Journal of Education and Practice www.iiste.org
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.4, No.8, 2013

Transformation Models as Predictive of Comparability Indices of Continuous Assessment in Nigeria.

Bandele, S.O[1] Ayodele, C.S.[2]*
1. Institute of Education, Faculty of Education, Ekiti State University, Ado Ekiti, Nigeria
2. Guidance & Counseling Department, Faculty of Education, Ekiti State University, Ado Ekiti, Nigeria
*E-mail: Sundayodele@yahoo.com

Abstract
This study investigated the predictive strength of Transformation models of continuous assessment scores in English Language, Mathematics, Integrated Science and Social Studies in Secondary Schools in Nigeria. The study employed survey and cross-sectional design. The sample consisted of 2,520 JSS III students selected from 36 secondary schools in 18 Local Government Areas based on multistage sampling technique. Data were collected directly from the Ministries of Education, Continuous Assessment Units with a proforma. The data collected were subjected to inferential analysis using regression. The results showed that the overall best predictor of comparability indices was True score followed by Predictive true score, T-score, derived true score while Z-score was the least. Based on the findings, it was recommended that secondary school teachers should be trained on the procedure to transform continuous assessment scores. True score should be used to standardize continuous assessment scores before finally used by examining bodies and continuous assessment units of Ministries of Education.
Keywords: comparability indices, continuous assessment, predictive true score, true score, derived true score

Introduction

The problem of non-uniformity in the quality of assessment instruments, consistency in assessment administrative procedure and procedure for scoring and grading which varies from teacher to teacher. Some schools seem to use this advantage to unduly inflate continuous assessment scores of the students to favour their schools. Beside, some uniform grades are attached to scores like A, B, C, D, E, F despite the fact that there are no uniform criteria or parameters by which such conclusions are made. Some school Registrars seem to manipulate continuous assessment scores with or without the knowledge of the subject teachers before submitting continuous assessment scores to ministries of Education to be used with JSS examination for the award of JSS certificate. These factors may constitute problem of comparability of standard of continuous assessment across secondary schools. But, it is assumed that transformed scores are useful in comparing continuous assessment scores.

Often in measurement, comparisons are made as part of judgment and decision making. For two or more sets of scores to be compared they must first be transformed into a comparable form (standard score). In his own contributions, Alonge (2004) pointed out that to facilitate meaningful analysis and interpretation, raw scores are usually transformed to other scale. Ojerinde and Falayajo (1984), Kolawole (2001) and Alonge (2004) said standard scores express individual raw score from the distribution mean in terms of the standard deviation of the distribution. In other words, it is a statement of the distance of standard deviation from the means; Z-score is one of the standard score thus defined as:

$$\text{Z-scores} = \frac{\text{raw score - mean score}}{\text{Standard deviation}}$$

The distribution of standard scores has a mean of zero and a standard deviation of one. In a normal distribution, Z-scores range approximately from -3σ to $+3\sigma$. In other words, Z-scores are almost half negative and half positive which make its interpretation very difficult for non test experts (Alonge, 2004). Anastasi (1976) stated that when Z-score is computed, negative values and decimals may occur which might tend to produce awkward numbers that are confusing and difficult to use for reporting purpose. To avoid confusion, she suggested a further, linear transformation to the scores into a more convenient form.

T-score eliminate the limitations of Z-scores, that is, the inclusion of negative values and decimals (Ananstasi, 1976, Bandele, 1999, Alonge, 2004, Gregrory, 2006 and Kolawole, 2005). Gregory (2006) argued that the mean of the transformed scores can be set at any convenient value, such as 50, 100 or 500, and the standard deviation at say 10, 15 or 100. Gregory (2006) noted that one popular kind of standardized score is the T-score and T-score scales are especially common with personality test. T-scores have a means of 50 and standard deviation of 10. The standard deviation is used as a multiplying factor and its means as addictive constant: T - Score = 10Z +50. One may then ask: why mean of 50 and standard deviation of 10? In spite of this Alonge (2004) said that T-Score has no negative values and this easy to interpret.

Afolabi (1999) argued that converting raw score to T-score is only a form of scaling and incongruity a mark from one scale to another cannot remedial incongruity or morbidity in the original score. The author added that if test scores of a class or moderation group in a subject are skewed, converting the raw scores to T-scores will still leave the distribution skewed. In the same development, Howell (2002) said linear transformation of numerical values, the data have not been changed in any way. The distribution has the same shape and observations continue to stand in the same relation to each other as they did before the transformation. In other words, it does not distort values at one part of the scale more than values at another part. It could be inferred that standard scores could put students on the same standard when scores are transformed which give room for comparison.

There is no psychological and educational measurements that are without errors. There are many reasons a student's test scores may vary. Mehrens and Lehmann (1978) affirmed that the causes of scores variability include trait instability, sampling error, administer error, scoring error and physiological variables of the students. The leniency, severity and central errors arise because rates do not use uniform standard (Verducci, 1980). Therefore, students' scores differ from each other with regard to both their true scores and their observed scores.

True score = observed score − error score

The variation in a person's scores is called error variance.

If we know nothing about an individual except his scores and group mean, and if his score as above or below the group mean, our best guess is that his obtained score contain positive or negative error of measurement respectively. The fewer and smaller the errors the more the accuracy of the measurement. Alonge (2004) said if it is possible to determine the amount of error in each examinee's score, it is possible to calculate that of standard deviation of the error score of the group. The standard deviation of the error score is called the standard error of measurement.

It can be expressed as: $S_{em} = S_x \sqrt{(1 - r_{xx})}$

Where; S_{em} is the standard error of measurement, S_x is the standard deviation of the score and r_{xx} is the reliability coefficient of the test scores.

The standard error of measurement is often used for what is called band interpretation (Mehrens & Lahmann, 1978). It is assumed that the errors are random, an individual's observed scores are normally distributed about his true scores. The standard error of measurement can also be used to define confidence interval (range) for which the student's true scores based on their obtained scores (Anastasi, 1976, Mehrens & Lehmann, 1978, & Alonge, 2004).

Whiston (2005) argued that standard error of measurement is more appropriate for interpreting individual scores because it provides the clients (students) with a possible range of scores which is more proficient method of disseminating results than reporting only one score. The author added that reporting a single score can be misleading because no instrument has perfect reliability. Gregory (2006) opined that one can never know the value of a true score with certainty but one can derive a best estimate of the true score. In the cognitive realm, a range of score is a better indicator of clients (students) capabilities (Whiston, 2005). One may ask how it would be possible to combine scores in range from difference subjects for the award of certificates or selection purpose.

The most commonly used confidence intervals are 68 percent confidence interval: Obtained score $\pm$ 1.0 $S_{em,}$ and 95 percent confidence interval: obtained score $\pm$ 2.0 S_{em} and 99 percent confidence interval: obtained score $\pm$ 2.6 S_{em}. Whiston (2005) said presenting the results as a range takes into consideration the fluctuation in depression and increase the probability that the clients (students) will find the results useful. It is very sure that if any confidence interval is used, the true score will be obtained within the range of confidence interval used. In this study therefore 95 percent confidence interval and lower limit was used for the analysis because error is regarded as mistake being made in measurement. According to Alonge (2004), error can be defined as any variable that is irrelevant to the purpose of the assessment and result in inconsistencies in measurement. The reviewed shown that error occurred as a result of the use of non-perfect instruments.

Any result from an assessment device is a combination of an individual true ability plus error. Many factors contribute to inconsistency in measurement. These include characteristics of the individual, test, or situation that have nothing to do with the attribute being measured. This factor represents the unavoidable nuisance of error

factor that contribute to inaccuracies in measurement.
However, it could be deduced from predictive true score and T-Score that.
Derived true score; D = P ± 0.1r (T-score – 50) SD
It is assumed that transformed scores are useful in comparing continuous assessment scores. The continuous assessment scores of a student tells one nothing about the student since one has no idea on how other students are scored. One does not know the total number of questions on the assessment instrument used to generate their scores nor whether some have easier questions than the others. True scores, predictive true scores, Z-scores, T-scores and Derived true scores are necessary for inter and intra individual interpretation. These transformation models will assist in a meaningful combination of continuous assessment scores in J.S.S. 3 to solve problem of comparability of standard.

Research Question:

Will transformation models predict comparability indices of continuous assessment scores for the selected school subjects among the sampled schools?

Research Hypothesis

None of the transformation models will best predict comparability indices of continuous assessment scores for selected school subjects among the sampled schools.

Methodology

This study employed survey and cross-sectional design. The population consisted of all public Junior Secondary School three (JSS3) students in South West Nigeria. These are students who were in JSS 3 in 2005/2006 and 2006/2007 sessions. The sample consisted of 2,520 Junior Secondary Schools three students' that were selected from 36 secondary schools in 18 Local Government Areas in three States based on multi-stage, stratified and simple random sampling techniques. Eight hundred and forty students were selected for the period of two years in each sampled State. Multi-stage and stratified sampling techniques were employed to select the States, Local Government Areas, Schools and Students who continuous assessment scores were used for the study. A Proforma titled "Continuous Assessment Scores Retrieval Format" was used to collect continuous assessment scores of the students selected for the study. These are continuous assessment scores in English Language, Mathematics, Integrated Science and Social Studies sent to the respective Ministries of Education in the various states for the 2005/2006 and 2006/2007 sessions. Regression analysis was used to test the hypothesis generated at 0.05 level of significance.

Results

Ho: None of the transformation models will best predict comparability indices of Continuous Assessment scores for the selected school subjects among the sampled schools.

Data were analyzed using regression analysis as presented in table.

Insert Table 1 here

The result showed that the multiple correlation coefficients for transformed English Language are 0.993, 0.993 for Mathematics, 0.991 for Integrated Science and 1.00 for Social Studies. All the coefficients were positive and high. The coefficient of multiple determination of transformed Continuous Assessment English Language was 0.985 which means that 98.5% of transformed Continuous Assessment English language could be accountable by the models from equation I, while 1.5% is due to random error. In equation II, the models could be accountable for 98.6% of transformed Continuous Assessment Mathematics while 1.4% came by chance. Also, in Integrated Science, the coefficient of multiple determinations is 0.983 i.e. 98.3% could be explained by the models in equation III while 1.7% cannot be explained. The coefficient of multiple determinations for transformed Continuous Assessment scores Social Studies was 0.999 that is 99.9% could be accountable by the models in equation IV while 0.1% cannot be explained.

The beta weight for true score was 0.902 (90.2%), predictive true score 0.144 (14.4%), T-score 0.566 (56.6%) and derived true score negative 0.738 (73.8% in opposite direction) in transformed Continuous Assessment score English language. True score was the highest contributor. Beta weight in transformed Continuous Assessment Mathematics for true score was –0.03 (3% in opposite direction), predictive true score was 0.686 (68.6%), Z-score 0.431 (43.1%) and derived true score –0.105 (10.5% in opposite direction), hence predictive true score was the highest contributor. In Integrated Science, beta weight of 0.13 (13%) was for true score, 0.186 (18.6%) for predictive true score, 0.822 (82.2%) for T-score and –0.218 (21.8% in opposite direction) for derived true score, hence T-score was the highest contributor. For social studies, beta weight for true score was 0.683 (68.3%), Z-score –0.71 (71% in opposite direction) and derived true score was 0.666 (66.6%), hence the highest contributor was true score. Garguilo (1987), Bandele (1993) and Akindehin (1997) maintain that the problem of comparability could be reduced through standardization.

The result revealed that the coefficients of multiple regressions for selected subjects (English language, Mathematics, Integrated Science and Social Studies) were positive and very high when the transformation models were used. Also, the coefficient of determination (adjusted R^2) was very high as well. That is, the models could be accountable for 98.5%, 98.6%, 98.3% and 99.9% variability in continuous assessment scores in English language, Mathematics, Integrated Science and Social Studies respectively. It indicated that the transformation models were highly related.

The findings further revealed that in Continuous Assessment scores for English language, the best predictor was True score followed by T-score next was Predictive true score and the least was derived true score which Z-score was excluded. Also, in Mathematics, Predictive true score was the best predictor followed by Z-score, next was True score and the least was Derived true score while T-score was excluded. The best predictor in Integrated Science was T-score followed by predictive true score, next was True score and the least was Derived true score while Z-score was excluded. In Social Studies, the best predictor was true score followed by derived true score and the least was Z-score, while T-score and Predictive true scores were excluded. Hence, the overall best predictor was True score followed by Predictive true score, next T-score, next Derived true score and Z-score was the least.

Based on the findings, it was recommended that subject teachers should be trained the procedure to transform Continuous assessment scores with the use of true score, Ministries of Education (continuous assessment unit) and Examination bodies like WAEC and NECO should use true score to transform continuous assessment scores sent to them.

References

Afolabi, E. R. (1999): ‘Six Honest men for continuous assessment, evaluating the equation of achievement scores in Nigeria secondary schools’. *Journal of Behavioural Research* 2 (1) 62-66.

Akindehin, F. (1997): ‘Ensuring quality of education assessment at the classroom level. Quality assurance in educational assessment’*, 18th Annual Conference proceedings of the Association for Educational Assessment in Africa (AEAA)* Abuja. 21st – 27th September.

Alonge, M. F. (2004): ‘Measurement and evaluation and psychological testing’ Ado-Ekiti, Adebayo printing (Nig) Ltd.

Anastasi, A. (1976): ‘Psychological testing’ (4th ed) New York: Macmillan Publishing C. Inc.

Ayodele, C.S. (2010): ‘Comparability of standard of continuous assessment in secondary school in Nigeria’. Unpublished *PhD Dissertation* University of Ado Ekiti, Nigeria.

Bandele, S. O. (1993): ‘Patterns of relationship between internal and external assessment of some junior secondary subjects in Nigeria’. *Journal of Educational leadership* (1).

Bandele, S. O. (1999): ‘Intermediate statistics. Theory and practice’ Ado-Ekiti, Petoa
Educational Publishers.

Garguilo, G. G. (1987): ‘The problems and practice of continuous assessment in Ondo State’*:* A seminar paper for state inspectors of education at St. Peters Unity School, Akure.

Gregory, R. J.(2006); ‘Psychological testing, history, principles and applications’. 4th ed Illinois, Wheaton College, Wheaton.

Howell, D. C.(2002); ‘Statistical methods for psychology’. (5th ed) USA. Thomson Higher education.

Kolawole, E. B. (2001): ‘Tests and measurement’ Ado-Ekiti. Yemi Prints and Publishing Services.

Kolawole, E. B. (2005): ‘Measurement and assessment in education’. Lagos Bolabay Publications.

Mehrens W. A. & Lehmann I. J. (1978): ‘Measurement and evaluation in education and psychology’. Holt, Rinehart and Winston, Inc.

Ojerinde D. & Falayajo W. (1984): ‘Continuous assessment. A new approach’. Ibadan, University Press Ltd.

Verducci, F. M. (1980): ‘Measurement concepts in physical education’. London. The C.V. Mosby Company.

Whiston, S.C. (2005): ‘Principles and applications of assessment in counseling’. (2nd ed). USA: Thomson Brooks/Cole.

Table 1: Regression Analysis on the predictive strengths of the transformation models.

	Models	Unstandardized coefficients B	Standardized coefficient β (Beta)	R	R^2	Adjusted R^2	Standard error of estimation
CA English	Constant	5.134		0.993	0.985	0.985	0.338
	True score	0.630	0.902				
	Predict true score.	0.117	0.144				
	Z – score	-	-				
	T- score	0.527	0.566				
	Derived True score	-0.325	-0.738				
CA Maths	Constant	10.465		0.993	0.986	0.986	0.314
	True score	-0.029	-0.030				
	Predict true score	0.493	0.686				
	Z – score	1.142	0.431				
	T- score	-	-				
	Derived True score	-0.068	-0.105				
CA Int. Sc.	Constant	2.442		0.991	0.983	0.983	0.374
	True score	0.113	0.130				
	Predict true score.	0.157	0.186				
	T- score	0.778	0.822				
	Derived True score	-0.066	-0.218				
CA Social Studies	Constant	-7.511		1.000	0.999	0.999	0.065
	True score	0.653	0.683				
	Predict true score	-	-				
	T-score	-	-				
	Z – score	-12.544	-0.710				
	Derived True score	0.606	0.666				

CA English = 5.134 + 0.63t + 0.117p + 0.527T – 0.325D

CA Math = 10.485 – 0.029t + 0.493p + 1.142z – 0.068D

CA Int. Science = 2.442 + 0.113t + 0.157p + 0.778T – 0.066D

CA SOS = -7.511 + 0.654t – 12.544z + 0.606D

Where t stands for true score, p is predictive true score, T is T-score, Z is Zed score while D is derived true score.