IFE JOURNAL OF THEORY AND RESEARCH IN EDUCATION

ISSN: 0794-6754

IJOTRE

JOURNAL OF THE INSTITUTE OF EDUCATION

Obafemi Awolowo University, Ile-Ife.

Bi-Annual

VOL 13 Nos. 1 & 2, 2011

CALL FOR PAPERS
IFE JOURNAL OF THEORY AND RESEARCH IN EDUCATION
(IJOTRE)
NOTES TO CONTRIBUTORS

The Institute of Education Journal of Theory and Research in Education (IJOTRE) disseminates information derived from research findings and theoretical topics in the areas of Nursery, Primary, Secondary and Higher Education for practitioners, educators, educationists, academia, researchers, curriculum planners and policy- makers with the main goal of improving educational status. IJOTRE is a bi-annual and peer-reviewed journal.

Articles are expected to focus on any of the following topics:

1. Contributions on theoretical, methodological and practical teaching aspects of education.
2. Research notes and projects reports.
3. Articles representing scholarly opinions on contemporary issues and trends stemming from any aspects of education.
4. Book reviews that are significant in the field of education.

Guidelines for Papers Submission

- Articles should not be longer than 15 A4-sized pages using Time New Roman, font size of 12. Longer articles will attract additional publication fee.
- Reference style should conform to the America Psychological Association format (6th Edition). This should be arranged in alphabetical order according to the surname of the authors.
- Footnotes are not allowed
- Manuscripts' cover should include the title of the paper, author(s)' name(s), institution affiliation and E-mail address.
- Abstract should not be more than 250 words
- Two hard copies of the manuscripts should be submitted for review. Articles can also be submitted by post to The Managing Editor, IJOTRE, Institute of Education, Obafemi Awolowo University, Ile - Ife. Nigeria or electronically via e-mail to insteduc@oauife.edu.ng or holubunmi@yahoo.com
- Assessment fee of NGN3,000.00 ($30).

It is a condition for publication that a manuscript submitted to Ife Journal of Theory and Research in Education (IJOTRE) has not been published and will not be simultaneously submitted or published elsewhere. Submissions are published at the editor's exclusive discretion. Submissions that do not conform to these guidelines may not be considered for publication

Bi - Annual
IJOTRE, Vol. 13, Nos. 1 & 2, 2011.

Item Local Independence in WAEC (SSCE) Economics in Ajeromi-Ifelodun Local Government Area, Lagos State.

Adams O.U Onuka, Ph.D
&
Nicholas W. Oke
Institute of Education,
University of Ibadan,
Nigeria
adamonuka@yahoo.com, ada.otuoze@gmail. or tomeoke@yahoo.com

Abstract

The differential scholastic achievement of students' assessment by examining bodies seems to be the "Holy Grail" in the education reform. If Nigeria would be among the best 20 economies, comparative standards must be maintained in our education sector before actualizing this dream. With an eye towards research and practice, this study reviewed and evaluated main trends that have contributed to the increasing use of assessment mechanisms to assess and evaluate students'. The goal of this paper was to highlight the major issues and challenges of massive failures recorded by examining bodies using of Item Response Theory (IRT) assumptions, which is Local Independence. The study adopted a survey research design and simple random sampling technique was used to select 750 students that were preparing for the 2012 WAEC (SSCE) Economics Examination in Lagos state. Two research questions were addressed using descriptive statistics, TSP, TID and Tetrachoric correlation to locate the extent at which WAEC items are Locally Independent. Findings from this study revealed that a larger percentage of the students' proficiency level on the test-item cut across different ability group, while the Tetrachoric correlation revealed that WAEC (SSCE) Economics objective items for 2010 met the assumptions of IRT on Local Independence. Thus, it is recommended that examining bodies, evaluators, assessors, tertiary institutions, parents and government should welcome the use of IRT in our educational system for quality assessment procedure, while measurement and evaluation courses need to be reviewed to inculcate more practical work than theory- based.

Introduction

In determining the quality of education in 1990s and 2000s, both national and international assessments have become extremely popular tools for assessing students' performance. This increase in popularity as posits by Greaney and Kellaghan (2008) reflects two important developments. First, it reflects increasing globalization and interest in global mandates, including Education for All (UNESCO, 2000). Secondly it represents an overall shift in emphasis in assessing the quality of education from a concern with inputs (such as students' participation rates, physical facilities, curriculum materials and teacher training/qualification) to a concern with outcome (such as knowledge and skills that students have acquired as a result of their exposure to schooling). Some academic economist and educationist according to Todaro and Smith

(2006) teach and research totally irrelevant sophisticated mathematical models of non-existent competitive economies, while problems of poverty, rural development, unemployment and education are considered less intellectually interesting in all these diverse professional activities. Performance criteria are often based not on contributions to national development but rather on praise from international community (professional mentors in the developed nations). In all, the educational system in Nigeria and of most developing countries is in need of great reform. This challenge is greatest in education, where the large body of empirical evidence linking education to economic growth indicates that improved enrollment and completion rates are necessary, but not sufficient conditions for poverty education. Instead, enhanced learning outcomes in the form of increased students' knowledge and cognitive skills are keys to alleviating poverty and improving economic competitiveness (World Bank, 1996). Despite growth in national and international assessment activity, a lack of appreciation still exists in many quarters about the worth and potential value of the data that assessment can provide, as well as a deficit in the skills required to carry out a technically sound assessment.

Greaney and Kellaghan (2008) were able to give a number of reasons that account for assessment not fully exploited;

1. The policy makers may have only been peripherally involved in the assessment and may not have been fully committed to it.
2. The results of analysis may not have been communicated in a form that was intelligible to policy makers.
3. Policy makers may not have fully appreciated the implications of findings for social policy in general or for educational policy in particular relating to curricular provision, the allocation of resources, and the practice of teaching and teachers professional development.

With so much emphasis today on high-stakes testing for promotion, graduation, teacher and administration accountability, and school certification/accreditation, it is critical that all educators understand concepts like standard error of measurement, reliability coefficient, confidence intervals and standard setting. It therefore follows that when assessment is integrated with instruction, it informs teachers about what activities and assignment will be most useful, what level of teaching is most appropriate and how summative assessment provides diagnostic information.

Cohen and Swerdlik (1999) noted that the development of a new test may be in response to a need to assess mastery in an emerging occupation or profession. For example, new test may be developed to assess mastery in fields such as environmental engineering, wireless communication, information technology and computer networking. As technology advances and teachers become more proficient in the use of technology, there will be increased opportunity for teachers and administrators to use computer-based techniques (item bank, electronic grading, computer-adapted testing, and computer-based simulations), internet resources and more complex detailed ways of reporting results. There is however, a danger that technology will contribute to the mindless use of new resources, such as using item-on-line developed by some companies without adequate training, exposure and evidence of reliability, validity and fairness and crunching numbers with the software without sufficient thought about weighting,

error and averaging.

Before now, psycho-metricians in the 1960s and 1970s were conversant with the use of Classical Test Theory (CTT) as a measurement framework in assessing students' performance level in a given test item. Although CTT has served the measurement community for most of this century, Item Response Theory (IRT) has witnessed an exponential growth in recent decades. The major advantages of CTT are its relatively weak theoretical assumptions which make CTT easy to apply in many testing situations and also its extension principle (generalizability theory) (Hambleton and Jones 1993). Despite the theoretical weakness of CTT in terms of its circular dependency of item and person statistics, measurement experts have worked out practical solutions within the framework of CTT for some otherwise difficult measurement problems. For example, test equating can be accomplished empirically with the CTT framework (e.g. equi-percentile equating). It is fair to say that to a great extent, although there are some issues that may not have been addressed theoretically with the CTT framework; many have been addressed through ad hoc empirical procedures. The major criticism for CTT is its inability to produce item/person statistics that would be invariant across examinee/item samples. This criticism according Xitao (1998) has been the major impetus for the development of IRT models and for the exponential growth of IRT research and applications in the recent decades.

Thorndike (1997) posits that Item Response Theory (IRT) or latent trait theory depends on the availability of computers and has in turn shaped the way in which computers are used in testing. Latent trait theory assumes the existence of a relatively unified underlying trait or characteristic that determines an individual's ability to succeed with some particular type of cognitive task- possible attribute might be knowledge of word meanings, arithmetical reasoning, or spacial visualizing.

Zenisky, Hambleton and Sireci (2003) also acknowledged that the probability that a test taker will provide a specific response to an item is a function of the test takers location on theta (è), and one or more parameters (depending on the IRT model chosen) describing the relationship of the item to è. Theta (è) distinguishes items with respect to difficulty and test takers with respect to proficiency. The relationship between ability level and passing an item of a given difficulty is not an-all-or-none matter, but, instead, it is a question of probability. Since IRT models are probabilistic, independence must be assumed conditional on è between responses to any pair of items. This conditional independence is called item local independence.

Ubi, Joshua and Umoinyang (2011) posits that, local independence of items conceptualizes that the probability of an examinee getting an examination item correct must not be dependent on the answers given to other items in the examination. This is because the ability which influences responses to any two items in a test is constant; thus, the relationship between the two items should not differ from zero (0). If it does, then responses to the item are influenced by factor(s) other than what the test instrument was designed to measure.

The basic principle involved in producing local item independence is that once the performance of examinees on some test items has been determined, then, there will be no additional factor(s) that can consistently affect such performance. In a factor analysis, such factors may not be desirable as they would not likely contribute to

load on a second factor. Such factors may be considered undesirable and without important dimensions of behaviour being measured. Yung-chen (2007) states that once a particular performance score is known, nothing changes it anymore. Rather, it is those factors that caused the performance that remain as the determinant factors. It is important to guard against factors such as; external interference, fatigue, exposure to the questions, same response format, item chaining, speediness of the test, leniency or victimization on the part of the examiner and uniqueness of item contents while preparing for a test and during the test administration to ensure local independence of items.

The IRT framework encompasses a group of models, and the applicability of each model in a particular situation depends on the nature of the test items and the viability of different theoretical assumptions about the test items. Since assessment assesses what the assessment procedure intends to assess, it is therefore the purpose of the assessment process to develop a tool or measurement device which when applied, evaluates what we intend to assess.

According to Kpolovie, Ololube and Ekwebelem (2011) evaluation agencies were set up to promote education, to co-ordinate educational programmes, and to control and monitor the quality of education in educational institutions. The essence of which is the organization of public examinations so as to provide uniform standards to all test takers, irrespective of the type or method of instruction they have received.

The implication and use of IRT by examining bodies should be taken into consideration in the following areas:

1. Test scoring and prediction: Testers can use IRT in test scoring to increase accuracy by taking into account the statistical characteristics of the particular items that the students answered correctly. Such scoring methods can be helpful in increasing score accuracy for low-scoring students who have taken multiple-choice test. The user of any test score should know the amount of measurement error it is likely to contain.
2. Item banking: Item banking is the collection of test items, "stored" with known item characteristics. Depending on the intended purpose of the test, items with desired characteristics can be drawn from the bank and used to construct a test with known properties.
3. Item development and construction: Item model is presently being used by a number of organizations in test development/construction. An item model can be used to create a pool of items that have known statistical characteristics including descriptions of how well each item is measuring students at each ability level.
4. Item selection: Item selection model is useful to the problem of item selection because they lead to item statistics, which are referenced to the same scale on which examinee abilities are defined. It should be noted that IRT provides a procedure for placing a cut-off score, which is normally set on a proportion-correct scale defined over a domain of items on the same scale as the test items and the examinees.
5. Equating: Emeke and Edward (2011) states that equating is the process by which scores from test forms are designed to be parallel are made comparable

so that they can be used interchangeably. It refers to a relationship between scores of different forms that are constructed according to the same content and statistical specifications.

Item response theory [IRT] has the under-listed uses in the educational assessment:

1. One of the most important application of IRT modeling is to analyze the performance of test- takers and test items for diagnostic purposes,
2. IRT ensures that students are not been cheated during the process of assessment.
3. IRT is used for research, comprise and equating studies undertaken after examinations. It ensures that standards are compared before examination.
4. It is used to obtain comparable score across the different test forms.
5. It is useful for detection of Differential Item Functioning (DIF) because DIF can be modeled through the use of estimated item parameters and latent traits, and different item functions between two groups can be described in a precise and graphical manner.

A Practical Explanation on the Graphical Pattern of IRT Curves

The following explains very practically the graphical pattern of IRT curves:

Item Characteristic Curve (ICC)

Item characteristic curve (ICC) is a graphical display of students' proficiency (ability) level based on the students' theta (è). After the probabilities of giving the correct answer across different levels of è are combined, the relationship between the probabilities and è is thus presented as an Item characteristic curve (Cohen and Swerdlik, 1999).

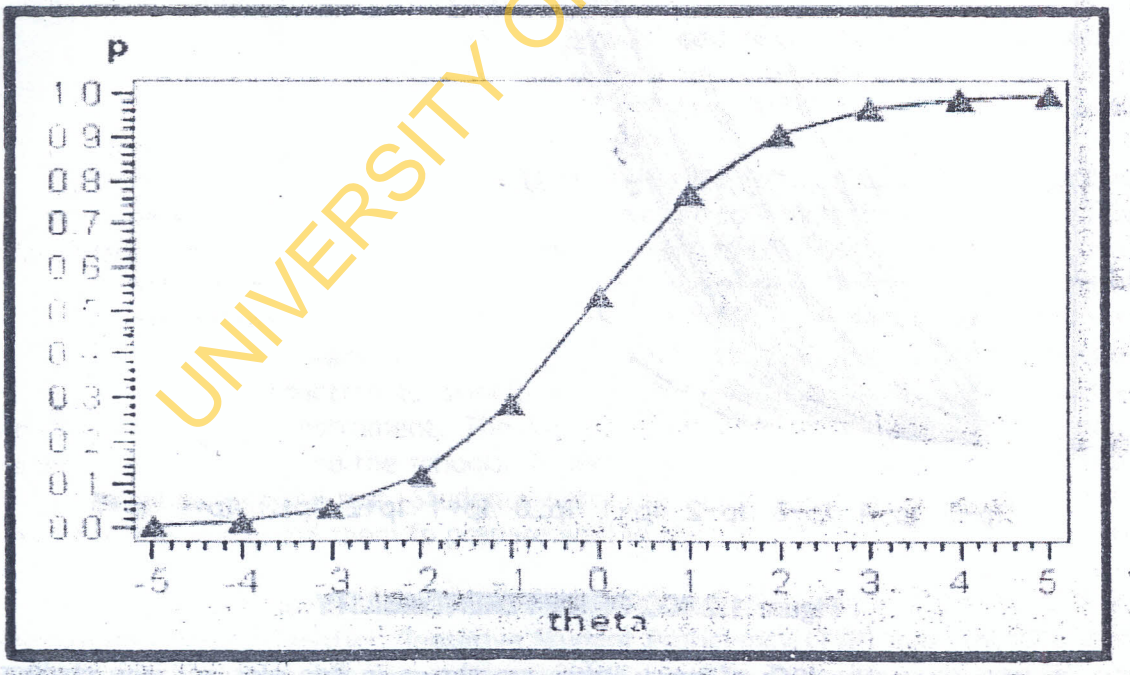

Figure 1.1 ITEM CHARACTERISTIC CURVE (ICC)

From the ICC curve above, it should be noted that the theoretical theta (proficiency) level of a person (performance) of an item ranges from -5 to +5. Chong (2010) found out that the ICC curve indicates that when è was zero (0), which is average, the probability of answering the item correctly is almost .5 and when è is -5, the probability is almost zero. When è is +5, the probability increases to .99. It should be noted that there may not be examinees who can reach proficiency level of +5 or fail so badly as to be in the -5 group.

Item characteristic curves of different ability are vividly illustrated as follows:

The curve shown below is known as the difficulty or threshold parameter and it tells us how easy or how difficult an item is. It is used in the one-parameter (1PLM) IRT model.

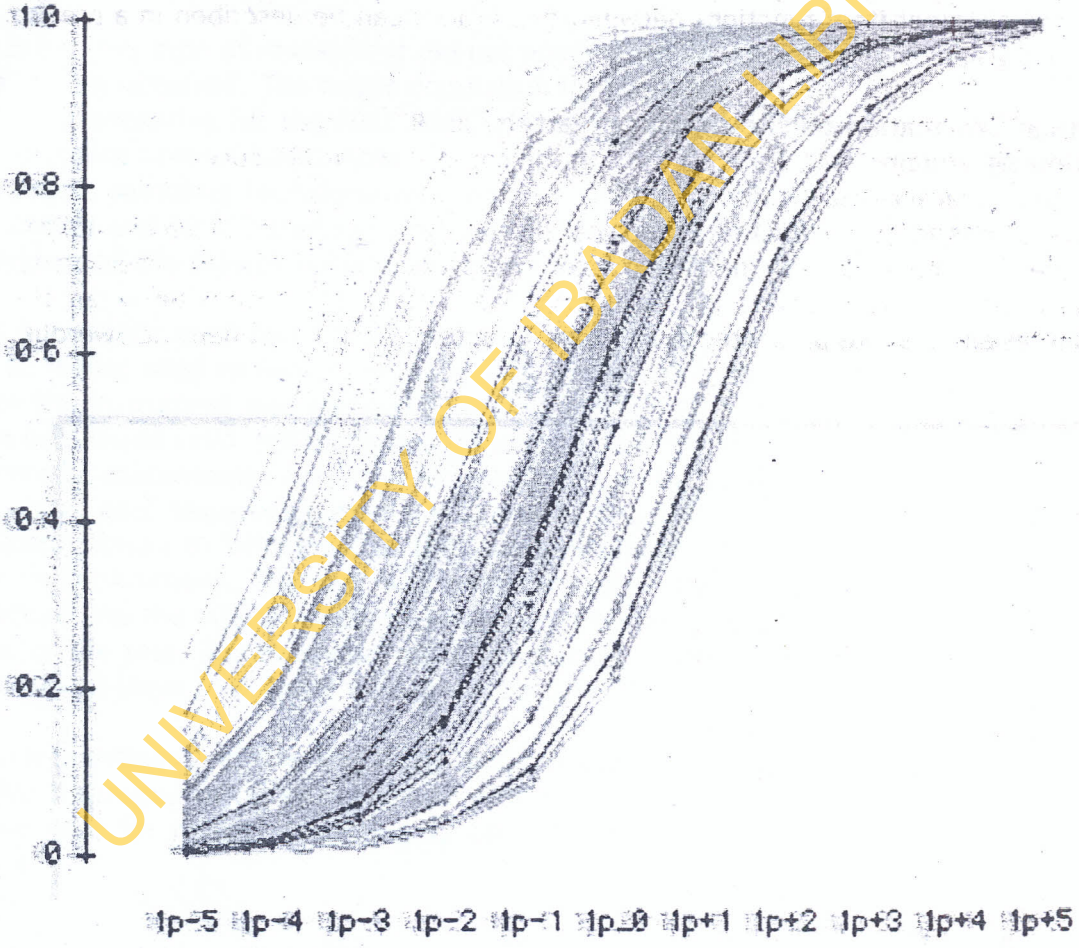

Figure 1.2 ICC OF DIFFERENT ABILITY

In this curve the ICCs of many items are shown in one plot and one obvious characteristic of this plot is that no two ICCs cross over each other.

The assessment of students by WAEC and other examining bodies over the years has been a controversial issue and concern to parents, stakeholders, government, psychometricians, and funding agencies on subjects such as mathematics and English language. The massive failure recorded in these two major subjects has drawn the attention of evaluators and assessor on the need to assess students in other subjects. The noble objective of secondary education can only be achieved if there is an effective and efficient evaluation mechanism that will assess students' without any form of bias. With the complexity in the use of designed software by companies and psycho-metricians to assess student's level on a computer scale, measurement purpose has posed a great threat in the assessment of students, either for grading, promotion, certification or placement purpose. This study therefore, investigated the use of Local Independence of items in WAEC (SSCE) 2010 Economics objective items and how best students can be assessed and evaluated effectively.

Research Questions

1. What is the extent of tentative item difficulty and tentative students' proficiency level in WAEC (SSCE) 2010 Economics objective items?
2. To what extent does local independence of items exist in WAEC (SSCE) 2010 Economics objective items?

Methodology

This study is a survey type of research; it did not manipulate the variables since they were studied as they occurred. The target population for this study consists of all SS3 students that were preparing for the 2012 WAEC (SSCE) Examinations in Ajeromi-Ifelodun Local Government Area of Lagos state.

A simple random sampling technique was used to select fifteen (15) public senior secondary schools and 50 SS3 students were also randomly selected from each school. The breakdown of the figures reflects a total of 350 males (46.7%) and 400 females (53.3%).

The major instrument used for this study is the WAEC (SSCE) 2010 objective Economics paper II. It was used to locate the extent to which the items are locally independent. Since the instrument was adopted from WAEC (SSCE) 2010 examination, the instrument was considered valid, reliable and standardized. The research instrument was adequately scored dichotomously. Each correct option shaded by an examinee was scored "1" (one) while, each incorrect option was scored "0" (zero). The researcher with a letter of introduction to principals of the sampled schools carried out the administration of the instrument. The test was administered in the schools under examination conditions and the schools/students were pre-informed by the researcher on the importance of the test. Students were also informed to respond positively to the test as this would aid them to prepare well in the 2012 WAEC (SSCE) Examination.

The research questions and the statistical analysis and procedure that were used are descriptive statistics, Tentative Student Proficiency (TSP) and Tentative Item Difficulty (TID) was used for research question 1, while Tetrachoric correlation was used for research 2.

Results and Discussion

Table 1.1Tentative Item Difficulty (TID) Summaries of Item Response by Examinees

Items	Right Response	Wron Response	TID	Items	Right Response	Wrong Response	TID
1.	618	132	.176	26.	183	567	.76
2.	155	595	.793	27.	216	534	.71
3.	401	349	.47	28.	246	504	.76
4.	424	326	.43	29.	274	476	.63
5.	169	581	.77	30.	198	552	.74
6.	257	493	.66	31.	186	564	.75
7.	94	656	.87	32.	221	529	.71
8.	81	669	.89	33.	225	525	.70
9.	367	383	.51	34.	176	574	.77
10.	175	575	.77	35.	296	454	.61
11.	234	516	.69	36.	256	494	.66
12.	166	584	.78	37.	228	522	.70
13.	417	333	.44	38.	493	257	.34
14.	285	465	.62	39.	280	470	.63
15.	495	255	.34	40.	185	564	.75
16.	265	485	.65	41.	164	586	.78
17.	123	627	.84	42.	141	609	.81
18.	133	617	.82	43.	186	564	.75
19.	270	480	.64	44.	310	440	.59
20.	151	599	.80	45.	155	595	.80
21.	222	528	.70	46.	174	576	.77
22.	230	520	.69	47.	146	604	.81
23.	400	350	.47	48.	128	622	.83
24.	144	606	.81	49.	153	597	.80
25.	203	547	.73	50.	267	483	.64

Table 1.2 Analysis of Person by Item Matrix Results on WAEC (SSCE) 2010 Objective Economics

ITEMS	1	2	3	4	5	6	7	8	9	...	...	...	45	46	47	48	49	50	TSP
Person 1	1	1	1	1	1	0	1	1	1	...	...	...	1	1	1	1	0	1	86%
Person 2	1	0	1	1	1	0	0	0	1	...	...	...	1	1	1	0	1	0	78%
Person 3	1	1	1	1	1	0	1	1	1	...	...	...	0	0	1	1	0	1	76%
Person 4	1	0	0	1	0	0	1	1	1	...	...	...	0	0	1	1	0	1	68%
Person 5	0	0	1	1	1	1	0	0	1	...	...	...	1	1	0	0	0	0	66%
Person 6	1	0	1	1	1	0	0	0	1	...	...	...	1	1	1	0	0	1	66%
Person 7	1	0	1	1	1	0	0	0	1	...	...	...	1	0	1	0	0	0	62%
Person 8	1	0	1	1	1	0	0	1	1	...	...	...	0	1	0	0	0	1	62%
Person 9	1	1	1	1	1	0	1	0	1	...	...	...	1	1	1	1	0	0	60%
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Person 744	0	0	0	0	0	0	0	0	0	...	...	...	0	0	0	1	0	1	12%
Person 745	0	0	0	0	0	1	0	0	0	...	...	...	0	0	0	0	0	0	12%
Person 746	0	0	0	0	0	0	0	0	0	...	...	...	0	0	0	0	0	1	12%
Person 747	1	0	0	1	0	0	0	0	1	...	...	...	0	0	0	0	0	0	12%
Person 748	0	0	0	0	0	1	0	0	0	...	...	...	0	0	0	0	0	0	10%
Person 749	1	0	1	0	0	0	1	0	0	...	...	...	0	0	0	0	0	0	10%
Person 750	0	0	0	0	0	0	0	0	0	...	...	...	0	0	0	0	0	0	8%
TID	.18	.79	.47	.43	.77	.66	.87	.89	.51	...	...	...	.80	.77	.81	.83	.80	.64	

Table 1.1 shows the TSP and TID on the items and obtainable score. It should be noted that in Rasch Measurement Framework, persons' proficiency is counted based on the number of successful answers, while for item difficulty; it is the number of failures that is counted. The portion of correct answer for each person is called **"Tentative Students' Proficiency (TSP)** and the pass rate for each item is called **"Tentative Item Difficulty" (TID).**It shows a descriptive analysis of the performance of students in the test. Questions 1, 3, 4, 13, 15, 23 and 58 are the easiest because more than 50% of the students got it right. Questions 7 & 18 are the hardest as less than 15% got it right. Questions like these are not useful in distinguishing students who possess high ability and students who possess low ability. Questions like these are therefore omitted during the process of test construction. Table 1.2 shows the item by person matrix on the results that is obtained by students in the test. The **TSP** on the item by person matrix table is similar to the **TSP** on the obtainable score on table 1.1. Chong (2010) stated that scores are tentatively in percentages because in IRT, there is another terminology and scaling scheme for proficiency. The second reason is that, we cannot judge a person's ability just on the number of correct items he obtained. Rather, the item attribute should also be taken into consideration.

In carrying out the analysis, the item scores for all the candidates were used to prepare an item person matrix. The test had a matrix of 750 students on 50 items and the responses of students were scored dichotomously.

Table 2.1 Frequency Distribution of Tetrachoric Correlation on WAEC (SSCE) 2010 Economics Objective Items.

Year	Correlation coefficient	Frequency	Percentage
2010	.5 and above	0	0
	.450 - .499	0	0
	.100 - .449	585	23.9%
	Approximately zero (0)	1865	76.1%

Table 2.2summary of Tetrachoric Correlation for WAEC (SSCE) 2010 Economics Objective Items

item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1																
2	.020	1															
3	.151˙	-.019	1														
4	.216	.029	.207¨	1													
5	-.011	.032	.068	.080˙	1												
6	-0.13	.027	.031	.027	.007	1											
7	-.047	.075˙	-.018	-.042	.018	-.053	1										
8	-.065	-.050	-.020	-.042	.018	-.016	.063	1									
9	.179¨	-.012	.074˙	.164*	.123**	.086˙	-.016	-.023	1								
10	.081˙	.100**	.091˙	.051	.034	-.060	.001	.011	.078˙	1							
11	.054	.097*	.005	.068	.133¨	.011	-.064	.044	.095˙	.003	1						
12	-.023	.037	.053	.079˙	.028	.014	.060	.011	-.001	-.051	.003	1					
13	.003	-.034	.049	.072˙	.032	.051	-.043	-.009	.177**	-.046	-.051	.057	1				
14	.088*	-.013	.064	.199˙	.058	.031	-.047	-.042	.140¨	.120˙	.018	-.093	.003	1			
15	.215¨	-.022	.160˙	.239˙	.077¨	-.045	-.094	-.086	.247¨	.030	.088	.037	.188	.208	1		
16	.027	-.047	.041	.052	-.005	.025	-.027	.030	.102¨	.067	.032	.056	.122	-.021	.136	1	
17	-.016	-.075	.059	.033	.054	-.009	.017	.020	.136¨	-.049	.005	-.054	.048	-.057	-.017	-.019	1
18	-.033	.056	-.106	-.001	.025	-.34	.003	-.027	.000	-.025	-.049	-.020	.007	-.040	-.072	-.051	.064
19	.121*	.070	.059	.170˙	.161¨	.032	-.032	.008	.210¨	.026	.118	..022	.161	.168	.263	.103	.023

The result on the analysis on table 2.1 shows the correlation coefficient among items reflecting the extent of local independence on the items. Using Tetrachoric correlation, the coefficient of the correlation was then used to determine the extent at which the items are locally independent. The 50 items used resulted to 2450 correlations among the items. Each item correlates perfectly with itself (1) but at different extent with others. From the table; the correlation between item 1 and item 2 was .020; item 2 and item 3 was -.019; item 3 and item 4 was .207¨; item 4 and item 5 was .080˙. Also the correlation coefficient between item 7 and item 8 was .063; item 9 and item 10 was .078˙. This inter item correlation process goes on through all the items.The answer to this research question has revealed that WAEC (SSCE) 2010 objective items on Economics were to a great extent locally independent. It is important to note that items with lower Tetrachoric correlations are better than those with higher local independence. The result shows that a significant number of the correlations were

approximately zero (0) and this implies that such items are not related and may not have acted as clues to one another during the testing session.

This study is in line with Zenisky, Hambleton and Sireci (2003) who analyzed items associated with reading passages and found that when the items were (improperly) treated as discrete, locally independent items, test information functions and reliability estimates were overestimated. This is a serious problem, especially in computer-adaptive testing (CAT) where the standard error of the estimate (SEE) is often used as the termination criterion.

The study is also in line with Lee & Frisbie (1999) who also computed average within and between-testlet correlations in their generalizability theory approach to assessing the reliability of tests composed of testlets. When testlet scoring was used on the sets of items in their research, the difference between the computed passage reliability and the generalizability coefficient was small, supporting the position that testlet scoring was the appropriate level of scoring to use as compared to dichotomous item scoring.

The findings from this study is also in line with Essen (2009) who carried out a study on item local independence and students' performance in NECO mathematics. In his study, he observed that there were a total of 2450 correlations, and of this number, about 1182 (48.24%) were significant at .01, about 262 (10.70%) were not significant at .05, while 1006 (41.06%) were not significant. The items in the test on the study he investigated are said not to have local independence.

Conclusion and Recommendations

Conclusion

With so much emphasis on teachers' effectiveness, personality style, teachers characteristics and students' attitude as correlate on different factors, researchers in the education sector should have a re-think on how best students' performance can be assessed effectively with the use of IRT.

Based on the findings of this study, it can be concluded that:

1. WAEC items are properly validated before being tested on examinees.
2. The "clarion call" on IRT will serve as an eye opener to evaluators, measurement experts, test developers/constructor, examining bodies and psycho-metricians.
3. The use of IRT is invaluable for evaluating the psychometric properties of WAEC, NECO and JAMB.

Recommendations

The importance of economics towards national growth and development in the real per capita income of a nation's economy, can lead to sustainability in the country's national income, policy and the international economy. The following recommendations were made for improved quality in the educational system:

1. Examining bodies, government, test developers, private firms, psychometricians and stakeholders should venture into the use of IRT in assessing students' and also in the construction and validation of test items.
2. IRT software's should be made available by examining bodies, Ministry of Education, Institutes of Education, Universities and evaluation agencies.

3. Measurement and evaluation course should be reviewed to inculcate more practice than theory base in our tertiary institutions of learning. Students should be sent to examining bodies to undergo some form of intern or training programme on test construction.
4. Professionals with experience in the field of measurement, evaluation and test development should be called in the training of teachers, lecturers, psychometricians on the general principles of testing and measurement through seminar and workshop.
5. It is strongly recommended that examination questions be validated by appropriate test experts, pilot tested to ensure maximum reliability coefficient, tried out as much as possible, and banked for use.

References

Chong Ho Yu, (2010). A Simple Guide to the Item Response Theory (IRT) and Rasch modeling.Retrieved from www.creative-wisdom.com on 23/11/2011.

Cohen, R.J & Swerdlik, M.E, (1999). *Psychological Testing and Assessment. An Introduction to Test and Mearsurement*. (4th edition), Mayfield publishing company, Mountain view, California, USA.

Emeke, E.A & Edward, M.O,(2011). Linking and Equating Tests; Needed Revisit by Large Scale Educational Assessment Bodies. A paper presented at the 29th conference of the Association for Educational Assessment in Africa, Nairobi, Kenya.

Essen, C.B (2009). Item Local Independence and Students' Performance in Multiple-Choice Mathematics in Ikot-Abasi LGA, Akwa-Ibom state. An M.Ed Unpublished research work, University of Calabar.

Greaney, V & Kellanghan, T, (2008). Assessing National Achievement Levels in Education. *A World Bank report on National Achievement Assessment of Educational Assessment in Africa*. Washington, DC

Hambleton, R.K & Jones, R.W, (1993). Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development. *Educational Measurement: Issues and Practice*, 12(3), 3847

Kpolovie, P. J., Ololube, N. P. & Ekwebelem, A. B. I.,(2011). Appraising the Performance of Secondary School Students on the WAEC and NECO SSCE from 2004 to 2006.

Lee, G., & Frisbie, D. A (1999).Estimating Reliability under a Generalizability Theory Model for test scores composed of testlets. *Applied Measurement in Education*, 12, 237-255.

Sireci, S.G., Thissen, D, & Wainer, H., (1991). On the Reliability of Testlet-based Tests.*Journal of Educational Measurement*. 28(3), 237-247.

Thordike, R.M, (1997). *Measurement and Evaluation in Psychology and Education*. (6thEdition). Prentice Hall Inc. publishing press, Upper Saddle River, New Jersey.

Todaro, M.P & Smith, S.C, (2006). *Economic Development*. (9th Edition). Pearson Educational Ltd, Edinburgh Gate, Harlow, England.

Ubi, I.O, Joshua, M.T & Umoinyang, I.E, (2011). Item Local Independence in selection Examination in Nigeria: implication for Assessment for Regional Integration. A paper presented at the 29th conference of the Association for Educational Assessment

in Africa, Nairobi, Kenya.

UNESCO, (2000). *The Dakar Framework for Action—Education for All (EFA): Meeting Our Collective Commitments.* Paris: UNESCO.

World Bank, (1996). Nigeria: Poverty in the midst of plenty. The challenge of growth with inclusion. A World Bank poverty report assessment report. Report No. 14733 UNI

Xitao, F, 1998. Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and Psychological Measurement*; 58(3); pg 357 (25).

Yung-chen Hsu, 2007. EZLID: A SAS Macro for local item dependence assessment. A paper presented by testing services. *American Council on Education*, Washington, DC &Tsuug-hsun Tsai, research league, LLC, Morganville, NJ.

Zenisky, A.L, Hambleton, R.K & Sireci, S.G, (2003). Effects of Item Local Dependence on the Validity of IRT Item. *Test and Ability Statistics*. University of Massachusetts, Amherst.