

UTILIZING R SOFTWARE FOR THE DEVELOPMENT AND VALIDATION OF ECONOMICS ACHIEVEMENT TEST: A CLASSICAL TEST THEORY APPROACH

Hager Atisi Eremina Amakiri

*Department of Educational Psychology, Guidance and Counselling,
Faculty of Education, Ignatius Ajuru University of Education, Nigeria.*

hager.amakiri@iaue.edu.ng

ORCID: 0009-0004-9262-3770

Ibiene Tandi Longjohn

*Department of Educational Psychology, Guidance and Counselling,
Faculty of Education, Ignatius Ajuru University of Education, Nigeria.*

tandilongjohn@gmail.com

ORCID: 0009-0000-8328-4800

Allwell Sunny Njigwum

*Department of Educational Psychology, Guidance and Counselling,
Faculty of Education, Ignatius Ajuru University of Education, Nigeria.*

allwellsn@gmail.com

ORCID: 0000-0001-8402-551X

ABSTRACT

This study investigated the utilization of R software for the development and validation of Economics achievement test: a classical test theory approach. The study adopted an instrumentation research design using Classical test theory to construct and validate an Economics Achievement Test for senior secondary schools in Rivers State, Nigeria. The initial test was administered to 400 Senior School One (SS1) students that were randomly selected from six (6) public schools during the third term of the 2020/2021 academic session in Etche Local Government Area, Rivers State. Three research questions guided the study. The EAT covered a total of fourteen (14) topics drawn from the SS1 Economics curriculum in Nigeria from which 150 items were generated. The content validity of the EAT was established using a test blueprint. The data collected were coded using SPSS and transferred to R-software for item analysis (item difficulty and item discrimination) under the CTT framework. The reliability of the initial test was estimated using Cronbach alpha and Kuder-Richardson-20 reliability methods in R-Studio software which produced a reliability coefficient of 0.963. The result indicated that 63 items (42.3%) were considered good items to form the final test, while 86 items (57.7%) were marked for rejection. From the selected items, the difficulty indices ranged from .30 to .70 while discrimination indices ranged from .22 to .55. The reliability of the final test produced a coefficient of 0.92 for both Cronbach alpha and Kuder-Richardson-20 reliability methods while split-half reliability produced a coefficient of 0.94 indicating that the selected 63 items were reliable. Also, a SEM of 3.29 and 2.911 was established based on the respective reliability coefficients. The EAT is therefore both valid and reliable hence, it is recommended for use to test achievement in Economics among secondary school students in Rivers State.

Keywords: Validation, Economics, Classical Test Theory, Achievement Test.

1. Introduction

The application of Economics as a discipline occupies a very key position in the everyday life of humans and society as a whole. Economics is considered an important subject and is taught at the senior secondary school level. It is a subject concerned with the efficient utilization

or management of limited productive resources for the purpose of attaining the maximum satisfaction of human wants (Ochuba in Chime 2012). Economics is a social science which studies human behaviour as a relationship between ends and scarce means which have alternative uses. Economics is social sciences that is concerned with

human behaviours such as how people earn their living and make alternative choices with regard to scarcity in order to satisfy their wants. Economics can also be defined as a social science concerned with the production, distribution and consumption of goods and services with main concerns on how individuals, businesses, governments and nations make choices on how best to allocate resources, (Hayes, 2020).

Ever since the introduction of Economics into the secondary school curriculum in Nigeria, the number of schools that teach Economics and students that take the examination have witnessed an unpredictable increase because economic issues concern every sphere of human life especially in modern society be it business, government, finance, production (Onyekachi, 2018). Jhingan (2010) noted that the curriculum has been designed by the Comparative Education Study and Adaptation centre (CESAC) to meet the requirements of economics in the new system. In light of this philosophy, the following are the set objectives for the Economics course:

To equip students with the basic principles of economics necessary for useful living and for higher education.

To prepare and encourage students to be prudent and effective in the management of scarce resources

To raise students' respect for the dignity of labour and their appreciation of economic, cultural and social values of our society and

To enable students to acquire knowledge for the practical solution of the economic problems of society; Nigeria, as developing country and the world at large.

For the teaching of Economics to be effective, much depends on the ability of the teacher who handles the teaching and learning process. The teacher being the principal actor in the classroom needs to be well equipped in the subject matter content as well as the pedagogy that will enable the students to develop a

sustained interest in the subject. Since learning is a change in behaviour, there is the need to assess on a regular basis the extent to which the learning has made for the needed change in behaviour as stated in the learning objectives by the teacher. Doing this requires the use of valid and reliable achievement test items that are of good psychometric properties. Achievement tests are administered to pupils to determine how much they have learned or mastered following the teaching-learning process (Chikwe, 2017). Achievement tests are designed to assess the degree to which educational objectives in a content or subject area have been met (Orluwene, 2012). An achievement test is either teacher-made or standardized. The classroom teacher creates teacher-made tests to assess specific topics presented by the teacher. While standardized tests are developed by subject specialists and test experts (psychometricians) under a uniform set of conditions for use on a larger scale, they have established psychometric qualities (validity and reliability) (Asuru, 2015). However, Asuru et al. (2023) affirmed that most locally developed teacher-made tests do not go through the rigorous validation procedure, hence lack known psychometric properties. Similarly, Longjohn et al. (2021) found that most classroom teachers lack the knowledge to create valid and reliable tests for their ongoing assessment as well as end-of-term internal examinations.

In order to construct achievement tests, Mahajon (2005) as cited by Obilor (2020) and Asuru (2015) recommended the following steps to be adopted: planning of the test, preparation of the test, administration of the test/pilot testing/test try out, final try out of the test, preparing final form of test and establishing reliability and validity. However, Onunkwo (2012) published a flow chart of the development and validation phases as follows:

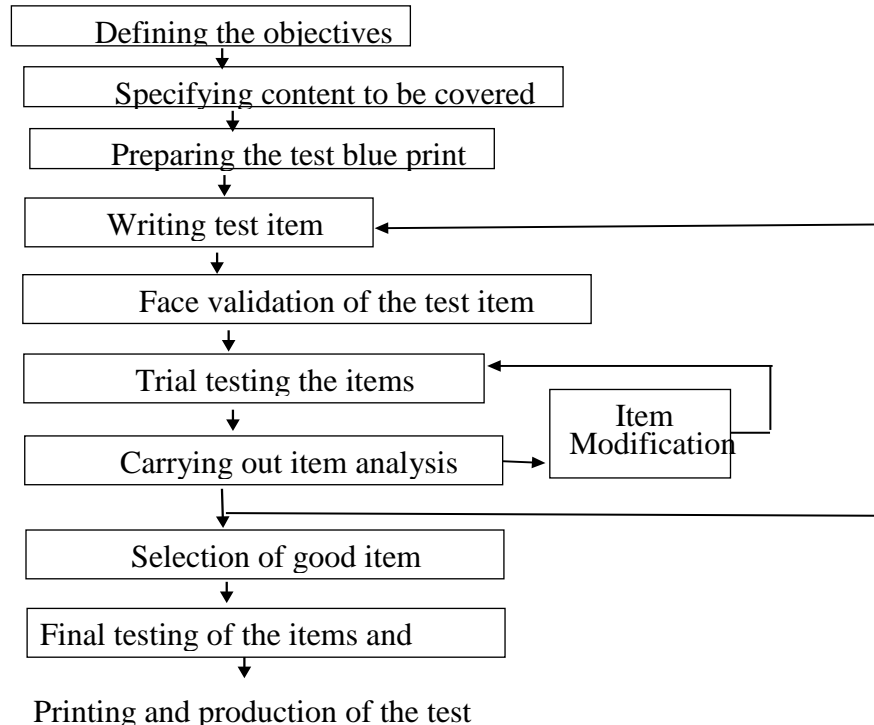


Fig.1: Flow chart of the development and validation phases

Longjohn, et al. (2021) noted that a test can be studied from different perspectives and the items in the test can be evaluated according to different theories. Two of such theories are the Classical Test Theory (CTT) and the Item Response Theory (IRT). These theories are among the major frameworks that are used in educational measurement to develop, evaluate, determine the reliability and validity of tests, as well as improve the quality of test items. These frameworks are based on different assumptions and use different statistical approaches. Classical test theory was originally the leading framework for developing and analysing standardized tests. Later, IRT was developed to compliment the role of CTT (Longjohn et al., 2023). However, the study will be anchored mainly on the use of CTT techniques. Classical test theory is based on the assumption that an examinee has an observed score and a true score. The observed score of a test-taker is usually seen as a combination of an estimate of the true scores of that test-taker, plus/minus some unobservable error. The true score reflects what the test-taker actually knows, but it is always contaminated by different sources of errors.

To estimate the range of true score, the

CTT applies reliability and standard error of measurement (SEm). The SEm is closely related to test reliability; that is, the higher the reliability, the smaller the SEm (If reliability is 1.00, then SEm = 0) and the higher the precision in the measures taken and scores obtained. The standard error of measurement (SEm) estimates how repeated measures of a person on the same instrument tend to be distributed around his or her "true" score. The SEm is also defined as the "standard deviation of measurement errors associated with test scores for a specified group of test takers (AERA, APA, & NCME 1985 in Glen 2023)." Because no measure can provide a precise reflection of the true score, the true score is always unknown. Statements concerning an examinee's actual score received on a test are couched in terms of a confidence interval — a range of scores that contains the examinees "true" score with a high probability.

Classical test theory utilizes measures of item characteristics, item difficulty and item discrimination, the values of which are dependent upon the distribution of examinee proficiency within a sample. Although the assumptions upon which classical test theory is based allow it to be

applied to an assortment of test construction situations, these same assumptions appear to create weaknesses in the classical test theory model. The CTT based statistical indices are easy to compute, manipulate and understand by lay persons, but they vary from sample to sample. While CTT has proven very useful in test development, the two statistics that form its cornerstones, item difficulty and item discrimination are both sample dependent.

The researchers reviewed some studies that applied test theories for construction and validation of achievement tests in Economics. Esonomu and Eleje (2017) developed and validated a diagnostic test to measure economics quantitative skill of secondary school students. The test consisted of 50 multiple choice test items which were constructed based on economics curriculum. Six research questions guided the study. Preliminary validation was done by two experienced teachers in secondary economics and two experts in test construction. The pilot testing was conducted to ensure grammatical checking and compatibility of the test items while test try-out on 517 students yielded the data for item analysis. Analysis was done through Bilog MG using Item Response Theory (IRT) three-parameter logistic model (3PL) to establish item difficulty, item discrimination, and the guessing values. The empirical reliability of the test was 0.87. The difficulty and discrimination level of the items indicated that 39 items be retained while 11 items be replaced or modified.

Eleje and Esonomu (2018) also constructed and validated an achievement test to measure quantitative skills in economics among secondary school students. The test is made up 20 multiple choice test items constructed based on quantitative economics sub-skills. Six research questions guided the study. Preliminary validation was done by two experienced teachers in secondary school economics and one expert in test construction. The pilot testing was conducted to ensure compatibility of the test items while test try-out on 515 students yielded the data for item analysis. Analysis was done with the aid of SPSS and Bilog MG using Item Response Theory (IRT)

three-parameter logistic model (3PL) to establish item difficulty, item discrimination, and the guessing value. The result revealed that eleven (11) items be retained, nine (9) items be revised and ten (10) items be rejected. The empirical reliability of the test was 0.86.

Eleje et al. (2017) developed an Economics achievement test (EAT) for assessing senior secondary two (SS2) achievements in economics. The study was guided by five research questions. Twenty (20) and 100 mid-senior secondary (SS2) economics students were used for the pilot testing and reliability check respectively. A sample of 250 students randomly drawn was used to subject the initial 80 objective test items for the test try-out that yielded the data for item analysis. A total of 50 items with difficulty indices ranged from 0.25 to 0.79 and discrimination indices of 0.20 to 0.58 were retained. Face and content validation of EAT was ensured by constructing items in line with the test blueprint, the use of subject experts in SS2 economics and two experts in test construction. The test reliability established through the Kuder-Richardson formula 20 gave a coefficient of 0.81. The test was found to be of good quality, valid and highly reliable.

Osadebe (2014) carried out a study to determine the construct validity and reliability of test in Economics for secondary school students. Two research questions were stated to guide the study. It was a 100 items multiple choice objective test with five options. A sample of 1000 students was randomly drawn for the study. Difficulty and discrimination indices were also determined. A difficulty index or p-value of 0.5 for each item was considered after applying the formula of correction for guessing. The discrimination index was established through point biserial statistics for each item with a correction coefficient of at least 0.3. The test has a reliability coefficient of 0.95 established through the use of Kuder-Richardson formula 20. Also, a SEM or error measurement of 4.6 was established.

All the works cited in the study were carried out in areas outside Rivers State while there was paucity of literature in this area of study in Rivers State. However, the researchers

observed that public schools in the state employ teacher-made tests for internal assessments of their students and these tests with unknown quality or psychometric properties are also used in preparation for external certificate examinations. This anomaly where unvalidated teacher made tests are used to prepare students for external examinations may be one of the causes of poor performance in external examinations. Asuru et al. (2023) observed that students are often afraid of external examination because of the supposed difficulty, and this cause them to explore illegal means of passing external examinations. Asuru et al. (2023) stated that among several factors that could lead to examination malpractice and failure is the fact that classroom teachers fail to prepare students with quality tests (with good psychometric properties) like those used by external examining bodies. Therefore, the study seeks to address the problem of lack of test construction skills amongst educationists with the application of innovative and easy approach to test construction using computer application (R software). This study investigated the application of R computer software for construction and validation of Economics achievement test in Etche LGA of Rivers State.

Research Questions

The following research questions guide this study:

What are the item parameters (difficulty and discrimination indices) of the Economic achievement test?

How many of the items of the EAT survived under CTT framework?

What is the reliability of the Economic achievement test (EAT)?

2. Methodology

Research Design

The design for the study is instrumentation research design. In this study, Economics Achievement Test was developed and validated using Classical test theory to measure the ability of SS1 students in Economics in Etche Local Government Area of Rivers State.

Population of the Study

The populations of the study consist of all

senior secondary school students offering Economics in 2020/2021 academic session in Etche LGA, Rivers State. The total number of Economics students in all the eighteen (18) public senior secondary schools for 2020/2021 academic session was 2000 students (Source: Etche L.G.A Zonal Education Authority, 2020).

Sample for the Study

A sample size of 400 was chosen for the study. This sample size is above the minimum sample size of 333 appropriate for a population statistically determined using the Taro Yamen Formula. The sample consists of 190 males and 210 females. A purposive sampling technique was used to select six (6) schools for the study while, a proportionate stratified random sampling technique was used to select 400 SS1 students from the sampled schools.

Instrument for Data Collection

The instrument for data collection is a self-structured multiple-choice test titled "Economics Achievement Test (EAT)" which consists of section (A and B). Section A dealt with the bio-data of the students, while section B captured questions on Economic Achievement Test with four multiple-choice options used to elicit information from the students. A pool of 150 items generated from 14 topics in the SS1 Economics curriculum was used for trial testing. The distribution of the test items was according to the content and the levels of cognitive domain of SS1 students as shown on the table of specification. The Economics Test Achievement (EAT) was administered to SSI Economics students during the third term of the 2020/2021 academic session in Etche Local Government Area of Rivers State. The test administration was done by one of the researchers, aided by two trained research assistants recruited by the researchers. At completion, the scripts were collected from the test takers.

Validation of the Instrument

The content validity was established through a test blueprint or table of specification. This is in agreement with Kpolovie (2010) in Asuru et al. (2023), who affirmed that validity of any

achievement test can best or most appropriately be ascertained or established through content validity using test blueprint. First, the researcher extracted the content of 14 topics from the Economics curriculum for SS1 as published by the Nigerian Educational Research and Development Council (NERDC) Also, the face and content

validities of the test were further established by subject specialists (using senior teachers in Economics). The experts were provided with the test form and test blue print for editing and vetting of the instrument. All suggestions and corrections made were effected before actual testing.

Table 1: A Test Blueprint for a 150-Item Economics Achievement Test

CONTENT	BEHAVIOURAL OBJECTIVES						Total %
	Remember 15%	Understand 25%	Apply 20%	Analyze 25%	Evaluate 10%	Create 5%	
Concept of economics	3	3	2	1	1	0	10 (6.7%)
Basic tools of economic analysis	4	3	2	1	1	1	11 (7.3%)
Demand and Supply	4	2	1	1	2	2	12 (8%)
Theory of Production	3	3	2	2	1	1	12 (8%)
Basic Economic problem of society	2	1	2	1	0	0	6 (4%)
Firms and Industry	3	2	1	1	1	0	8 (5.3%)
Population	4	2	1	2	2	2	13 (8.7%)
Nature of the Nigerian Economy	2	1	2	2	1	1	9 (6%)
Agriculture	2	2	2	2	1	0	9 (6%)
Mining	3	3	2	2	0	0	10 (6.7%)
Financial institution	2	2	2	1	2	2	11 (7.3%)
Money	3	3	3	2	2	2	15 (10%)
Channels of distribution	4	3	2	2	2	0	13 (8.7%)
Instrument of business finance	3	2	2	2	1	1	11 (7.3%)
Total (%)	41 (27.3%)	32 (21.3%)	26 (17.3%)	22 (14.7%)	17 (11.4%)	12 (8%)	150 (100%)

Reliability of the Instrument

The reliability of the test was established through Cronbach Alpha and Kuder Richardson-20 statistics using R-software and SPSS version 27 for initial testing. Both Cronbach Alpha and Kuder

Richardson-20 produced reliability coefficients of 0.96. For the final test a coefficient of 0.92 was obtained for both Cronbach alpha and Kuder-Richardson-20 reliability methods while split-half reliability produced a coefficient of 0.94 indicating

that the selected 63 items were reliable.

Item Analysis

A good item which fulfils the criteria for an acceptable discrimination index (0.2 and above) was selected while those that have discrimination indices of less than 0.2 were dropped. Items that have difficult indices of 0.3 to 0.7 were selected while low difficulty indices less than and above were dropped. The selected items formed the final test.

Method of Data Analysis

For method of data analysis, SPSS was used for data coding and cleaning while, several R packages were employed for item analysis using CTT techniques. The following R commands were utilized for CTT analysis:

```
#Item difficulty and Item discrimination
library(psychometric)
EATpara <-
item.exam(EATdata,y=NULL,discrim=TRUE)
```

```
# Split-Half reliability
library(hemp)
#For half test:
split_half(EATdata2, type = "alternate")
#For full test:
split_half(EATdata2, type = "alternate", sb = TRUE)
#Cronbach-alpha, and KR20 reliability
library(validateR)
kr20(EATdata2)
alpha(EATdata2)
```

3. Results

Research Question 1: What are the item parameters (difficulty and discrimination indices) of the Economic achievement test?

To answer this question, the test items were subjected to psychometric analysis using R programming language. Here, the package ‘psychometric’ and ‘item.exam’ function were utilized.

Table 4: Item parameter estimates of Economic achievement test

Items	Difficulty Index	Remark	Discrimination Index	Remark
Item 1	0.75	Easy	0.32	Good
Item 2	0.79	Easy	0.29	Good
Item 3	0.55	Moderate	0.32	Good
Item 4	0.72	Easy	0.44	Good
Item 5	0.77	Easy	0.35	Good
Item 6	0.73	Easy	0.29	Good
Item 7	0.77	Easy	0.25	Good
Item 8	0.75	Easy	0.45	Good
Item 9	0.56	Moderate	0.41	Good
Item 10	0.70	Easy	0.28	Good
Item 11	0.75	Easy	0.33	Good
Item 12	0.46	Moderate	0.28	Good
Item 13	0.75	Easy	0.39	Good
Item 14	0.74	Easy	0.38	Good
Item 15	0.40	Moderate	0.29	Good
Item 16	0.75	Easy	0.42	Good
Item 17	0.75	Easy	0.35	Good
Item 18	0.60	Moderate	0.38	Good
Item 19	0.72	Easy	0.22	Good
Item 20	0.73	Easy	0.37	Good
Item 21	0.77	Easy	0.28	Good
Item 22	0.60	Moderate	0.34	Good
Item 23	0.74	Easy	0.18	Bad
Item 24	0.70	Easy	0.32	Good

Item 25	0.75	Easy	0.32	Good
Item 26	0.60	Moderate	0.49	Good
Item 27	0.74	Easy	0.31	Good
Item 28	0.71	Easy	0.43	Good
Item 29	0.71	Easy	0.36	Good
Item 30	0.45	Moderate	0.32	Good
Item 31	0.77	Easy	0.35	Good
Item 32	0.71	Easy	0.44	Good
Item 33	0.60	Moderate	0.36	Good
Item 34	0.75	Easy	0.37	Good
Item 35	0.72	Easy	0.35	Good
Item 36	0.60	Moderate	0.40	Good
Item 37	0.73	Easy	0.39	Good
Item 38	0.72	Easy	0.40	Good
Item 39	0.73	Easy	0.43	Good
Item 40	0.61	Moderate	0.30	Good
Item 41	0.74	Easy	0.45	Good
Item 42	0.72	Easy	0.51	Good
Item 43	0.59	Moderate	0.51	Good
Item 44	0.70	Easy	0.31	Good
Item 45	0.60	Moderate	0.41	Good
Item 46	0.75	Easy	0.45	Good
Item 47	0.71	Easy	0.45	Good
Item 48	0.72	Easy	0.36	Good
Item 49	0.70	Easy	0.42	Good
Item 50	0.74	Easy	0.42	Good
Item 51	0.55	Moderate	0.46	Good
Item 52	0.76	Easy	0.34	Good
Item 53	0.70	Easy	0.45	Good
Item 54	0.70	Easy	0.31	Good
Item 55	0.33	Difficult	0.41	Good
Item 56	0.72	Easy	0.43	Good
Item 57	0.72	Easy	0.40	Good
Item 58	0.35	Difficult	0.39	Good
Item 59	0.42	Moderate	0.32	Good
Item 60	0.75	Easy	0.37	Good
Item 61	0.72	Easy	0.45	Good
Item 62	0.73	Easy	0.34	Good
Item 63	0.70	Easy	0.39	Good
Item 64	0.43	Moderate	0.37	Good
Item 65	0.70	Easy	0.37	Good
Item 66	0.72	Easy	0.37	Good
Item 67	0.31	Difficult	0.36	Good
Item 68	0.73	Easy	0.30	Good
Item 69	0.60	Moderate	0.38	Good
Item 70	0.73	Easy	0.40	Good
Item 71	0.72	Easy	0.39	Good

Item 72	0.45	Moderate	0.37	Good
Item 73	0.73	Easy	0.46	Good
Item 74	0.78	Easy	0.34	Good
Item 75	0.70	Easy	0.40	Good
Item 76	0.53	Moderate	0.30	Good
Item 77	0.72	Easy	0.47	Good
Item 78	0.76	Easy	0.31	Good
Item 79	0.75	Easy	0.38	Good
Item 80	0.40	Moderate	0.42	Good
Item 81	0.72	Easy	0.38	Good
Item 82	0.75	Easy	0.34	Good
Item 83	0.35	Difficult	0.40	Good
Item 85	0.75	Easy	0.40	Good
Item 86	0.77	Easy	0.40	Good
Item 87	0.60	Moderate	0.47	Good
Item 88	0.76	Easy	0.28	Good
Item 89	0.72	Easy	0.45	Good
Item 90	0.78	Easy	0.36	Good
Item 91	0.51	Moderate	0.45	Good
Item 92	0.75	Easy	0.36	Good
Item 93	0.52	Moderate	0.43	Good
Item 94	0.77	Easy	0.34	Good
Item 95	0.48	Moderate	0.38	Good
Item 96	0.76	Easy	0.31	Good
Item 97	0.54	Moderate	0.39	Good
Item 98	0.73	Easy	0.42	Good
Item 99	0.72	Easy	0.44	Good
Item 100	0.34	Difficult	0.35	Good
Item 101	0.73	Easy	0.35	Good
Item 102	0.74	Easy	0.43	Good
Item 103	0.74	Easy	0.34	Good
Item 104	0.53	Moderate	0.47	Good
Item 105	0.74	Easy	0.33	Good
Item 106	0.71	Easy	0.44	Good
Item 107	0.60	Moderate	0.38	Good
Item 108	0.73	Easy	0.36	Good
Item 109	0.44	Moderate	0.34	Good
Item 110	0.73	Easy	0.38	Good
Item 111	0.30	Difficult	0.45	Good
Item 112	0.74	Easy	0.38	Good
Item 113	0.40	Moderate	0.44	Good
Item 114	0.76	Easy	0.31	Good
Item 115	0.42	Moderate	0.39	Good
Item 116	0.75	Easy	0.33	Good
Item 117	0.61	Moderate	0.55	Good
Item 118	0.71	Easy	0.45	Good
Item 119	0.56	Moderate	0.31	Good

Item 120	0.74	Easy	0.47	Good
Item 121	0.59	Moderate	0.41	Good
Item 122	0.48	Moderate	0.47	Good
Item 123	0.74	Easy	0.33	Good
Item 124	0.59	Moderate	0.42	Good
Item 125	0.74	Easy	0.34	Good
Item 126	0.43	Moderate	0.41	Good
Item 127	0.73	Easy	0.42	Good
Item 128	0.35	Difficult	0.33	Good
Item 129	0.44	Moderate	0.37	Good
Item 130	0.68	Easy	0.41	Good
Item 131	0.73	Easy	0.27	Good
Item 132	0.73	Easy	0.33	Good
Item 133	0.44	Moderate	0.26	Good
Item 134	0.74	Easy	0.41	Good
Item 135	0.70	Easy	0.40	Good
Item 136	0.70	Easy	0.36	Good
Item 137	0.60	Moderate	0.46	Good
Item 138	0.75	Easy	0.38	Good
Item 139	0.75	Easy	0.37	Good
Item 140	0.60	Moderate	0.31	Good
Item 141	0.73	Easy	0.42	Good
Item 142	0.75	Easy	0.34	Good
Item 143	0.53	Moderate	0.46	Good
Item 144	0.71	Easy	0.39	Good
Item 145	0.74	Easy	0.35	Good
Item 146	0.55	Moderate	0.33	Good
Item 147	0.72	Easy	0.38	Good
Item 148	0.76	Easy	0.38	Good
Item 149	0.34	Difficult	0.44	Good
Item 150	0.51	Moderate	0.36	Good

The result from Table 2 presents the CTT item parameters of the 150 test items used for the study, and the parameters include discrimination and difficulty indices. The item difficulty indices ranged between 0.30 (item 111) to 0.78 (item 2). The result showed that 98 (65.7%) items were easy, 43 items (28.9%) had moderate difficulty and 8 items (5.4%) are difficult. On the whole, the items were relatively easy. Similarly, the item discrimination indices ranged between 0.18 (item

23) to 0.55 (item 117). The result showed that one (1) item possessed poor discrimination, 12 items were between the range of 0.2 to 0.29 while 137 items were between the range of 0.30 and above. On the whole, the items possessed excellent discrimination.

Research Question 2: How many items in the Economic achievement test (EAT) survived under CTT framework?

Table 3: Items deleted and accepted under CTT Framework

Item parameter	Number Deleted	Items Accepted
Item Difficulty	86	3, 9, 10, 12, 15, 18, 22, 24, 26, 30, 33, 36, 40, 43, 44, 45, 49, 51, 53, 54, 55, 58, 59, 63, 64, 65, 67, 69, 72, 75, 76, 80, 83, 87, 91, 93, 95, 97, 100, 104, 107, 109, 111, 113, 115, 117, 119, 121, 122, 124,

126, 128, 129, 130, 133, 135, 136, 137, 140, 143, 146, 149, 150

Item Discrimination	1	
Total	86 (57.7%)	63 (42.3%)

The Result from Table 5 presents the CTT item parameters of the 150 test items, calibrated using the standard criteria. The criteria for item analysis under the CTT framework for item selection was based on the discriminative index ($r_{pbs} \geq 0.20$) and difficulty index ($0.30 \leq p \leq 0.70$) as recommended by Metibemu (2016) and Kaplan and Saccuzo (2009) respectively. This range tends to maximize information about the differences among testees and it also accommodates test items with moderate difficulty which produces the best quality. Therefore, 86 out of the 150 items were marked as bad items, one was omitted, while, 63 items were selected as good items to

make the final test form. This implies that the CTT approach retained 42.3% of the Economic Achievement test items, showing good psychometric properties and fit for measuring students' Economic Achievement for SS 1 in public secondary schools in Etche Local Government Area, Rivers state.

Research Question 3: What is the reliability of the Economic achievement test (EAT)?

The reliability of the test was determined using three reliability methods, namely; Cronbach alpha method, Split-half method and Kuder-Richardson 20. The 'psychometric', 'hemp' and 'ValidateR' packages were utilized for reliability testing.

Table 3: Reliability coefficients using Cronbach alpha, split-half and Kuder-Richardson

Reliability Method	Coefficient	SEm
Cronbach alpha	0.922	3.292
Split-half	0.939	2.911
Kuder-Richardson-20	0.922	3.292

SEM: $SD \cdot \sqrt{1-r}$, where scale SD = 11.788, Mean score 44.93, r = reliability coefficients

After computation, the following coefficients were obtained as the measures of internal consistency of the test. Since the coefficient of the EAT is high enough, it can therefore be affirmed that the test was reliable, since its value is above 0.7 which is the benchmark for a reliable instrument. Also, CTT assumes that measurement error or standard error of measurement (SEM) is the same for every examinee and it is used to determine the precision of the observed score to the true score. The SEM of the test at a reliability coefficient (r) of 0.922 is 3.29 and for r of 0.94 is 2.911.

4. Discussion

The first result of the study showed the item parameters of the 150 test items and the findings indicated that on the whole, two third of the items were easy while about a third of the items were moderately difficult and leaving only five percent of difficult items. The finding also revealed that over 90% of the items possessed

excellent discrimination. Indeed, the item parameters place a quality tag on newly developed items and this piece of information provides the test user with the characteristics of each item which is a precondition for selection based on the purpose of the test.

Similarly, the second result of the reported that under the CTT framework, 63 items out of the initial 150 EAT items were selected as good items (most of which are moderately difficult) to make the final test form. The criteria for item selection under the CTT framework corroborated by Metibemu (2016) and Kaplan and Saccuzo (2009). Although, the number of items retained is low as compared to the studies of Eleje and Esomonu (2017) who retained 78% of the test items and Eleje et al. (2017) who retained 63% of the test items. However, the finding still affirms the fact that the best items are within the moderate difficulty range. The 42.3% item retention further supports the rule of thumb that

test experts are required to develop thrice the number of items intended for a test. This approach makes more items available for use and forestalls the situation of repeating the process because of insufficient good items.

The third result of the study reported that the reliability of the final test is 0.92 and 0.94 for the various reliability methods with a SEM of 3.29 and 2.911 for the respective reliability coefficients.

This result implies that the test possesses a high precision in estimating the true score of test takers on repeated attempts. Suppose an individual scores 80 on the EAT with a SEM of 3.29. Applying the recommended confidence intervals (CI) formulas for the true score (Glen, 2023; Fur, 2018): $68\%CI = \text{Score} \pm \text{SEM}$, $95\%CI = \text{Score} \pm (1.96 * \text{SEM})$ and $99\%CI = \text{Score} \pm (2.58 * \text{SEM})$. The true score could be calculated at 95% confidence interval as: $(80 - 2 * 3.29, 80 + 2 * 3.29) = (73, 87)$. Therefore, the “true score” of the examinees who scored 80 on the EAT varies between 74 (lower bound) and 86 (upper bound) at 95% CI. The result is similar to the study of Osadebe (2014) who reported a reliability of 0.95 and a SEM of 4.6. The higher value of SEM is due to the length of the items (100) which has impacted the Standard Deviation of the observed score. However, the few other studies on the subject reported reliability coefficients below 0.90 and they also failed to establish a SEM for their test (Eleje & Esomonu, 2017; Eleje et al., 2017).

5. Conclusion

In all, the study has advanced an easy and cost-effective way of developing and validating achievement test using a few editable commands on the R software- a free and open source computer software capable of any conceivable statistical analyses. The study indicated that 42.3% of the items were retained to make the final test with difficulty indices ranging from 0.30 to 0.70 and discrimination indices that ranged from 0.22 to 0.55. The study affirmed a high reliability coefficient of above 0.90 with an established SEM for the test. The findings show that the EAT is both valid and reliable and thus, is fit for measuring students' academic achievement in Economics for

SS 1. Therefore, the EAT will be useful for teaching and learning, research purposes and cognitive testing.

6. Recommendations

Based on the findings of the study, the following recommendations were made:

The EAT can be used as a valid and reliable classroom test for measuring students' proficiency in SS1 Economic in public secondary schools in Etche LGA of Rivers State. Educationists and researchers should explore the use of innovative approaches such as R software for test development and validation. The EAT possess good psychometric properties and is recommended for standardization using a larger sample size.

References

- Asuru, V. A. (2015). *Measurement and evaluation in education and psychology* (2nd ed.). Pearl Publishers International Ltd.
- Asuru, V. A. & Njigwum, A. S. & Nna-kue, L. N. (2023). Development and validation of Basic Science Achievement test using classical test approach: equipping teachers for global competitiveness. A paper presented at the 2022 Committee of Deans of Education in Nigerian Universities (CODENU) National Conference held at the Faculty of Education, Ignatius Ajuru University of Education, Port Harcourt, from 11th – 14th January, 2023.
- Kaplan R. M. & Saccuzzo, D. P. (2009). *Psychological testing principles, applications and issues* (7th ed.). Thomson Wadsworth.
- Bichi A. A (2016). Classical Test Theory: An Introduction to Linear Modeling Approach to Test and Item Analysis. *International Journal of Social Sciences* 2(9), 27-33. <https://doi.org/10.26643/ijss.v2i9.6690>
- Esomonu, N. P. M. & Eleje, L. I. (2017). Diagnostic Quantitative Economics Skill Test for Secondary Schools: Development and Validation Using Item Response Theory. *Journal of Education and Practice*, 8(22),

- 110-125.
- Eleje, L. I., Abanobi, C. C., & Obasi, E. (2017). Development and Validation of Economics Achievement Test for Secondary Schools. *Asian Journal of Education and Training*, 3(1), 6–17. <https://doi.org/10.20448/journal.522/2017.3.1/522.1.6.17>
- Eleje, L.I. & Esomonu, N. P. M. (2018). Test of Achievement in Quantitative Economics for Secondary Schools: Construction and validation using item response theory. *Asian Journal of Education and Training*, 4(1), 18-28.
- Frey, F. (2017). Test theory, classical test theory. In J. P. Matthes, R. Potter, & C. S. Davis (Eds.). (in prep.). *International Encyclopedia of Communication Research Methods*. Wiley Blackwell. <https://dx.doi.org/10.1002/9781118901731.iecrm0>
- Furr, R. M. (2018). *Psychometrics: An introduction* (3rd ed.). SAGE.
- Hayes, A. (2020). *Guide to Economics*. <https://www.investopedia.com/terms/e/economics.asp#:~:text=Economics%20is%20a%20social%20science,about%20how%20to%20allocate%20resources>.
- Iowa State University (2001). Revised Bloom's taxonomy <https://www.celt.iastate.edu/teaching/effective-teaching-practices/revised-blooms-taxonomy/>
- Jhingan, M. L. (2010). *Microeconomic theory*. Vrinda Publications (P) Ltd.
- Longjohn, I. T., Njigwum, A. S. & Ibeh, C. P. (2021). Validation of Basic Science achievement test using R software. *International Journal of Innovative Research & Development*, 10(9), 77-87. <http://www.doi10.24940/ijird/2021/v10/i9/SEP21025>.
- Longjohn, I. T., Njigwum, A. S., Osila K., Anunaobi, J. C. & Amadi, C. P. (2023). Construction and validation of Mathematics aptitude test for Junior Secondary school students in Rivers State using R programming technology. A paper presented at the 8th Annual National Conference of Faculty of Education, 8th – 12th May 2023, Ignatius Ajuru University of Education, Port Harcourt, Nigeria.
- Metibemu, M. A. (2016). Comparison of classical test theory and item response theory in the development and scoring of senior secondary school physics tests in Ondo State. Unpublished Ph.D thesis. Institute of Education. University of Ibadan.
- Mehjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., & Rizvi, M. (2017). Difficulty index, discrimination index and distractor efficiency in multiple choice questions. *Annals of Pakistan Institute of Medical Sciences*, 4, 310–315. <https://www.researchgate.net/profile/>
- Obilor, E, I & Akpan, U. T. (2020). Construction and Validation of Basic Science Achievement Test for Junior Secondary Three (JS3) Students in Public Secondary Schools in Akwa Ibom State. *International Journal of Innovative Education Research*, 8(4), 124-135.
- Osadebe, P.U. (2014). Construction of Economics Achievement Test for Assessment of Students. *World Journal of Education*, 4(2), 58-64. <https://doi.org/10.5430/wje.v4n2p58>
- Salkind, N. J. (2010). Guessing parameter. In *Encyclopedia of Research Design*. <https://dx.doi.org/10.4135/9781412961288.n172>
- Statology (2021). Standard error of measurement: Definition & example. <https://www.statology.org/standard-error-of-measurement/>