RELIABILITY INDEX OF 2015 WEST AFRICA EXAMINATION COUNCIL ECONOMICS OBJECTIVE ITEM BASED ON NUMBER OF OPTIONS USING KUDER-RICHARDSON FORMULA.

Jimoh Kasali[a] & Yinusa Akintoye Faremi[b]

jimoh.bukola@yahoo.com

[a]Department of Educational Foundations and Counselling, Obafami Awolowo University, Ile-Ife

[b]Department of Educational Foundations and Management University of Eswatini Kwaluseni Campus Eswatini

Abstract

The study examined the reliability index of 2015 West Africa Examination Council Economics objective items. It determined the reliability coefficient of 3, 4 and 5 options using Kuder-Richardson Formula. Three objectives were raised and three research questions were used to guide the study. The study adopted survey design. The population of the study consisted of secondary school students in Osun State. The study sample consisted of 360 student's selected using multistage sampling procedure. The instrument used for this study was an adapted version of the 2015 West Africa School Certificate Examination (WASCE) Economics test items. The 4-options WASCE Economics items were adopted while its 3- and 5-options items were adapted. Data collected were analysed using Kuder Richardson Formula (KR-20) and Fisher's Z-Test with aid of FZT compotator. The results of the study showed that the reliability coefficient of 3-options items was 0.26; while the reliability index of 4-options item was 0.68 and 5-options had reliability coefficient of 0.62. This implies that 4-options Economics test items are more reliable than 3 and 5-options when number right scoring procedure were used. The study therefore concluded that four option items especially in multiple choice Economics tests should be encouraged but if five options items should be used more attention should be given to psychometric properties of the tests.

Keywords: Economics, Objective Item, Number right, Kuder Richardson Formula 20

Introduction

Economics is a social science subject. According to Dwivedi (2004), subject of Economics is a vital discipline in any nation's socioeconomic growth. Individuals, students, groups, associations, political classes, and even governments, according to Davies in Oleabhiele (2012), agrees that Economics can be used to make decisions, distribute, and economize resources for the common good. Accordingly, the Federal Republic of Nigeria (2004) claims that teaching economics in secondary schools will provide students with the understanding of how to allocate scarce resources, make decisions, and make reasonable decisions on critical economic issues. The Comparative Education Study and Adaptation Centre (CESAC) created the subject's curriculum in 1985, and it is presently reviewed by the National Education Research Development Centre (NERDC). The curriculum is founded on the premise of providing recipients with the fundamental information and skills necessary to understand the nature of economic problems in any society and to effectively prepare them for the difficulties that the Nigerian economy presents (NEDRC, 2008). Because of their capacity to accurately and efficiently measure dimensions such as ability and achievement, multiple-choice items have continued to dominate educational testing. For obvious reasons, measurement specialists and testing companies favour multiple-choice item formats to others (such as short-answer, essay, and constructed-response). This can be a very effective item format if item writers are thoroughly taught and things are quality ensured. First and foremost, students do better on multiple choice exams than on any other

sort of objective examination when they are taught how the item format works and the fallacies around the assessment type are debunked. On many tests, reliability has been proven to improve as the number of questions on the test increases, and with adequate sampling and attention to case specificity, overall test reliability of multiple-choice items can be raised even more (Murayama, 2009).

Multiple choice exams take less time to administer for a given amount of content than tests requiring written replies, allowing for more questions to be included in the assessment without increasing the time required; this allows for a more thorough examination of the candidate's knowledge. The utilization of online examination delivery software can help you achieve even more efficiency. Multiple choice tests are ideal for developing objective evaluation items. Because this type of test does not require a teacher to mark the offered answers arbitrarily, test-takers are scored only on their choices, reducing the chance of results including teacher-student bias. Scoring is objective because it is unaffected by factors such as the examinee's poor handwriting. Multiple-choice questions are the most widely employed to assess linguistic competency, according to Anna and George (2015), since they are quick, inexpensive, and simple to grade. A typical multiple choice test item has two parts: a stem (a question or a problem to be solved) and a list of alternative solutions, which usually includes one correct (or "best") answer and several erroneous ones. Importantly, unlike essay questions, the examinations do not explicitly disadvantage students with poor reading skills. Misreading one multiple-choice question

can cost a student a small percentage of their grade; however, misreading an essay question can cost a lot of money. As a result, students with strong intellectual and conceptual abilities but poor writing abilities are not disadvantaged when it comes to multiple-choice questions. In comparison to essays that may focus on a certain topic, multiple-choice questions can provide a greater range and depth of coverage of content.

The key and 3-4 distractors are usually included in multiple choice test items. However, research has shown that using fewer options is more effective. It is difficult to construct test items with more than one plausible distractor, according to Haladyna and Downing (1993), resulting in items with a right answer and one alternative, also known as the alternate choice (AC) format. Several research have looked into the time savings in test administration when there are fewer options. According to Costin (1972), pupils can finish things with three options faster than those with four possibilities. This makes logical because the amount of time spent reading and evaluating should be reduced. A meta-analysis of the impact of different test item features on test scores and test completion durations was undertaken by Aamodt and McShane (1992). They discovered that three-option tests took much less time to complete than four-option assessments. Schneid et al. (2014) used a computerized testing approach to collect data on time to finish each item in a pharmacology exam,the authors discovered that students answered three-option MCQs five seconds faster than four- or five-option questions.

The consistency of particular instruments in delivering the same result in repeated

measurements is referred to as reliability. Reliability is an instrument used to analyze identical measurements, if it produces the same result every time Sabri (2013). The classic definition of reliability is concerned with the consistency and repeatability of measurements over time and among various people. A test can be highly trustworthy in one situation and completely unreliable in another (Gilbody, Morley & Snaith., 2006). Internal consistency indicates that the items in a measure are related to one another; test-retest reliability is established by administering the same measure to a group of people on two occasions, separated by a specified period of time; and inter-rater reliability is established by comparing the ratings of two or more independent evaluators.

The reliability estimates developed by Kuder and Richardson (1937) have been frequently employed by test manufacturers among the different statistical methods for determining a test's internal consistency dependability. The Kuder-Richardson reliability estimates only require the administration of a single test and eliminate any biases that may develop when a test is split in any of a number of ways, such as the split-half approach. The most precise Kuder-Richardson formula, also known as K-R 20, is as follows:

$$KR20 = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum pq}{\sigma 2}\right) \quad \text{..... equ (i)}$$

Where, k is the number of items contained in the test, p is the proportion of examinees who got the item correctly, q is the proportion of examinees who got the item wrongly and σ^2 is the variance of the total test score. The problem of reliability estimation, and in particular Cronbach's alpha, has sparked a lot of discussion in the psychometric literature recently. Alpha has been chastised for making assumptions that aren't based on reality (McNeish, 2018; Schmitt,1996; Sijtsma,

2009).

The case for three-option items, the four- and five-option MCQ remains the prevailing choice for high-stakes (e.g., credentialing and education) testing, according to the literature review. Despite research indicating no significant changes in item discrimination, item complexity, or test reliability for tests using the three-option MSQ format vs the four- or five-option MSQ formats, this is still the case (Schneid et al., 2014). As a result, the dependability index of 2015 West Africa Examination Council Economics objective items was investigated in this study.

Objectives of the Study

The objectives of the study are to:

a. Determine the reliability index of 3-options 2015 WAEC Economics objective items using Number right scoring procedure
b. Determine the reliability index of 4-options 2015 WAEC Economics objective items using Number right scoring procedure
c. Determine the reliability index of 5-options 2015 WAEC Economics objective items using Number right scoring procedure

Research Questions

a. What the reliability coefficient of 3-options 2015 WAEC Mathematics objective items using Number right scoring procedure

b. What the reliability coefficient of 4-options 2015 WAEC Mathematics objective items using Number right scoring procedure
c. What the reliability coefficient of 5-options 2015 WAEC Mathematics objective items using Number right scoring procedure

Methodology

The study adopted survey design. The population of the study consisted of secondary school students in Osun State. The study sample consisted of 360 students selected using multistage sampling procedure. The three senatorial districts in the State include Osun Central Senatorial District, Osun East Senatorial District and Osun West Senatorial District. From the three senatorial districts in the State, two Local Government Areas (LGAs) were selected using simple random sampling technique. From each of the two LGAs selected, three schools were also selected randomly to make a total of 18 schools. From each school 20 Senior Secondary two (SSII) were selected using purposive sampling technique, being best 20 students in a pre-test in each school for the study. Data collected were analysed using Kuder and Richardson Formula 20. Prior to the analysis of the data, the responses of the examinees was scored dichotomously

Results

Research Question 1: What the reliability coefficient of 3-options 2015 WAEC Mathematics objective items using Number right scoring procedure

To answer this research question, the Kuder and Richardson Formula 20 reliability index, an internal consistency of measurements with dichotomous choices (i.e. correct versus

incorrect) with the formula;

$$KR20 = (\frac{k}{k-1})(1 - \frac{\sum pq}{\sigma 2})$$

Where, k is the number of items contained in the test, p is the proportion of examinees who got the item correctly, q is the proportion of examinees who got the item wrongly and σ^2 is the variance of the total test score.

In order to estimate the reliability of each of three options multiple choice test items when scored using number right scoring, the p, q, sum of all pq's and σ^2 determined and finally the reliability estimates were determined using Microsoft Excel Package. The differences in the estimated reliabilities were determined using Fishe'sr Z-Test with aid of FZT compotator. The result is as presented in Table 1

Table 1: Reliability index of 3-Options score using Number Right

Item Number	P	Q	Pq	Item Number	P	Q	pq	Variance	rhoKR20
1	0.43333	0.56667	0.24556	26	0.16667	0.83333	0.13889		
2	0.13333	0.86667	0.11556	27	0.65	0.35	0.2275		
3	0.25	0.75	0.1875	28	0.25	0.75	0.1875		
4	0.3	0.7	0.21	29	0.5	0.5	0.25		
5	0.36667	0.63333	0.23222	30	0.28333	0.71667	0.20306		
6	0.36667	0.63333	0.23222	31	0.36667	0.63333	0.23222		
7	0.2	0.8	0.16	32	0.11667	0.88333	0.10306		
8	0.36667	0.63333	0.23222	33	0.23333	0.76667	0.17889		
9	0.41667	0.58333	0.24306	34	0.48333	0.51667	0.24972		
10	0.33333	0.66667	0.22222	35	0.45	0.55	0.2475		
11	0.23333	0.76667	0.17889	36	0.4	0.6	0.24		
12	0.21667	0.78333	0.16972	37	0.28333	0.71667	0.20306		
13	0.5	0.5	0.25	38	0.48333	0.51667	0.24972		
14	0.21667	0.78333	0.16972	39	0.4	0.6	0.24		
15	0.46667	0.53333	0.24889	40	0.45	0.55	0.2475		
16	0.66667	0.33333	0.22222	41	0.46667	0.53333	0.24889		
17	0.38333	0.61667	0.23639	42	0.55	0.45	0.2475		
18	0.46667	0.53333	0.24889	43	0.58333	0.41667	0.24306		
19	0.43333	0.56667	0.24556	44	0.41667	0.58333	0.24306		
20	0.36667	0.63333	0.23222	45	0.36667	0.63333	0.23222		
21	0.56667	0.43333	0.24556	46	0.38333	0.61667	0.23639		
22	0.56667	0.43333	0.24556	47	0.28333	0.71667	0.20306		
23	0.38333	0.61667	0.23639	48	0.2	0.8	0.16		
24	0.25	0.75	0.1875	49	0.4	0.6	0.24		
25	0.36667	0.63333	0.23222	50	0.41667	0.58333	0.24306		
							10.9261	14.64972	0.25936

The result showed that the reliability coefficient estimated using KR20 was 0.26., under 3-option 2015 WAEC Economics Objective items when number right scoring was used

Research Question 2: What the reliability coefficient of 4-options 2015 WAEC Mathematics objective items using Number right scoring procedure

To answer research question 2, the Kuder and Richardson Formula 20 reliability index, an internal consistency of measurements with dichotomous choices (i.e. correct versus incorrect) with the formula;

$$KR20 = (\frac{k}{k-1})(1 - \frac{\sum pq}{\sigma 2})$$

Where, k is the number of items contained in the test, p is the proportion of examinees who got the item correctly, q is the proportion of examinees who got the item wrongly and σ^2 is the variance of the total test score.

In order to estimate the reliability of each of three options multiple choice test items when scored using number right scoring, the p, q, sum of all pq's and σ^2 determined and finally the reliability estimates were determined using Microsoft Excel Package. The differences in the estimated reliabilities were determined using Fishe'sr Z-Test with aid of FZT compotator. The result is as presented in Table 2

Table 2: Reliability index of 4-Options score using Number Right

Item Number	P	Q	Pq	Item Number	P	Q	pq	Variance	rhoKR20
1	0.45763	0.54237	0.24821	26	0.4237	0.576	0.24418		
2	0.32203	0.67797	0.21833	27	0.4576	0.542	0.24821		
3	0.42373	0.57627	0.24418	28	0.2203	0.78	0.17179		
4	0.33898	0.66102	0.22407	29	0.3559	0.644	0.22924		
5	0.49153	0.50848	0.24993	30	0.3898	0.61	0.23786		
6	0.62712	0.37288	0.23384	31	0.2542	0.746	0.1896		
7	0.38983	0.61017	0.23786	32	0.6102	0.39	0.23786		
8	0.23729	0.76271	0.18098	33	0.3559	0.644	0.22924		
9	0.64407	0.35593	0.22924	34	0.339	0.661	0.22407		
10	0.25424	0.74576	0.1896	35	0.4407	0.559	0.24648		
11	0.28814	0.71186	0.20511	36	0.3051	0.695	0.21201		
12	0.32203	0.67797	0.21833	37	0.5763	0.424	0.24418		
13	0.52542	0.47458	0.24935	38	0.4746	0.525	0.24935		
14	0.44068	0.55932	0.24648	39	0.2881	0.712	0.20511		
15	0.49153	0.50848	0.24993	40	0.3898	0.61	0.23786		
16	0.30509	0.69492	0.21201	41	0.3729	0.627	0.23384		
17	0.37288	0.62712	0.23384	42	0.4237	0.576	0.24418		
18	0.54237	0.45763	0.24821	43	0.3729	0.627	0.23384		
19	0.35593	0.64407	0.22924	44	0.6441	0.356	0.22924		
20	0.45763	0.54237	0.24821	45	0.4407	0.559	0.24648		
21	0.42373	0.57627	0.24418	46	0.4576	0.542	0.24821		
22	0.42373	0.57627	0.24418	47	0.3559	0.644	0.22924		
23	0.45763	0.54237	0.24821	48	0.4407	0.559	0.24648		
24	0.27119	0.72881	0.19764	49	0.4068	0.593	0.24131		
25	0.45763	0.54237	0.24821	50	0.4068	0.593	0.24131		
							11.5806	34.6133	0.679

The result revealed that the reliability of the 2015 WAEC Economics objective test for 4-options scoring method was 0.68.

Research Question 3: What the reliability coefficient of 5-options 2015 WAEC Mathematics objective items using Number right scoring procedure

To answer this research question 3, the Kuder and Richardson Formula 20 reliability index, an internal consistency of measurements with dichotomous choices (i.e. correct versus incorrect) with the formula;

$$KR20 = (\frac{k}{k-1})(1 - \frac{\sum pq}{\sigma 2})$$

Where, k is the number of items contained in the test, p is the proportion of examinees who got the item correctly, q is the proportion of examinees who got the item wrongly and σ^2 is the variance of the total test score.

In order to estimate the reliability of each of three options multiple choice test items when scored using number right scoring, the p, q, sum of all pq's and σ^2 determined and finally the reliability estimates were determined using Microsoft Excel Package. The differences in the estimated reliabilities were determined using Fishe'sr Z-Test with aid of FZT compotator. The result is as presented in Table 3

Table 3: Reliability index of 5-Options score using Number Right

Item Number	P	Q	Pq	Item Number	P	Q	pq	Variance	rhoKR20
1	0.40678	0.59322	0.24131	26	0.1864	0.814	0.15168		
2	0.25424	0.74576	0.1896	27	0.5763	0.424	0.24418		
3	0.20339	0.79661	0.16202	28	0.1864	0.814	0.15168		
4	0.42373	0.57627	0.24418	29	0.2373	0.763	0.18098		
5	0.40678	0.59322	0.24131	30	0.1525	0.847	0.12927		
6	0.32203	0.67797	0.21833	31	0.2373	0.763	0.18098		
7	0.23729	0.76271	0.18098	32	0.2881	0.712	0.20511		
8	0.30509	0.69492	0.21201	33	0.2373	0.763	0.18098		
9	0.22034	0.77966	0.17179	34	0.1525	0.847	0.12927		
10	0.15254	0.84746	0.12927	35	0.3559	0.644	0.22924		
11	0.23729	0.76271	0.18098	36	0.2456	0.754	0.18529		
12	0.27119	0.72881	0.19764	37	0.1186	0.881	0.10457		
13	0.23729	0.76271	0.18098	38	0.1864	0.814	0.15168		
14	0.33898	0.66102	0.22407	39	0.1579	0.842	0.13296		
15	0.35593	0.64407	0.22924	40	0.3158	0.684	0.21607		
16	0.33898	0.66102	0.22407	41	0.1404	0.86	0.12065		
17	0.33898	0.66102	0.22407	42	0.2632	0.737	0.19391		
18	0.16949	0.83051	0.14076	43	0.386	0.614	0.237		
19	0.25424	0.74576	0.1896	44	0.5439	0.456	0.24808		
20	0.28814	0.71186	0.20511	45	0.1228	0.877	0.10773		
21	0.23729	0.76271	0.18098	46	0.386	0.614	0.237		
22	0.25424	0.74576	0.1896	47	0.2105	0.789	0.16621		
23	0.20339	0.79661	0.16202	48	0.3333	0.667	0.22222		
24	0.23729	0.76271	0.18098	49	0.5614	0.439	0.24623		
25	0.27119	0.72881	0.19764	50	0.2456	0.754	0.18529		
							9.43685	24.16	0.623

The result revealed that the reliability of the 2015 WAEC Economics objective test for 5-options scoring method was 0.62.

Discussion of the Findings

The goal of this study was to use the Kuder-Richardson Formula to establish the dependability coefficient of 3, 4, and 5 alternatives. The study's findings revealed that the reliability coefficient for three-option items was 0.26, while the dependability index for four-option items was 0.68, and the reliability coefficient for five-option items was 0.62. When the number right scoring system is applied, this means that 4-option Economics test items are more reliable than 3 and 5-options. This outcome is consistent with Owen and Froman's (1987) findings, which indicated no differences in item discrimination, item difficulty, or test scores between the three and five-option MSQ tests. Sidick et al. (1994) discovered no practical changes in psychometric qualities between three- and five-option job examinations. The outcomes of this study further support Costin's (1970) observation that mean discrimination indices for the three-option item test testing student knowledge of psychology were greater than for the four-option item test. Rodriguez (2005) observed increases in item discrimination and reliability for three-option versus four-option MC tests in a meta-analysis covering eighty years of research on multiple-choice items.

Conclusion and Recommendations

Based on the finding of this study the four option multiple choice items are using Kuder-Richardson 20 was more reliable that three and five options. Hence, the authors recommended that that four option items especially in multiple choice Economics tests should be encouraged but if five options items should be used more attention should be

given to psychometric properties of the tests

References

Aamodt, M., & McShane, T. (1992). A metea-analytic investigation of the effect of various test item characteristics on test scores and test completion times. Public Personnel Management, 21(2), 151-160

Costin, F. (1970). The optimal number of alternatives in multiple-choice achievement tests: Some empirical evidence for a mathematical proof. Educational and Psychological Measurement, 30, 353-358

Costin, F. (1972). Three-choice versus four-choice items: Implications for reliability and validity of objective achievement tests. Educational and Psychological Measurement, 32, 1035-1038.

Dwivedi, D. N (2004) Management Economics. (6th edition). London: Martins Press Inc

Federal Republic of Nigeria. (2004). National Policy on Education. Lagos: NERDC Press.

Gilbody S, Morley S, Snaith P (2006) Principles of psychological assessment. In Freeman C, Tyrer P (eds), Research methods in psychiatry, 3rd ed., pp. 160-182. The Royal College of Psychiatrists

Haladyna, T. M. (1993). Developing and validating multiple-choice test items (3rd Ed.). Mahwah, NJ: Lawrence Erlbaum.

Kuder, G. F., & Richardson, M. W. (1973).Theory of the test reliability, Psychometrika 1(2), 151-160.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. Psychological Methods, 23 (3), 412-433.

Murayama, K (2009). Improving your test questions. Center for Innovation in Teaching and Learning, University of Illinois.

Owen, S., & Froman, R. (1987). What's wrong woth three option multiple choice items? Educational and Psychological Measurement, 47, 153-522

Oleabhiele, E. O (2012) Effect of individualised and cooperative teaching/learning

strategies on secondary school students' achievement in Economics. Unpublished Ph.D thesis**.** Ebonyi State University.

Rodriquez, M. C. (2005). Educational Research Newsletter. Retrieved 25th March, 2005 with the Rasch model. Journal of Language Teaching and Research, 2(5). 234-236

Sabri, S (2013). Item analysis of student comprehensive test for research in teaching beginner string ensemble using model based teaching among music students in public universities. International Journal of Education and Research, 1(12) 1-14

Schmitt, N. (1996). Uses and abuses of coecient alpha. Psychological Assessment, 8, 350-353.

Schneid, S., Armour, C., Park, Y., Yudkowsky, R., & Bordage, G. (2014). Reducing the number of options on multiple-choice questions: response time, psychometrics, and standard setting. Medical Education, 48, 1020-1027. Doi: 10.1111/medu.12525\

Sidick, J., Barrett, G., & Doverspike, D. (1994). Three-alternative multiple choice tests: An attractive option. Personnel Psychology, 47(4), 829-835

.Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of cronbach's alpha. Psychometrika , 74 (1), 107 -120.

.