



ASSEREN

JOURNAL OF EDUCATIONAL RESEARCH AND DEVELOPMENT (AJERD)

VOL. 3, JULY 2016

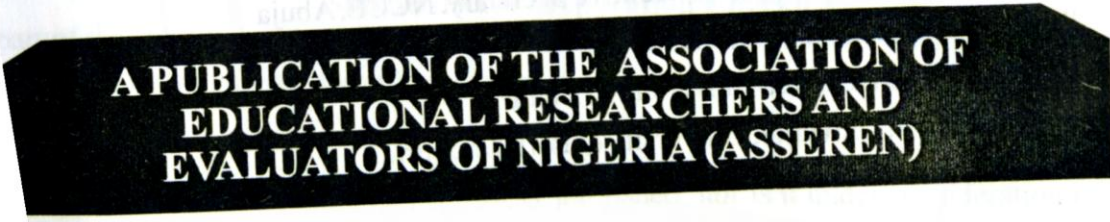
ISSN 2536-6899



ASSEREN

Journal of Educational Research and Development

VOLUME 3, JULY, 2016



**A PUBLICATION OF THE ASSOCIATION OF
EDUCATIONAL RESEARCHERS AND
EVALUATORS OF NIGERIA (ASSEREN)**

EDITORIAL BOARD

Editor – in – Chief

Professor A.O.U. Onuka
Institute of Education, University of Ibadan, Nigeria.

Deputy Editor –in – Chief

Monica N. Odinko, Ph. D
Institute of Education, University of Ibadan.

Associate Editor

E.O. Babatunde, Ph. D
Institute of Education, University of Ibadan.

Members

D.O.K. Omole, Ph.D, NERDC, Abuja
B.A. Falaye, Ph.D. OAU, Ile-Ife
Mustapha Yusuf, Ph.D., University of Jos
J.N. Odili, Ph.D. Delsu, Abraka.

Editorial Assistants:

Esther O. Durowaju, Ph.D
S.O. Alieme, Ph.D
T.G. Atsua

Editorial Advisers

1. Emeritus Prof. PAI Obanya, University of Ibadan.
2. Professor Adenike Emeke, University of Ibadan.
3. Professor Monday T. Joshua, Executive Secretary, NCCE, Abuja
4. Professor E.R.I. Afolabi, OAU, Ile-Ife
5. Professor O.A. Adenike University of Benin
6. Professor R.P.I. Ukwuijie, University of Port Harcourt.

TABLE OF CONTENTS

VOLUME 3

1.	The Nexus Between Purpose of the Study and Sampling Technique in Educational Research Prof. (Mrs) Alice E. Asim, Dr Eni I. Eni & Dada, Oluseyi A.	1
2.	Types and Procedures of Assessment Instruments Used in the Ramat Polytechnic Maiduguri, Borno State, Nigeria Terhemba Godwin Atsua & Umaur Abdullahi	14
3.	Action Research in Chemistry Classroom: Suggested areas Chemistry Teacher should be familiar with - Lawrence Achimugu	23
4.	Development, Validation and Use of Criterion Referenced Achievement Test in Biology for Senior Secondary Schools in Abia State. Collens I. Odo, Charles Ndukwu Agbaegbu	33
5.	Assessment of National Youth Service Corps Members' Influence on Junior Secondary School Students' in Bauchi State Aminu Yusuf & Ishaq Abdulsalam	42
6.	Development of Economics Mock Examination for Senior Secondary School Students in Imo State Ezechukwu, Ifeoma R. & Amaechi, Chidera Emmanuel	50
7.	Construction and Validation of Teacher-Pupil Interaction Instrument for Primary Schools in Nigeria Etobro, A. Benjamin & Adefowora, T. Adesola	59
8.	Assessment of Lecturers' Knowledge of observation as A Tool for Data Collection of Non-cognitive Traits in Research Amaechi C. Ifeakor & Mary O. Akujieze	70
9.	Effect of Convergent and Divergent Methods of Teaching on Students' Mathematics Achievement in Senior Secondary School in Umuahia Education Zone Ifeanacho A. O. & Osuagwu, C. T.	79

10.	Development and Validation of An Achievement Test in Agricultural Science for Senior Secondary Schools in Imo State Adolphus Onuabuchi Madu, Kelechi Joshua Princewill Ihechu & Akajiuba Flora Ugochi	88
11.	Portfolio As Assessment Tool for Mathematics Teaching Practice Dr Ihekweba Chinyere Ngozi, Unamba Eugene Chukwuemaka & Ikeanumba Michael.O.	98
12.	Evaluating the Achievement Level of the Basic Education needs of selected Primary and Junior Secondary Schools in Imo State From 2013-2015 Ihekwoaba Celestine K.C., Ekeada Chidi & Akpaka Blessing	108
13.	Assessment of Teachers' Understanding of Criterion Related Validity in Secondary Schools in Owerri Jiwueze, Celestine Chibuike	118
14.	Teacher Perception and Pupils Attitude to use of ICT Resources In Ibadan North Lga Of Oyo State, Nigeria. Ofem Ubi Arikpo & Monica N. Odinko (Ph.D)	123
15.	Development of Mathematics Achievement Tests for Senior Secondary Schools using The Rasch Model P. M. C. Ogomaka, Monica C. Nosike & Thecla A. Nwanebo	135
16.	Computer Literacy and Computer Anxiety as Predictors of Candidates' Performance in Unified Tertiary Matriculation Examination (UTME) Computer Based Test Opara, Ijeoma Margaret & Egbejule Ebizimor	144
17.	Counselling Undergraduate Students towards carrying out High Quality Research Work Orji Tobias Ihiegbulem & Mrs. Victoria Nkechi Ihiegbulem	152
18.	Survey of Secondary School Teachers' Knowledge of Soft Skills for Innovative Examination in Nigeria Omole, D.O.K. & Takema Susan Rumun	158
19.	Equating of the Senior Secondary School WASSCE Economics Multiple-choice Papers (2011 and 2015) A. R. Tomilola; H.O. Owolabi & D. S. Daramola	167

EQUATING OF THE SENIOR SECONDARY SCHOOL WASSCE ECONOMICS MULTIPLE-CHOICE PAPERS (2011 AND 2015)

A. R. Tomilola;

H.O. Owolabi

&

D. S. Daramola

Abstract

The non-stability of the percentages of failure and passes each year has necessitated the comparability of test scores from different test forms. Since the test scores serve as the major yardstick through which students' achievement is determined, it is important that test scores be comparable from one test form and year to another. This study equated the Senior Secondary School WASSCE Economics Multiple-choice papers (2011 and 2015) as well as investigated the (i) ability levels; (ii) difficulty levels of 2011 and 2015 WASSCE multiple-choice Economics items, (iii) result of the linking items and (iv) horizontal equating of 2011 and 2015 WASSCE Economics multiple-choice test items. The Non-Equivalent Anchor Test Design of test score equating was used and the sample consisted of 650 senior secondary three students selected from 14 secondary schools in Kwara Central Senatorial District. The 2011 and 2015 multiple-choice Economics papers were adopted and used as instruments in this study. The data were analysed using WINGEN and IRTEQ IRT software packages. The examinees who took the 2011 and 2015 WASSCE Economics test items with mean ability of 2.16 and 2.93 respectively had the same ability; (ii) 2011 and 2015 WASSCE items were found to be of moderate difficulty at 0.083 and 0.098 respectively; (iii) the mean difficulty of linking items was moderate at 0.41% difficulty coefficient and (iv) it was revealed that a score of 4 in 2011 was found to be equivalent to a score of 2.33 for the MS in 2015, a score of 10 in 2011 is equivalent to a score of 7.54 for the MS in 2015, also, a score of 23 in 2011 is equivalent to a score of 22.63 for the MS in 2015. Thus, IRT equating method should be

employed in equating test scores since it is an improvement of the Classical Test Theory and also developed to overcome the limitations of classical test theory (CTT)/traditional method of equating.

Introduction

Assessment as a crucial part of educational system is used among others for selecting candidates for a specific educational level or programme, as well as for the certification of successful participants on the completion of a programme or level of education. Assessment is a very useful tool for monitoring and evaluating educational system and it serves as great motivation for teaching and learning process. Test as one of the assessment techniques is a means of eliciting behaviour for observation with a view to assessing or making some inference about a specific attribute of the respondent (Abiri, 2007).

Test as one of the assessment techniques which is used for obtaining information from students is a regular procedure to sample behaviour as well as to describe the behaviour in terms of scores or categories (Adeleke 2013). Owolabi (2004) defined a test as a sample of behaviour drawn to ascertain the presence or absence of certain traits, characteristics or skills and the extent to which these traits, characteristics or skills are present or absent in an individual or a group of persons. Test is

The West African Examinations Council (WAEC), National Examinations Council (NECO) and National Business and Technical Examination Board (NABTEB) are examination bodies that conduct standardized tests for Nigerian secondary school students and these tests according to Abiri (2007) are known to possess psychometric properties before it could be judged worthwhile because they have been carefully prepared and standardized for use using large samples which helps to ascertain its validity and reliability. Psychometrics is primarily concerned with the construction and validation of measuring instrument such as test, questionnaires, personality inventories etc. and the usefulness of tests depend on its quality. The psychometric properties of multiple-choice items include distractor effectiveness, difficulty and discrimination indices. According to Schumacher (2005), classical test theory (CTT) utilizes traditional item difficulty, item discrimination estimates, distractor analysis, test item inter correlations and a variety of related statistics. The two statistics that form the corner stone of classical test theory are item difficulty and item discrimination.

Olatunji (2007) described difficulty index of an item as the extent to which an item has been answered correctly by the testees. That is, the percentage of the testee that chooses the right option. The closer to one (1) the value of the difficulty index the easier the item and the closer the value to zero (0), the more difficult the item. Discrimination power of multiple-choice items, on the other hand, is the ability of the test items to discriminate between the brilliant students and the poor students (Olatunji, 2007). Discrimination power of a test item ranges from zero to one (0-1). The closer the value is to one (1), the better the item. The effectiveness of a Multiple-choice test is judged by how plausible the distractor is if five percent of the testees selected it.

Since the test scores serve as the major yardstick through which student achievement is determined, it is critical that test scores be comparable from test to test and year to year. Scores on tests often are used as one piece of information in making important decisions. Regardless of the type of decision that is to be made, it should be based on the most accurate information as possible: *All other things being equal, the more accurate the information, the better the decision.* Making decisions in many of these contexts requires that tests be administered more than once. Thus, in order to address the issue of test security, administering different forms of a test is necessary. A test form is a set of test questions that is built according to content and statistical *test specifications* (Millman and Greene, 1989). The use of different test forms on different periods which tests are taken leads to the concern that the forms might differ somewhat in terms of difficulty level. This challenge is usually addressed through test equating.

Equating is a statistical process that is applied to adjust scores on test forms so that scores on the forms can be used interchangeably. Equating adjusts for differences in difficulty among forms that are built to be similar in difficulty and content. *Equating* is a process of making test scores across **different forms of the same test** interchangeable. When test forms are created to be similar in content and difficulty, *equating* adjusts for differences in difficulty. Test forms are considered to be essentially the same, so scores on the two forms can be used interchangeably after equating has adjusted for differences in difficulty.

The need for *test equating* arises when there are two or more tests of the same construct or subject that can yield different scores for the same examinee. The goal of test equating is to allow the scores on different forms of the same tests to be used and interpreted interchangeably.

Test equating, as it is currently practiced, requires some type of control for differential examinee ability, or proficiency, in the assessment of, and adjustment for, differential test difficulty. The term equating refers to the procedure for establishing comparable scores from different tests or test forms. It is used to provide solutions to the challenge of maintaining comparability of scores from different forms of a test and/or from different test administrations (Doran, Moses and Eignor, 2010).

The process of equating is used in situations where such *alternate forms* of a test exist and scores earned on different forms are compared to each other. Even though test developers attempt to construct test forms that areas similar as possible to one another in content and statistical specifications, the forms typically differ somewhat in difficulty. Equating is intended to adjust for these difficulty differences, allowing the forms to be used interchangeably. *Equating adjusts for differences in difficulty and not for differences in content.* After successful equating, for example, two examinees who earn a score of 26 on the different forms of a test respectively could be considered to have the same achievement level (Kolen and Brennan, 1995).

Test equating is a statistical process providing the inter conversion of the scores obtained from different test forms measuring the same content. In the event that the students are simultaneously given different forms of a test and the difficulty level between the forms are not equal, it is possible that the individuals given the difficult test get lower scores than the individuals given the easy test. This situation makes it difficult to compare the points obtained from different test forms. Test equating prevents the possible unfairness against individuals given the difficult test and eliminates the bias problems caused by the difference in the difficulty levels of the test forms. Equating is only necessary because a

standardized educational assessment uses numerous test forms that tend to differ in difficulty although they are built to the same specifications.

Peterson, Cook and Stocking (1989) defined equating as the process of ensuring that scores resulting from the administration of the multiple forms can be used interchangeably. Angoff (1971) has defined the equating of tests as a process of converting the system of units of one form to the system of units of the other so that the scores obtained from one form could be compared directly with the scores obtained from the other form. In many situations in test administration, more than one form of a test is used for security reasons (Jaeger, 1980; Jaeger, 1981; Kolen and Whitney, 1982; Cook and Eignor, 1991; Kolen, 1994).

When reporting evidence of change in performance between two forms of a test, it is important that a distinction be made between differences in difficulty of the test forms used to assess the students and real gains or losses in achievement between the two years. The purpose of equating is to adjust for test difficulty differences so that only real differences in performance are reported (Xuan and Rochellen 2011).

Psychometric approaches to equating test scores are based on the traditional approach known as Classical Test Theory (CTT) and Item Response Theory (IRT).

There are two ways by which equating could be done which are the Vertical (also referred to as Across-grade) and Horizontal (also referred to as Within-grade) equating.

Statement of the Problem

The use of different tests aimed at measuring the same constructs from year to year, by different examination bodies in Nigeria (such as the Joint Admissions and Matriculations Board JAMB, West African

Senior School Certificate examination WAEC, National Examinations Council NECO and National Business and Technical Examination Board NABTEB) which conducts final examination for students at the end of their third year of Senior Secondary School education raises the issue of comparing scores from these different examinations.

A common feature of the results being released by these public examining bodies is non-stability of the percentages of failure and passes each year. Reports have shown that there has been a continuous decline in the performance of candidates sitting for SSCE examinations. Charles Eguridu (2015) noted that in 2012, 2013 and 2014, 38.31%, 36.57% and 31.28% of the respondents had credits in five subjects respectively. Adeyemi (1990) gave several reasons which can be attributed to the cause of the decline in students' performance one of which is the difficulty level of the test item which is one of the motivating factors for carrying out this present study. Also, series of studies have been carried out on comparability of test scores from different test forms. For example, Adokoniyi (2014) carried out a study on equating of Kwara State Joint Mock Economics Multiple-choice Papers. Also Olatunji, (2015) also carried out a study on analysis of linear and equipercentile equating of senior school certificate Economics Multiple-choice Papers and findings from the study showed that WASSCE Economics multiple choice items are standard and therefore should maintain standards. These previous researches have only worked on equating scores from different examination bodies as well as Mock examination but the gap left out is that none has compared WASSCE over years and also with the use of the Item Response Theory (IRT) approach which necessitated this study.

Research Questions

1. What are the ability levels of students in WAEC 2011 and WAEC 2015 multiple choice Economics items?
2. What are the difficulty levels of WAEC 2011 and WAEC 2015 multiple-choice Economics items?
3. What is the result of the linking items of 2011 and 2015 WAEC Economics multiple-choice Examinations?
4. What is the result of horizontal equating of 2011 and 2015 WAEC Economics multiple-choice Examination?

Methodology

The Non-Equivalent Anchor Test design was used for this study. The population for this study consisted all Secondary School Students in Kwara Central Senatorial District, the target population consisted all SS3 Students in Kwara Central Senatorial District and the sample for this study consisted 650 SS3 students which were sampled through the use of multistage sampling technique.

At the first stage, proportionate sampling technique was used to select Schools from the four local governments in Kwara Central Senatorial District. Ten percent of the schools in each Local Government were used because the number of schools in each of the local governments is unequal. As at the period of this study, Kwara Central Senatorial District had 95 public Senior Secondary Schools. Asa Local Government Area had 21 public senior secondary Schools, Ilorin East Local Government Area had 26 public secondary schools, Ilorin South Local Government had 22 public senior secondary schools and Ilorin West Local Government had 26 public senior secondary schools (Kwara State Ministry of Education, 2015).

At the second stage, purposive sampling technique was used to select SS3 students from

the chosen schools. Senior Secondary three students were chosen because of their level of preparedness toward their examination.

At the third stage, proportionate sampling technique was used to select students who offer Economics from the three levels (Art, Commercial and Science class) in each school.

Instrumentation

The 2011 and 2015 May/June WASSCE Multiple-choice Economics items were adopted in its original form for this study.

Data Analysis Techniques

The data collected from this study was scored dichotomously and analysed with respect to the research questions raised in the course of this study. The ability index of each

student who responded to the test items and the item parameter (item difficulty) were generated with the use of Wingen IRT software package and analysed to provide answers to research questions one and two. Also, IRTEQ IRT software package was used to carry out horizontal equating of the two tests which was used to answer research question three and four.

Findings and Discussion

Research Question 1: What are the ability levels of students in 2011 and 2015 WASSCE Economics multiple-choice items?

The ability index (theta) of each examinee who responded to the tests items was generated and presented in Appendix IV. Under Item Response Theory (IRT), ability levels range from -3 to +3. The estimated examinees' abilities are summarised in table 1.

Table 1: Examinees' Ability (Theta) in 2011 and 2015 Economics Multiple -Choice Items

	WASSCE	N	Mean	S.D.	Minimum	Maximum
2011	325	0.121	1.083	-3.04	2.86	
2015	325	0.128	0.970	-2.89	2.93	

Table 1 shows the summary of estimated abilities of the students who responded to the test items. The mean ability (theta) of the examinees in WASSCE 2011 is 0.121 with a range of -3.04 to 2.86 and standard deviation (SD) of 1.083 while the mean ability (theta) of the examinees in WASSCE 2015 is 0.128 with a range of -2.89 to 2.93 and SD of

0.970. This shows that there is no difference in the ability of the SS3 examinees who responded to the test items of 2011 and 2015 WASSCE and that they are also of medium ability.

Research Question 2: What are the difficulty levels of 2011 and 2015 WASSCE Economics multiple-choice items?

Table 2: Mean Difficulty Index of 2011 and 2015 Economics Multiple-Choice Items

	WASSCE	No. of Items	Mean
2011	42	0.083	
2015	420	0.098	

Table 2 presents the mean difficulty index of 2011 and 2015 WASSCE multiple-choice Economics items. The table reveals that WASSCE 2011 Economics items had mean difficulty of 0.083 while WASSCE 2015 Economics items had mean difficulty of 0.083. This shows that the 2011 and 2015 WASSCE Economics items are moderately difficult.

Research Question 3: What is the result of the linking items of 2011 and 2015 WASSCE Economics multiple-choice Examination?

The result of linking items of 2011 and 2015 WASSCE Economics multiple-choice Examinations were generated through IRTEQ Equating Software and presented in Table 3.

Table 3: Mean Difficulty of the Linking Items of 2011 and 2015 Economics Items

WASSCE	No. of Item	Mean	SD	Min	Max	Diff Co
2011	8	0.174	1.057	0	7	0.41
2015	8	0.181	1.062	0	7	

With mean difficulty of 0.174 for the linking items of WASSCE 2011 and 0.181 for the linking items of WASSCE 2015 and 0.41 difficulty coefficient of the linking items, it can be said that the 2011 and 2015 WASSCE Economics items were moderately difficult.

WAEC Economics multiple-choice Examination?

The result of horizontal equating of 2011 and 2015 WAEC Economics multiple-choice Examinations were generated through IRTEQ Equating Software and presented in Table 4. The Mean-sigma equating parameter was generated for each corresponding score.

Research Question 4: What is the result of horizontal equating of 2011 and 2015

Table 4: Summary of Horizontal Equating of 2011 and 2015 Economics Examination

Scores	Mean Sigma
Test 1 (2011)	(Test 2-2015 rescaled)
1	0.48
4	2.33
10	7.54
23	22.63
32	33.59
49	49.36
50	49.99

From Table 4, it is shown that a score of 4 in 2011 is equivalent to a score of 2.33 for the MS in 2015, (this also means that an examinee who earns a score of 4 in 2011 is considered to be at the same achievement level as an examinee who earns a score of 2.33 (MS) in 2015), a score of 10 in 2011 is equivalent to a score of 7.54 for the MS in 2015, (an examinee who earns a score of 10 in 2011 is considered to be at the same achievement level as an examinee who earns a score of 7.54 (MS) in 2015) also, a score of 23 in 2011 is equivalent to a score of 22.63 for the MS in 2015 (also means an examinee who earns a score of 23 in 2011 is considered to be at the same achievement level as an examinee who earns a score of 22.63 (MS) in 2015).

Conclusion

Based on the findings from this study, it can be concluded that the examinees who took the 2011 and 2015 WASSCE Economics test items were of the same ability, specifically, they were of medium ability.

WASSCE 2011 and WASSCE 2015 items were moderately difficult. The mean difficulty of linking items in test 1 (WASSCE 2011) and test 2 (WASSCE 2015) also revealed that the two items were moderately difficult at 0.41 difficulty coefficient.

The result of horizontal equating of 2011 and 2015 WASSCE Economics multiple-choice Examinations generated through IRTEQ Equating Software revealed that the scores on the two test forms can be used interchangeably given the Mean-Sigma equating procedure.

Recommendations

Based on the findings and conclusions drawn in this study, the following recommendations are made:

- i. Giving that the SS3 students are of medium ability levels, items being constructed and administered to students by WAEC should have moderate difficulty.
- ii. WASSCE Economics items are moderately difficult and therefore WAEC should maintain its standards over years by consistently ensuring that the difficulty levels of the examination items are moderate so as to give room for the students to pass the examination.
- iii. IRT equating method should be employed by prospective researchers in equating test scores since it is an improvement of the Classical Test Theory and also developed to overcome the limitations of classical test theory (CTT)/traditional method of equating.
- iv. Students should be educated on how scores can be used interchangeably and so they should be more serious and determined with more seriousness to do well.

References

- Abiri, J. O. O. (2007). *Elements of evaluation measurement and statistic techniques in education*. Unilorin Press: University of Ilorin, Nigeria.
- Adeleke (2013). Development and Validation of Scientific Literacy Achievement Test
- Angoff, W. H. (1971). *Scales, norms and equivalent scores*. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education. Belmont, CA: Wadsworth Group, 508-600.
- Cook, L. L. & Eignor, D. R. (1991). An NCME instructional module on IRT equating methods. *Educational Measurement: Issues and Practice*, 10, 37-45.

- Doran, N. J., Moses, T. P. & Eignor, D. R. (2010). *Principles and practices of test score equating*. New Jersey: ETC Princeton.
- Grolund, N. E. (1998). *Assessment of student achievement* (7th ed.). Boston: Allyn and Bacon.
- Jaegar, R. M. (1981). *Some exploratory indices for selection of a test equating method*. *Journal of Educational Measurement*, 18, 23-38.
- Jaeger, R. M. (1980). *Some exploratory indices for selection of a test equating method*. *Journal of Educational Measurement*, 18, 23-38.
- Kolen, M. J. & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York: Springer Verlag.
- Kolen, M. J. & Whitney, D. R. (1982). *Comparison of four procedures for equating the tests general educational development*. *Journal of Educational Measurement*, 19(4), 279-293.
- Kolen, U. (1994). *Conditional standard errors of measurement for scale scores using IRT*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans.
- Millman, J. & Greene, J. (1989). The specification and development of tests of achievement and ability. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.), pp.335-366). New York: American Council on Education & Macmillan.
- Olatunji, D. (2015). *Analysis of Linear and Equipercentile Equating of Nigerian Senior School Certificate Examination Economics Multiple Choice Papers in Kwara State, Nigeria*. Unpublished Ph.D. Thesis, University of Ilorin, Ilorin.
- Olatunji, D. S. (2007). *Effects of Number of Options on Psychometric Properties of Multiple Choice Tests in Economics*. Unpublished M.Ed. Thesis, University of Ilorin, Ilorin.
- Owolabi, H. O. (2004). *Assessment in the classroom*. In E.O. Ogunsakin (Ed). *Teaching in Tertiary Institutions*. (pp.93-99). Faculty of Education, University of Ilorin, Ilorin.
- Petersen, N.S. Cook, L.L. & Stocking, M.L. (1989). *IRT versus Conventional Equating Methods: A Comparative Study of Scale Stability*. *Journal of Educational Statistics*, 8, 137-156.
- Schumacker, R. E. (2005). *Test equating*. Retrieved March 21st 2009 from <http://www.appliedmeasurementassociation.com/white%20papers/TEST%EQUATING.Pdf>.
- Xuan, T. & Rochellen, M. (2011). *Why do standard testing programme report scaled score*: Retrieved from <http://www.ets.org/understandingtesting/glossary/>.