			
Unpractised test items	30					
		12	4.26			

NB: N=Number of Items, $\overline{X}$ = Mean score, p-value= Significant value
* Significant at 0.01.

From Table 5 given that $(t_{29=3.970}, p<0.01)$, it can be concluded that there exists a significant difference in the performance of students on the practiced and unpracticed tests items on the EAT.

Research Hypothesis 3: There is no significant difference between the performance of students in compromised and practiced group of the EAT items.

To test this hypothesis, the scores of the students in the two groups on the compromised test items and the practiced test items on the EAT were obtained and were subjected to independent t-test. The result is shown in Table 6.

Table 6: Independent t-test of scores on compromised and practiced test items

Tests items	Sample size	N	$\overline{X}$	df	t-value	p-value
Compromised test items	**30**			**58**	**4.33**	**0.667**
		13	10.97			
Practiced test items	**30**					
		13	11.73			

NB: N=Number of Items, $\overline{X}$ = Mean score, p-value= Significant value

From Table 6 given that ($t_{58=\ 4.33,}$ p>0.01), it can be concluded that there does not exist a significant difference in the group performance of students on the compromised and practiced tests items on the EAT.

Research Hypothesis 4: There is no significant difference between the performance of students across the test items compromised, test items practiced experimental groups and control groups when moderating effect is controlled on the validity of test scores obtained on EAT.

To test this hypothesis, the scores of the students in the test items compromised, test items practiced experimental groups and control groups on the EAT were obtained and were subjected to ANCOVA using the uncompromised, unpractised test items in the three groups used as moderating effect. The result is shown in Table 7

TABLE 7
Summary of ANCOVA result of difference between the performances of students across the test items compromised, test items practiced experimental groups and control groups when moderating effects are controlled.

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	1623.133[a]	3	541.044	30.883	.000
Intercept	24.306	1	24.306	1.387	.242
Uncompromised & Unpracticed score	1609.644	1	1609.644	91.879	.000
Group	73.039	2	36.519	2.085	.131
Error	1506.656	86	17.519		
Total	14397.000	90			
Corrected Total	3129.789	89			

a. R Squared = .519 (Adjusted R Squared = .502)

Table 7 reveals that there is no significant difference in the scores obtained across the test items compromised, test items practiced experimental group and control groups on the EAT after using the uncompromised, unpractised test items in the three groups used as moderating effect (F_2=36.52, P>0.05).

Discussion

The results of this study show that performance of students in the compromised test items is better than the performance of students in the uncompromised test items of the Economics Achievement Test (EAT) although with a minimal variation in performance of students in both the compromised test items and uncompromised test items. Similarly, this study found out that performance of students in the practiced test items is better than the performance of students in the unpracticed test items of the EAT with a significant variation in performance of students in both the practiced test items and unpracticed test items. This finding is in agreement with the study carried out by Jurich, DeMars, & Goodman (2012) where they investigated the impact of compromised anchor items on IRT equating under the non-equivalent anchor test design and found that cheating artificially increased the equated scores of the entire examinee group that was administered in the compromised form.

Also, this study concluded that there exists a significant difference in the performance of students on the compromised and uncompromised test items on the EAT as well as a significant difference which exists in the performance of students on the practiced and unpracticed test items on the EAT. These findings imply that the nature of compromise and practice test items produced an inflated score which can be described as cheating because it gives an unfair advantage to the test takers. The inflated score (produced as a result of test item compromise and practice) would essentially be a misrepresentation of that individual's performance, thus, yielding an inaccurate estimate of performance (Royal & Puffer, 2012). This assertion from the finding is in agreement with the claim of a number of researchers such as: Cizek, (2001) who defines cheating as any action that violates the rules for administering a test, any behaviour that gives an examinee an unfair advantage over other examinees, or any action on the part of an examinee or test administrator that decreases the accuracy of the intended inferences arising from the examinee's test score or performance. Similarly, Isangedighi concluded in Idika and Joshua (2005), that cheating is a deliberate act of commission made by a candidate alone, or jointly with others, in order to obtain a grade that is better than he is capable of obtaining. Hence, when test item compromise and practice occur, the estimates of an examinee's performance are no longer accurate; which obviously is a threat to score validity (Royal & Puffer, 2012).

The study further concluded that there does not exist a significant difference in the group performance of students on the compromised and practiced test items on the EAT. This implies that when examination items are compromised or practiced, it could give those

groups of test takers with access to such items a significant advantage which may be similar. Also, Royal and Puffer (2012) asserted that regardless of how the test is constructed, if a single item has been compromised (or practiced), it could result in some examinees receiving a score that misrepresents their actual estimates of performance which of course is a threat to the validity of the examination. Hence, the effect of a compromised or practiced item is significantly similar as they both create bias in the test score unlike when the items remain uncompromised or unpractised.

Finally, the study concluded that there is no significant difference in the scores obtained across the test item compromised, test item practiced experimental groups and control groups on the EAT after using the uncompromised, unpractised test items in the three groups used as moderating effect. This implies that when administering a test, the combination of compromised or practiced test items with uncompromised or unpractised test items will provide a moderated score that will still adequately represent the students' learning outcome. Hence, the combination will contribute positively to the validity of the score obtained from such tests when used for assessment.

Conclusion and Recommendations

The research concluded that students' scores were inflated on compromised and practiced test items which contributed to the score invalidity and that the combination of either compromised or practiced test items with uncompromised or unpractised test items will allow for moderation effect and improve score validity.

As a result of the findings, it was recommended that security of test items should be considered vital before and during test administration process and to maintain the validity of a continuous testing system, such as computerized adaptive testing (CAT), items should be monitored to ensure that the performance of test items has not gone through any significant changes during their lifetime in an item pool due to item exposure. Also, examination bodies should tighten up the security of live question papers and copyright the previously used items in case of future usage due to item rotation. Finally, the reuse of items previously used or practiced should be combined with fresh items to provide a moderated score that will be a true estimate of the test takers' knowledge.

REFERENCES

Afolabi, E. R. I. (2012). Tests and Measurement: A Tale Bearer or True Witness? *Inaugural lecture series 253.* Obafemi Awolowo University, Ile-Ife. Nigeria.

Cannell, J. J. (1988). National normed elementary achievement testing in America's public schools: How all-50 states are above the national average. *Educational Measurement: Issues and Practice*, 7 (2), 5-9.

Cizek, G. (2001). *An overview of issues concerning cheating on large-scale tests. Paper presented at the annual meeting of the National Council on Measurement in Education. Seattle.*

Cizek, G. J. (1998). *Testing in American schools: Getting the right answers*. Lawrence Erlbaum Associates, Mahwah, NJ.

Davey, T., & Parshall, C. G. (1995). *New Algorithms for selection and exposure control with computerized adaptive testing.* Paper presented at the annual meeting of the American Educational Research Association. San Francisco.

Drasgow, F., Nye, C. D., Guo, J., & Tay, L. (2009). Cheating on proctored tests: The other side of the unproctored debate. *Industrial and Organizational Psychology*, 2, 46-48.

Idika, D. (2015). Parents' concern about the use of Computer-Based Testing (CBT) for UTME in Cross River State. *Nigerian Journal of Educational Research and Evaluation.* 14 (3), 1-9.

Idika, D. O. & Joshua, M. T. (2005). Psycho-social variables as predictors of cheating tendency among students in Calabar Education Zone. *Global Journal of Educational Research*, 4 (1&2), 33-37.

Joshua, M. T. (2005). *Fundamentals of Tests and Measurement in Education*. Calabar - Nigeria: The University of Calabar Press.

Jurich, D. P., DeMars, C. E. & Goodman, J. T. (2012). **Investigating the Impact of Compromised Anchor Items on IRT Equating Under the Nonequivalent Anchor Test Design.** *Applied PsychologicalMeasurement. 36: 291-308, doi:10.1177/0146621612445575.*

Kane T. M. (2013)**.** *Validating the Interpretations and Uses of Test Scores.* National Council on Measurement in Education. doi: 10.1111/jedm.12000.

Koretz, D. (2005. *Alignment, High Stakes, and the Inflation of Test Scores*. University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Koretz, D. M., McCaffrey, D. F. & Hamilton, L. S., (2001). Toward a Framework for Validating Gains Under High-Stakes Conditions. *Centre for the Study of Evaluation: Technical Report 551*. Los Angeles, CA: University of California.

Lievens, F., & Burke, E. (2011). Dealing with threats inherent in unproctored Internet testing of cognitive ability: Results from a large-scale operational test program. *Journal of Occupational and Organizational Psychology*, 84, 817-824.

McLeod, L. D., Schnipke, D. L. (1999). *Detecting Items That Have Been Memorized in the Computerized Adaptive Testing Environment*. Paper presented at the Annual Meeting of the National Council on Measurement in Education. Montreal, Quebec, Canada.

Naglieri, J. A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A. (2004). Psychological testing on the Internet: New problems, old issues. *American Psychologist*,   59, 150–162.

Ojerinde, D. (2014). *Registrar's note. Computer- Based Test (CBT)*: Here to Stay. The Annual newsletter for candidates of the UTME, Press Pass, 21, UTME.

Ojerinde, D. (2015). Innovations in Assessment: JAMB Experience. *Nigerian Journal of Educational Research and Evaluation.* 14 (3), 1-9.

Popham, W.J. (2002). *Classroom Assessment: What teachers need to know*. (3rd ed.), Allyn & Becon Publication Company, U.S.A

Royal, K. D.&Puffer, J. C. (2012). Cheating: its implications for ABFM examinees. *American Board of Family Medicine.* 10(3) 274-275. doi: 10.1370/afm.1408.

Shogbesan,Y. O., & Faleye, B. A. (2016). Teaching to the Test: An Innovative Assessment for Learning Strategy? *International Journal of Education and Research*, 4(8), 13-22. Available online at www.ijern.com

Shogbesan, Y. O. (2017). Effect of test facets on the construct validity of economics achievement tests in Osun State secondary schools. *An Unpublished Master Thesis, Faculty of Education,* Obafemi Awolowo University, Ile-Ife, Nigeria.

Stecher, B. M. (2002). *Consequences of large-scale, high-stakes testing on school and classroom practice. Cited in Hamilton, L.S., Stecher, B. M. & Klein, S. P., (2002)*, Making Sense of Test-Based Accountability in Education. RAND Corporation.

Sympson, J. B., & Hetter, R. D. (1985). *Controlling item exposure rates in computerized adaptive testing.* In proceedings of the 27th Annual Meeting of the Military Testing Association. 937-977. San Diego, C A: Navy Personnel Research and Development Centre.

Tippins, N. T. (2009). Internet alternatives to traditional proctored testing: Where are we now? *Industrial and Organizational Psychology*, 2, 2-10.

Zara, A. & Pearson, V. (2006). *Defining Item Compromise.* Paper presented at the 2006 annual   meeting of the National Council on Measurement in Education, San Francisco.