



DEVELOPMENT AND VALIDATION OF INSTRUMENT FOR ASSESSING STATISTICAL LITERACY IN OF POST GRADUATE STUDENTS

Uju Patricia Okonkwo
and

Prof. Romy Okoye
Department of Educational Foundations
Nnamdi Azikiwe University Awka
Email: ouju3577@gmail.com

ABSTRACT

This study aimed at the development and validation of statistical literacy scale for postgraduate (PG) students in faculties of education of public universities in South East, Nigeria. One hundred and twenty (120) items were generated, guided by a table of specifications. These constituted the first draft of the instrument. This draft was given to two experts in Educational Measurement and Evaluation and an expert in statistics, all from Nnamdi Azikiwe University Awka. They were requested to critically examine the items of the draft instrument in terms of coverage and quality of items. After this initial validation, 110 items survived. These items made up the Second draft of the instrument. Six research questions were raised to guide the study. Instrumentation design was adopted. The study was carried out in South East Nigeria, with the population made up of all 2969 postgraduate students offered admission in 2020/2021 session in the public universities in South East, Nigeria. A sample of 1,000 students, was used for the study, made up of 400 that were used for trial- testing and 600 that were used for ascertainment of reliability and validity. Each of the two samples was obtained, using multi-stage sampling technique. After the initial validation, the 110 items that survived were administered on 400 students. Scores obtained from the students were subjected to item analysis. Consequently, the difficulty, discrimination and distracter indices for each item were computed. Forty items, which were found to be acceptable in terms of each of the three indices constituted the final instrument. This final instrument was then administered on the second sample of 600 postgraduate students. scores obtained from this second testing were used to ascertain evidences of reliability and validity for the instrument. The result showed that a reliability coefficient of 0.87 was obtained when Kuder –Richardson 20 method was used while a coefficient of 0.81 was obtained when split-half method was used. Scores obtained with the instrument converged with those obtained with Statistical Reasoning Test (SRT) and diverged with those obtained using Test Anxiety Questionnaire {TAQ}. Based on the above, it was concluded that the new instrument (Postgraduate Students' Statistical Literacy Scale) was reliable and valid.

Keywords: Statistical literacy, instrument development, validation, reliability, statistical assessment, measurement.



Introduction

Testing has become one of the most critical parameters by which a society judges the product of her educational system. Testing has been an integral part of the school system, to the extent that even the habitual absentees normally turn up to school and present themselves for testing on examination days. The reason for testing is to identify the latent ability of examinee. Testing has been accepted in most modern societies as the most objective method of decision making in schools, industries, and government agencies. A test is a set of questions or statement that have been structured and graduated to elicit and measure responses about an attribute (Kunuba and Anigbo, 2023). Tests are used for gathering valuable data on which educational decisions are based (Carik and Ari, 2022). These decisions are valid and centred on the learner so that such decision should be carefully based on authentic and accurate data. Authentic and correct data collected from testing helps the lecturer to ascertain the extent to which the instructional objectives have been achieved, provides feedback on the efficiency of his teaching methods and materials. On the part of the learner, it increases their motivations to learn and helps to determine one's area of strengths and weaknesses. Teacher-made achievement test and standardized achievement test are frequently used in achieving these ends.

Achievement tests are the tests designed to evaluate how much learning that has taken place due to exposure to a given instruction of learning. Karl (2022) defined achievement test as an evaluation of past or present learning by measuring progress, which students have made as a result of instruction or training. A teacher-made achievement test is that test which is designed by classroom teachers (Fajobi, 2019). They had observed that teacher-made achievement tests lack uniformity in administration and scoring. Therefore, there is a need for well-constructed, standardized test to help reduce the deficiencies of the teacher-made test in statistics. Moreover, the teacher-made tests that are always in use for evaluation of learning outcomes lack test norms and test manuals. According to Obilor (2019), because they do not have norms and test manuals, informal achievement test scores are "limited in comparison and interpretation. Hence, the need for standardized test.

A standardized achievement test is that test, which is designed by a test expert and has been subjected to certain statistical analysis to ascertain the psychometric properties of the test. Test norms enable us to understand how students perform in a given test. Test standardization is a crucial stage in test development. According to Fajobi (2019), test standardization is a relatively difficult task which requires experience, concentration, a thorough knowledge of the subject matter, and a good deal of patience. Standardized tests are usually the product of experienced teachers, curriculum specialists and test experts. Standardized test covers a wider area of skills and knowledge than the teacher made test. A standardized test is usually based on the course content, and objectives common to many schools across a country, state or a designated geographical area and the preparation, therefore, requires more skills. Hence, the highly skilled nature of the personnel required for its construction and the time available is usually long. As observed by Carik and Ari (2022), most classroom teachers cannot or will not be able to undergo the rigours involved



in constructing and validating achievement test. Obilor (2019) stated that this is because classroom teachers are not skilled in the act of construction and standardization of tests. The procedures adopted in their test development are haphazard, narrow in scope and covers few topics taught within a specified period (Carik and Ari, 2022). They went on to say that such tests are designed to suit the convenience of the locality where they are produced, and staff member of the school usually administer such test. achievement test. In most cases, such classroom test lacks validity and reliability that are essential characteristics of a good test. Planning the test, preparing the test, experimental try out and revision, administering the standardized edition and provision of norms to guide interpretation of test scores.

A standardized test is usually tested, and the responses are subjected to item analysis (Haw and Sharif, 2020). Haw and Sharif went further to say that such test could only be prepared by subject teachers who work with specialists in test development. According to them, this will make it possible to determine the psychometric qualities of the items such as item difficulty, discriminative and distracter capacities as well as the validity and reliability of the test. It should be noted that the teaching of advance statistics is not complete until every aspect of the concepts taught has been adequately evaluated.

One of the roles of a university is to prepare graduates with skills for their future, particularly in the workplace. An important aspect of such learning skill, is statistical literacy. Man, now lives in a data-drenched society, and it is the responsibility of educators and statisticians to promote, enhance and develop statistical literacy within the immediate community. Statistical literacy, as a concept, includes the ability to read and interpret statistical data in one's daily activities (Ukar, 2012). Examples of statistical data include graphs, tables, statements, statistical surveys and studies. It is also considered a facilitator of students' navigation of the depths of the data-driven world.

Therefore, students are expected to have the ability to not only read statistical results and data summaries, but to understand and critique them (Watson, 2006). Statistical literacy is often touted as an expected outcome of schooling and as a necessary component of adults' numeracy and literacy (Garfield & Ben-Zvi, 2007). Budgett and Pfannkuch (2007) explain it to be the ability of an individual to view the world and its phenomena from a statistical perspective, after being able to statistically construct and synthesize sound reports from analyzing original sources. It thus suffices to infer that statistical literacy refers to a student's capacity to interpret measured phenomena in logical, numerical, and analytical forms. In Nigeria, the higher-level aspect of statistical literacy, which include: critical thinking and statistical reasoning have appeared in some of the policies and practices in the Universities (Okebukola, 2017). Critical thinking and statistical reasoning are among the major competencies a post graduate student is expected to have.

The post graduate students in the universities are usually prepared to acquire the attributes and skills a student is expected to possess after their 3-year/4-year PG programme. The list of specific



attributes includes knowledge of and an ability to apply scientific principles and concepts; solving problems through the application of scientific knowledge and methods; demonstration of analytical abilities, and the development of skills for lifelong learning. Others are ability to understand, evaluate, access and critically review new information; the ability to think independently; ability to demonstrate in-depth knowledge of and skill in his or her major subject; and ability to effectively access and use information relevant to the subject. The rest of the attributes are the demonstration of numeracy skills, and ability to work collaboratively on tasks, among others.

Statistical literacy measuring tools such as Schield's (2008) Statistical Literacy Skills Survey, Watson's (2003) framework for statistical literacy and that of Gal (2002) have been in use in practice and in research (Keosayian & Johnson, 2014) to determine the statistical literacy of university students. These tools were developed for use with the ultimate purpose of making curricula adjustments to enhance statistical literacy and reasoning. The problem with a good number of these measurement tools is that most of them are meant for undergraduate students. Statistical literacy is of great importance to students offering Advanced Educational Statistics in higher institution. It has been demonstrated that students with good knowledge of statistical literacy will do well in Advance Educational Statistics. In other words, there is strong relationship between statistical literacy and students' academic achievement in advanced educational statistics. Karl (2022) also stated that a strong correlation between two similar constructs is evidence of convergent validity while a weak correlation between two different constructs is evidence of divergent validity. Convergent validity refers to strength of the relationship between the scores that are obtained from the two different measurement constructs you are interested in. In order to establish convergent validity, the strength of the relationship between the scores from the two different measurement procedures, is assessed. This is usually achieved by calculating a correlation between the scores from two constructs. A strong relationship, helps to demonstrate convergent validity. Divergent validity also helps to establish construct validity by demonstrating that the construct you are interested in is different from other constructs that might be present in your study. The extent to which divergent validity has been demonstrated is the strength of the relationship between the scores that are obtained from the two different constructs. In divergent validity, we are interested in the extent to which the scores diverge. That is little or no relationship should be found between the scores between the two different constructs

Considering the relevance of Advanced Educational statistics as a are subject in the curriculum of the university education, and the fact that most of the existing instruments for measuring statistical literacy, known to the researchers are meant for undergraduate students, there is need for a valid instrument for measuring statistical literacy of university postgraduate students. The primary concern of this study therefore is to develop and validate a statistical literacy scale for university postgraduate students



Purpose of the Study

The main purpose of this study was to develop and validate an instrument for assessing statistical literacy of post graduate student in faculty of Education in South East Nigeria.

Specifically, the study sought to:

1. identify the major dimensions of statistical literacy relevant to [target population];
2. develop items for measuring the identified dimensions;
3. establish evidence of the content validity of the instrument;
4. determine the empirical characteristics of the items;
5. determine the construct validity of the instrument; and
6. establish the reliability of the final instrument.

Research Questions

To address the above problem the following research question were posed

1. `What are the difficulty indices of PGSSLS?
2. What are the discrimination indices of PGSSLS?
3. What are the distracter indices of PGSSLS?
4. What are the items that made up the final instrument?
5. What are the evidences of reliability of the instrument?
6. What are the evidences of validity of the instrument?

Literature Review

Concept of Statistical Literacy

Statistical literacy involves the knowledge and skills required to understand and make appropriate use of statistical information. It involves more than computation because statistical information is usually presented within a context.

A statistically literate individual should be able to read tables and graphs, understand percentages and averages, interpret statistical statements, recognize misleading presentations and communicate conclusions based on statistical evidence.

Components of Statistical Literacy

For the purpose of this study, statistical literacy may be represented through the following dimensions:

- understanding basic statistical concepts;
- interpreting tables and graphs;
- interpreting measures of central tendency;
- understanding percentages, ratios and proportions;
- interpreting variability; evaluating statistical claims;
- drawing conclusions from statistical information; and
- communicating statistical information.

The final dimensions should, however, be determined from the literature, curriculum and expert review relevant to the target population.



Instrument Development Procedure

The first step in constructing the instrument was to define the construct being measured and identify the various levels of educational objectives as stipulated by Bloom. Thereafter, a table of specifications or test blueprint was constructed based on the course contents of Advance Educational Statistics taught to postgraduate students in faculties of education in public universities in South East Geopolitical Zone. The course content included review of basic statistical concept, descriptive statistics, inferential and non-parametric statistics, t-test statistics, ANOVA, ANCOVA, correlation and regression analysis, chi-square and statistical techniques post-hoc analysis.

Items measuring the statistical literacy of postgraduate students of faculty of education were generated from various sources, such as education statistics selection tests, standard achievement tests, textbooks, and from day-to-day experiences as a researcher. The first draft of the instrument consisted of one hundred and twenty multiple choice test items with four options-one key (correct answer) and three distracters.

The first draft instrument which was made of 120 items was presented to two experts in Measurement and Evaluation, from Nnamdi Azikwe University, Awka and one statistics expert from the Department of Statistics in the same University. The experts were requested to scrutinize the instrument in terms of language, level of testees, clarity of language and relevance; others include appropriateness of stems, options and spacing of items directions for administration and scoring among others. The experts were also requested to indicate if the items were relevant to the construct being measured. Their suggestions and corrections were effected on the first draft instrument. After this stage of validation, the number of items were reduced from 120 to 110 to give rise to a second draft.

Thereafter, the second draft of the instrument was subjected-testing to the trial on a sample of 400 PG students obtained from the study area through the help of Advanced Educational Statistics Lecturers. The Advanced Educational Statistics Lecturers that were involved in the administration of the instrument were guided by the researcher to ensure uniformity in the administration of the instrument.

The responses obtained from the students were used to do the item analysis. Consequently, item difficulty index, item discrimination index and distractor indices were obtained for each of the 110 items. In order to obtain these indices, the scripts of the respondents were arranged from that of the highest scorer to that of the lowest scorer. There after the scripts of the 27% of the high scorers and those of the 27% of the bottom scorers were identified. The responses of the top 27% and bottom 27% of the scorers were used to compute the indices (108 scripts for upper group and 108 scripts in lower group).

After the item analysis of the 110 items, 40 items survived the analysis. The 40 items that had the desired qualities were assembled. To constitute the final instrument. The final instrument along



with two other instruments statistical reasoning test and test Anxiety Questionnaire, were administered on another sample of 600 postgraduate education students, obtained from universities in the study area, outside the ones that were used for trial –testing. Scores obtained from those instruments were used to ascertain evidences of reliability and validity of the instrument.

Method

Research Design

The design of this study was instrumentation research design. Eze (2003) defined instrumentation research as one in which the main thrust of the study is geared entirely towards the development and standardization of an instrument whose different psychometric properties have been empirically determined. Similarly, Chime (2012) described instrumentation research design as a study which aims at investigating and introducing new or modified procedures, technologies or instrument for educational practice. The design is appropriate for the study because it is on the development and validation of instrument for assessing statistical literacy.

Area of the Study

The study was conducted in the South East geopolitical zone of Nigeria. South-East bordered in the north-by-North Central Geopolitical Zone, in the West and South by South South Geopolitical Zone and in the East by Cameroon Republic. The South Eastern Nigeria has five states which are: Abia, Anambra, Ebonyi, Enugu and Imo states. These states have common boundaries with each other. The people of the South East, Nigeria, are predominantly of Igbo speaking ethnic group. They are mostly known for their commercial activities.

The Zone also adds a rich agricultural importance to the economy of Nigeria with crops like yam, maize, potatoes, rice, cashew, plantain and cassava extensively cultivated. The region is also rich in natural resources, including crude oil, natural gas, lead, zinc, iroko, mahogany, obeche, rubber trees and oil palms. The region is well known for higher human capital because of presence of several public universities: Abia State University, Utulu. Chukwuemeka Odumegwu Ojukwu University Igbariam, Nnamdi Azikiwe University, Awka, Ebonyi State University, Abakaliki, Enugu State University of Science and Technology Enugu, University of Nigeria Nsukka, Imo State University Owerri, Mcheal Okpara University, Umudike, Alex Ekwueme Federal University Ndufu-Alike and Federal University of Science and Technology, Owerri.

Population of the Study

The population of the study comprised 2969 postgraduate students (made up of 1197 PhD and 1772 masters students offered admission in 2020/2021 academic session (Source: Exams and Records of PG Schools of public universities in South East, Nigeria) in faculties of education of public universities in South East Geopolitical Zone, these were Abia State University, Chukwuemeka Odumegwu Ojukwu University Igbariam, Nnamdi Azikiwe University, Awka, Ebonyi State University, Abakaliki, Enugu State University of Science and Technology Enugu,



University of Nigeria Nsukka, Imo State University and Mcheal Okpara University, Umudike, Alex Ekwueme Federal University Ndufu-Alike and Federal University of Science and Technology, Owerri.

Sample and Sampling Technique

Two different samples were used in this study. The first sample was used for trial testing of the initial draft of the instrument after preliminary validation had been done, while the second sample was used for final testing of the instrument to establish validity and reliability of the final scale. The first sample consisted of 400 PG students in faculties of education from three universities in the study area. The researchers made use of a combination of multi-stage and stratified random sampling procedures. At the first stage simple random sampling technique was used to obtain three universities in the study area. The faculties of education in those three universities were used. At the second stage, four departments, from each of the three faculties were obtained through simple random sampling. At the third stage, students in each of the four sampled departments were stratified according to their programmes of study (Ph.D. and masters.) fifty percent (50%) of the students in each of the two programmes in each of the departments, were obtained through simple random sampling. The exercise gave rise to 400 PG students made up of 176 Ph.D. and 224 master's students.

The second sample consisted of 600 PG students from the faculties of education in four sampled universities. Any university previously used for trial-testing was excluded during this second sampling. A combination of multi-stage and stratified sampling procedures was again used. At the first stage, simple random sampling procedure was used to obtain four universities outside those that were used for trail testing. The faculties of education in these four universities were used for the study. At the second stage, four departments, from each of the four faculties, were obtained through simple random sampling. At the third stage, students in each of the four departments sampled were stratified according to programme of study and fifty percent of students in each programmed obtained of study and fifty percent of students in each programmed obtained through simple random sampling. The exercise gave rise to 600 PG students made up of 224 Ph.D. and 336 master's students.

Method of Data Analysis

Research Question 1 to 3 were answered by categorising the items according to the values of the indices.

Research Question 4 was answered by listing the items that were acceptable in terms of each of the indices. Research Question 5 was answered using Kuder- Richardson and Split-Half methods. To answer Research Question 6, the researchers ascertain evidences of construct validity. This was done by determining how data obtained with the instrument converged or diverged with measures of other constructs. In other words, evidences of convergent and divergent validity were ascertained by correlating scores obtained with the Statistical Literacy Scale with scores obtained with Test Anxiety Questionnaire (TAQ), and Statistical Reasoning [SRT]



Result and Discussion

Research Question 1: What are the difficulty indices of the items of the draft PG Students Statistical Literacy Scale?

The results are presented in table 1

Table.1: *Summary of Difficult Indices of Draft PG Students Statistical Literacy Scale.*

Range of indices of items		Items	Remarks
Below.30	43	4, 5, 7, 8, 9, 10, 11, 14, 15, 16, 17, 18, 20, 23, 25, 26, 27, 28, 29, 30, 32, 34, 36, 37, 38, 39, 43, 46, 50, 57, 58, 65, 68, 70, 73, 77, 81, 86, 97, 100, 101, 105, 108	Too difficult and rejected
.30-.70	64	1, 2, 3, 6, 12, 19, 21, 22, 24, 31, 33, 35, 41, 42, 44, 45, 47, 48, 49, 51, 52, 53, 54, 55, 56, 59, 60, 61, 62, 63, 64, 66, 67, 69, 71, 72, 74, 75, 76, 78, 79, 80, 82, 83, 84, 85, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 98, 99, 102, 103, 104, 106, 107, 109	Good and retained
Above .70	3	13, 40, 110	Too easy and rejected

The finding shows that 64 of the items had difficulty indices falling between 0.3 and 0.7. This finding is in line with that of Gourav (2015), who constructed and validated Achievement Test in Economics, and found that 70 items had difficulty values (DV) ranging from 0.20 to 0.75. An achievement test must include varying difficulty levels and should provide reasonable opportunity for students of different intellectual levels (Sidhu, 2007). Sidhu also opined that for a five – option multiple choice objective test items, the acceptable level of difficulty should fall between the range of .30 to .70. Asuru (2015) was of the opinion that in norm-referenced test, an acceptable item difficulty range is between 0.40 and 0.60 or optimum of 0.50. This finding is similar to the findings of Balogun and Mustapha (2014) who constructed and validated an achievement test for upper primary science. The 102 items had difficulty indexes which ranged from 0.30 to 0.70 and a discrimination index of 0.20 to 1.00. Also, the finding is similar to the study result of Okeme (2011). In the study, 50 test items were developed in Mathematics achievement test. The difficulty indices ranged between 0.30 and 0.8.

Research Question 2: What are the discrimination indices of the draft PG Students Statistical Literacy Scale?

The results are shown in table 2.

Table .2: *Summary of Discrimination Indices of the Draft Instrument*

Range of Indices items	No of item	Items	Remark
.30 and above	55	1, 6, 9, 12, 16, 17, 21, 22, 24, 27, 30, 32, 35, 36, 42, 44, 45, 49, 51, 52, 53, 54, 55, 56, 60, 61, 64, 67, 69, 72, 73, 74, 75, 76, 77, 78, 79, 82, 83, 85, 87, 89, 91, 93, 94, 95, 96, 98, 99, 102, 104, 105, 106, 107, 109	Good and retained
Below .30	55	2, 3, 4, 5, 7, 8, 10, 11, 13, 14, 15, 18, 19, 20, 23, 25, 26, 28, 29, 31, 33, 34, 37, 38, 39, 40, 41, 43	low and rejected



46, 47, 48, 50, 57, 58, 59, 62, 63, 65, 66, 68, 70,
71, 80, 81, 84, 86, 88, 90, 92, 97, 100, 101,
103, 108, 110

The study shows that (fifty-five) 55 items had discrimination indices of 0.3 and above. The items are good and were retained. Fifty-five items had discrimination indices below 0.3, and therefore were considered unacceptable. According to Sidhu (2007), the discrimination index of an item indicates whether or not the item is measuring the same ability as the test measures and shows how well an item discriminates between the able and weak students. Jayanthi (2014) and Sidhu (2007) were of the view that items would be considered acceptable if the facility indices were satisfactory and discrimination index 0.25 and better in excess of 0.40.

Research Question 3: What are the distracter indices of the draft PG Students Statistical Literacy Scale?

The results are shown in table 3. For an item to be acceptable all the distracters must have indices greater than 0.05

Table 3: *Summary of Distracter Indices of PG Students Statistical Literacy Scale*

Range of indices	No of items	Items	Remarks
One or more distracters with indices less than 0.05	46	2, 11, 15, 18, 20, 23, 25, 26, 28, 29, 33, 34, 35, 36, 37, 43, 44, 47, 48, 51, 57, 58, 59, 62, 65, 66, 68, 70, 71, 73, 77, 79, 80, 81, 84, 86, 88, 90, 92, 97, 100, 101, 103, 104, 108, 110	Poor and rejected
All distracters with indices more than 0.05 each	64	1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 16, 17, 19, 21, 22, 24, 27, 30, 31, 32, 38, 39, 40, 41, 42, 45, 46, 49, 50, 52, 53, 54, 55, 56, 60, 61, 63, 64, 67, 69, 72, 74, 75, 76, 78, 82, 83, 85, 87, 89, 91, 93, 94, 95, 96, 98, 99, 102, 105, 106, 107, 109	Good and retained

Table 3 shows that 46 items had one or more distracters with indices less than 0.05 and 64 items, each with all indices greater than 0.05. All the 64 items were therefore fit for inclusion in the final instrument. An effective distracter is a plausible wrong option of multiple-choice test meant to entrap testees who do not possess the knowledge and competence being assessed. The above finding is similar to that of Asuru (2015) who reported that good items should have plausible than to be too retained.

Research Question 4: What are the items that constitute the final PG Students Statistical Literacy Scale?

Table 4: *Items that Made the Final PG Students Statistical Literacy Scale*

Items:

1, 6, 12, 21, 22, 24, 42, 45, 49, 51, 52, 53, 54, 55, 56, 60, 61, 61, 67, 69, 72, 74, 75, 76, 78, 82, 85, 87, 89, 91, 93, 94, 95, 96, 98, 102, 104, 106, 107, 109



Research Question 5: what are the Evidences of reliability of the instrument?

In order to answer this question, Kuder- Richardson 20 and split- half methods of reliability estimation were employed.

The results are shown in table 5: Table 5: *Reliability Coefficients of the Scale, Using Kuder Richardson 20 and Split-Half Techniques.*

Reliability	Reliability	Remark
Kuder Richardson 20	Coefficient	
Split- Half	0.81	High
Kuder Richardson 20	0.87	High

The reliability coefficient of the final instrument obtained using the Kuder Richardson 20 technique is 87 while its coefficient when the split-held technique was applied is 0.81. The differences could be due to the fact that in the use of the split half technique, the error may come as a result of the fact that tow values separated may not produce the same result as the variances may get wider. However, Kuder Richardson 20 and Split Half method of reliability estimation demonstrated that the final instrument is reliable in a similar manner, the studies by Effiong (2006), Nkpone (2001) and Inosmiesa (1988) has reliability coefficients of 0.99,089 and 0.87 respectively using Kuder Richardson formula. The value of reliability is also similar to value of reliability coefficients obtained by Admu (2016), who conducted a study on development and preliminary validation of an instrument for assessment of psychomotor skills in physics which was found to be 0.87. Onah (2014) conducted research on development and standardization of agricultural science achievement test for senior secondary school. The reliability value was found to be 0.92. Okereke (2008), found the reliability index for mathematics achievement test to be 0.80. Kazeni (20005) reported a study on development and validation of a test in integrated science process skills for further Education and Training Learners. The reliability index of the 47-item achievement test was 0.81. The values of reliability coefficient were considered high reliability; thus, the present study is equally considered to have a high level of reliability coefficient, thereby making the instruments reliable.

In order to obtain evidences of convergent and divergent validity sure obtained with the instrument were correlated with scores obtained with statistical Reasoning test (SRT) and scores obtained with Test Anxiety Questionnaire {TAQ} The result is shown in table 6

Research Question 6: What are the evidences of validity of the instrument?

In order to obtain evidences of convergent and divergent validity, scores obtained with the instrument were correlated with scores obtained with Statistical Reasoning Test (SRT) and scores obtained with Test Anxiety Questionnaire (TAQ). The result is shown in table 6



Table 6: Multi- Trait Mono Method Correlational Analysis of PGS SLS, TAQ and SRT

Variables	PGSSLS	SRT	TAQ
PGSSLS	1	1	
SRT	.76		
TAQ	-.04	-.12	1

The PG Students Statistical Literacy Scale (PGSSLS) showed a positive correlation with the Statistical Reasoning Test (SRT) ($r = .76$), demonstrating evidence of convergence validity but showed an insignificant negative relationship with the Test Anxiety Test ($r = -.04$) demonstrating evidence of divergence validity. More so, the Statistical Reasoning Questionnaire (SRT) showed an insignificant negative relationship with the Statistical Anxiety Questionnaire ($r = -.12$). According to Cohen (1988) a coefficient of 0.66 is a large correlation. Based on this, the new instrument shows evidence of divergent and convergent validity. Fajobi (2019) suggested that convergent validity points out what other variables are correlated to the construct of interest and how they are correlated. A measure can accurately represent a variable if it correlates or converges with other supposed measures of that variable. It demonstrates that a variable is not just an accident of a particular measure. Zwick (2013) described divergent validity as evidence that a measure is not assessing something it is not supposed to assess. Kunuba and Anigbo (2023) stated that discriminability means that one can empirically differentiate the construct from another construct that may be similar, and that one could point out what is unrelated to the construct. Fajobi (2019) indicated that, as an aspect of construct-related evidence of validity, discrimination is evaluated by considering the measure's correlation with variables that the theory of the construct postulates to be irrelevant to that construct. A comparison construct postulated as irrelevant should not be just any construct, but rather one that at least represents some potential aspect of the construct under investigation (Obilor, 2019).

Discussion of Findings

This finding has shown that the PG students Statistic Literacy Scale is very appropriate and useful for the assessment of students' Statistics Literacy. This finding is in line with the finding of Gourav (2015), who constructed and validated Achievement Test in Economics, and found that 70 items having difficulty value (DV) ranging from 0.20 to 0.75. An achievement test must include varying difficulty levels and should provide reasonable opportunity for students of different intellectual levels (Sidhu, 2007). Sidhu also opined that for a five – option multiple choice objective items, the acceptable level of difficulty should be between the range of .30 to .70. Asuru (2015) was of the opinion that in norm-referenced test, an acceptable item difficulty range is between 0.40 and 0.60 or optimum of 0.50. This finding is similar to the findings of Balogun and Mustapha (2014) who constructed and validated an achievement test on upper primary science. The 102 items had a difficulty index which ranged from 0.30 to 0.70 and discrimination indices ranging from 0.20 to



1.00. Also, the finding is similar to the study result of Okeme (2011). In the study, 50 test items were developed in Mathematics. The difficulty indices of ranged firm 0.30 to 0.8.

They study shows that 63 items have discrimination indices of .20 and above. The items are good and were retained. 27 items have a discrimination index between 0 and .19. Those items have low discriminatory power and were rejected; while 20 items have in the new presentation there were no negative indices. These results are similar to the findings by Obilor and Obubere (2020) who found in their study that 58 items out of 70 items had item difficulty indices ranging from 0.43 to 0.61. In the opinion of Riaz (2008), item discrimination refers to the ability of an item to differentiate among students on the basis of how well they know the material being tested. According to Sidhu (2007), the discrimination index of an item indicates whether or not the item is measuring the same ability as the test measures and shows how well an item discriminates between the able and weak students. Jayanthi (2014) and Sidhu (2007) were of the view that items would be considered acceptable if the facility indices were satisfactory and discrimination index 0.25 and better in excess of 0.40.

They study Does not agree with what is in Table 4.3 .05. The items were rejected as their distracters were not plausible Does not agree with info in Table 4.3 retained in the final instrument as their distracters were not plausible. An effective distracter is a plausible wrong option of multiple-choice test meant to entrap testees who do not possess the knowledge and competence being assessed. The above finding is similar to that of Asuru (2015) who reported that good items should have plausible distracters for them to be retained.

The PG Statistical Literacy Scale (PGSLS) showed a positive correlation with the Statistical Reasoning Test (SRT) ($r = .76$) demonstrating evidence of convergence validity but showed an insignificant negative relationship with the Statistical Anxiety Test ($r = -.04$) demonstrating evidence of divergence validity. According to Cohen (1988) a coefficient of 0.66 is a large correlation and this evidence justifies the convergent validity. Fajobi (2019) suggested that convergent validity points out what other variables are correlated to the construct of interest and how they are correlated. A measure can accurately represent a variable if it correlates or converges with other supposed measures of that variable. It demonstrates that a variable is not just an accident of a particular measure. Zwick (2013) described divergent validity as evidence that a measure is not assessing something it is not supposed to assess. Kunuba and Anigbo (2023) stated that discriminability means that one can empirically differentiate the construct from another construct that may be similar, and that one could point out what is unrelated to the construct. Fajobi (2019) indicated that, as an aspect of construct-related evidence of validity, discrimination is evaluated by considering the measure's correlation with variables that the theory of the construct postulates to be irrelevant to that construct. A comparison construct postulated as irrelevant should not be just any construct, but rather one that at least represents some potential aspect of the construct under investigation (Obilor, 2019).



The reliability coefficient of the final instrument obtained using the Kuder Richardson 20 technique is .87 while the coefficient of the statistical literacy scale when the split - half technique was applied is 0.81. The differences could be due to the fact that in the use of the split half technique, the error may come as a result of the fact that two values separated may not produce the same exact result as the variances may get wider. However, Kuder Richardson 20 and Split Half methods of reliability estimation demonstrated that the final instrument is reliable. In a similar manner, the studies by Effiong (2006), Nkpone (2001) and Inosmiesa (1988) had reliability coefficients of 0.99, 0.89 and 0.87 respectively using Kuder Richardson formula. The value of reliability is also similar to value of reliability coefficients obtained by Adonu (2016), who conducted an intensive study on development and preliminary validation of an instrument for assessment of psychomotor skills in physics which was found to be 0.87. Onah (2014) conducted research on development and standardization of agricultural science achievement test for senior secondary school. The reliability value was found to be 0.92. Okereke (2008), found the reliability index for mathematics achievement test to be 0.80. Kazeni (2005) reported a study on Development and Validation of a Test in Integrated Science Process Skills for Further Education and Training Learners. The reliability index of the 47 items achievement test to was 0.81. These values of reliability coefficient were considered high reliability; thus, the present study is equally considered to have a high level of reliability coefficient, thereby making the post Graduate students' Statistical Literacy Scale very valuable.

Conclusion

This study focused on the development and validation of an instrument for assessing statistical literacy among [target population]. The development process involved defining the construct, identifying its dimensions, generating items, obtaining expert judgement, pilot testing, conducting item analysis and establishing evidence of validity and reliability. Following the validation process, [number] items were retained in the final Statistical Literacy Assessment Instrument. The instrument demonstrated [insert actual evidence] of validity and reliability and may therefore provide a useful measure of statistical literacy among the target population. However, the instrument should be used within the population and context for which evidence has been established. Further studies may examine its performance with larger and more diverse populations.

Recommendations

Based on the findings, the following recommendations are made:

1. Teachers and researchers should assess statistical literacy through items that require interpretation as well as computation.
2. The developed instrument may be used by researchers for assessing statistical literacy among the target population after appropriate administration procedures.
3. Statistics educators should emphasize interpretation of tables, graphs and statistical statements.



4. Further studies should validate the instrument with students from other geographical areas and educational levels.
5. Researchers who adapt the instrument to a substantially different population should conduct additional validity and reliability studies.

REFERENCES

- Adamu (2016), similar value of reliability coefficient obtained by Effiong [2006], Nkpone [2001] and Inosmiesa[1988]
- Asuru (2015), was of the opinion that in norm-referenced test an acceptable item difficulty range is between 0.40 and 0.60 or optimum of 0.50
- Arah, A.N. (2020). Reality Television and Participatory Development in Nigeria's Niger Delta: Dawn in the Creeks as a Paradigm. *Creative Artist: A Journal of Theatre and Media Studies* 14(1).
- Arah, A.N. (2017). Reality Television and Participatory Development in Nigeria's Niger Delta: Dawn in the Creeks as Example. *Department of Theatre and Film Studies, faculty of Arts, Nnamdi Azikiwe*.
- Arah, A. N. (2017). Reality Television, User-generated Content and Quality Assurance". *Journal of Humanities and Cultural Studies (JHUCS)*, 2(1). Awka: Centre for Arts, Culture and Humanities. Tetfund, Nnamdi Azikiwe University, Awka, pp 54-72.
- Arah, A. N. (2017). Reality Television and the Change Imperative: A Critical Appraisal of Jeta Amata's Dawn in the Creeks". *The Creative Artist: A Journal of Theatre and Media Studies*, 13(2), pp 80-90.
- Arah, A. N. (2025). Reverse-Acculturation and Identity: An Overview of Recent Tittle-taking Trend among Diaspora Onitsha Indigenes". *Journal of Humanities and Leadership Studies: Journal of Faculty of Arts, Dennis Osadebey University, Asaba, Delta State*, 1(2), 2025, 2971-5741.
- Batamero, C. and Diaz, C. (2006). *Methodological and didactical controversies around statistical inference. Actes du 36iemes. Journees de la Societe Franqaise de ¹ Statistique. Paris: Societe Francaisede Statistique.*
- Ben-Zvi, D., and Garfield, J. (2004). *The challenge of developing statistical literacy, reasoning and thinking*. Netherlands: Kluwer Academic Publishers
- Binnie, N. (2002). Using projects to encourage statistical thinking. In B. Phillips" (Ed.) *Proceedings of the sixth International Conference on the Teaching of Statistics*. Cape Town: International Association for Statistical Education



- Budgett, S. and Pfannkuch M. (2007) *Assessing students' statistical literacy*. International Association for Statistical Education (IASE/ISI Satellite, 2007 Portugal), <http://www.StatLit.org/pdf/2007BudgettPfannkuchIASE.pdf>
- Carik, H. and Ari, A. G. (2022). Development and validation of an achievement test in biology, *Journal of Social Sciences and Education*, 3(5), 64-75,
- Carole L. K., and Winterstein, A. G. (2014). Validity and reliability of measurement instruments used in research. *American Journal Health-System Pharmacy*, 65(1), 21-27.
- Chime, U. M. (2012). *Development and validation of economics achievement test for senior secondary school students*. Unpublished M.Ed. Dissertation, University of Nigeria, Nsukka. .
- Cohan, J. (1988) *Statistical power analysis for the behavioural science (seconded Hill shade, N. J Lawrence Erlbaum Associate publisher*
- Cohen, R. J., Swerdik, M. E. and Sturman, E. D. (2010). *Psychological testing and assessment: An introduction to tests and measurement* (7th Edition). New York: McGraw-Hill Inc.
- Colin, R. and Hay, O. (2009) Development and validation of the students' self-efficacy for statistical literacy scale. *Structural Equation Modeling*, 9, 413-430.
- Doyle, P. (2008). *Developing statistical literacy with students and teachers in the secondary mathematics classroom*. Unpublished Masters' Thesis, School of Education, University of Waikato
- Ebel, R. L. (2009). *Essentials of educational measurement*. Engelwood Cliffs, N.J: Prentice Hall
- Ebo, E. C. (2009). *Social and economic research: Principles and method*. Enugu: African Institute for Applied Economics.
- Effiong, E. J . (2006); *Development validation of alternative to practical test for measurement skills in electronic devices and circuit on technical college*. Unpublished PhD thesis, University Nigeria, Nsukka.
- Erguven, M. (2014). Two approaches to psychometric process: Classical test theory and item response theory. *Journal of Education*, 2(2), 23-30.
- Eze, M. J., and Onah, W. M. (2005/ *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.
- Ezeh, D.N. (2003). Reliability and validation of tests. /«"B. G. Nworgu (Ed.), *Educational measurement and evaluation: Theory and practice* (pp. 123-135). Nsukka: University Trust Publishers.
- Fajobi, O. O. (2019). The development and validation of mathematics multiple choice test items for Junior Secondary School Three (JSS 3) Students in Oyo State, Nigeria. *Global Scientific Journal*, 7(11), 23-43.
- Fergusson, G. (2011). *A guide to language testing: Development, evaluation, and research*. Cambridge Mass: Newberry House Publisher



- Gal I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25
- Garfield, J. B., (2003). Assessing statistical reasoning. *Statistics Education Research*, 5(3), 78-89.
- Garfield, J., and Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372-396
- Gessell, N. R. (2004). *Classroom management: Principles and practice*. London: George Allen and Unwin.
- Gonda D, Pavlovicova G, Duris V, and Tirpakova A. (2022). Implementation of pedagogical research into statistical courses to develop students' statistical literacy. *Mathematics*, 10(11), 45-50.
- Gourav, M. (2015). Construction and validation of achievement test in economics. *International Journal of Humanities & Social Science Studies*, 1(IV), 54-60.
- Hahi, J. (2009) Statistical literacy scale development for undergraduate students. *Jpurba; of Social Science Research*, 23. 567-589
- Harbor-Peters, V. F. (1999). *Noteworthy points on measurement & evaluation*, Enugu: Snap Press Ltd.
- Haw, L. H. and Sharif, S. B. (2020). The development and validation of science achievement test. *Journal of Education and Practice*, 11(20), 103-109.
- Ifeakor, A. C. (2011). *Psychological measurement & evaluation in Education: Issues and application*. Onitsha: Folmech Printing and Publishing Co.Ltd.
- Inomiesa, E. A (1988); the development validation and use of standardized instrument for the continuous assessment of pupils achievement in upper primary science .unpublished PhD thesis. Department of science education, University of Nigeria, Nsukka.
- Iseri C. S (2008). Validation and reliability of mathematical achievement test journal of mathematics education 8(2) 64-68
- Jayanthi, J. (2014). *Development and validation of an achievement test on Mathematics*. International Journal of Mathematics and Statistics Invention (IJMSI), 2(4), 40-60.
- Kabaca, T. and Y., Erdogan, (2007). "Fen bilimleri, bilgisayar ve matematik alanlarında akıtezhahşmalarinin istatistiksel ağıdan incelenmesi", Pamukkale University, Eğitim Fakultesi Dergisi, 22(54), 54-63.
- Kame, I. (2011). Development and validation of psycho-productive skills multiple choice items for students in agricultural science in secondary school; unpublished PhD thesis submitted to the department of educational teacher education, University of Nigeria, Nsukka.
- Karl, O. (2022). *Development and validation of mathematics test for senior secondary schools*, International Journal of Scientific and Research Publications, 12(7), 342-356.
- Kazeni M. M. (2005). Development and validation process skills for the further education and training of learners. Msc. Thesis Pretoria: University of Pretoria



- Keosayian, S. and Johnson, D. (2014) *Development of statistical literacy in undergraduate students*. *Jo urnal of Statistics Education*, 14(2), 56-60.
- Kiregyera, B. (2010/ The evidence gap and hs impact on public policy and decision-making. http://iaseweb.org/documents/papers/icots8/ICOTS8_IB3_KIREGYERAI.pdf
- Kunuba, A. C., and Anigbo, L. C, (2023). Assessing the validity of Christian religious studies multiple-choice items of Enugu State basic education certificate examination. *Godfrey Okoye University International Journal of Education*, 5(1), 11-17.
- MacGillivray, H. L. (2009). Learning support and students studying mathematics and *statistics*. *Inter national Journal of Mathematical Education in Science and Technology*, 40(4), 455-472.
- Magno, C. (2016). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment*, 7_s(1), 1-11.
- Man, P. O (2016) *Education in Nigeria a futuristic perspective*. Retrieved from [.csd.uwa.edu.au/icedu2002](http://csd.uwa.edu.au/icedu2002)
- Meredith, S., Joyce, N. and Walter, O. (2007). *Basic concepts in item and test*. Retrived from files.eric.ed.gov/fulltext/ED406441.pdf.
- Nkpone, H. C. (2001). *Development and standardization of physics achievement test for senior secondary students, unpublished MED thesis, faculty of education, University of Nigeria, Nsukka*
- Obilor, E. I. and Obubse S. H. (2020) *Development and validation of economics achievement test*. *Journal insurance education* 8(2) 64-69
- Obilor, E. I. and Akpan, U. T. (2020). Construction and validation of basic science achievement test for junior secondary three students in public secondary schools in Akwa Ibom State, *International Journal of Innovative Education Research* 5(4):124-135.
- Obilor, I. E. (2019). *Essentials of measurement and evaluation*. Port Harcourt: Sabcos Printers & Publishers.
- Odo, C. I. and Ugwoji, J. N. (2016). Development and validation of biology achievement test for assessment of students in Enugu State. *International Journal of Innovative and Applied Research*, 4(5), 23-28.
- Okafor, D. P. (2007). *Psychological testing: Principles, applications and issues*. Pacific Grove: Brooks Cole Pub. Company
- Okebukola, P.A.O. (2017). *Global competitiveness in education: Implications for educational assessment in Nigeria*. 2017 National Examinations Council (NECO) Annual Guest Lecture.
- Okoye, R. O (2015). *Educational and psychological measurement and evaluation*. Awka : Erudition Publishers.
- Olaitan, S. O. (2005). *Understanding curriculum*. Nsukka: Ndudim Press.



- Onunkwo, G .1 .N. (2006). *Fundamentals of education measurement and evaluation*. Owerri: Cape Publishers Int'l Ltd.
- Onunkwo, W. M. (2002). *Introduction to measurement theory*. Monterey, CA: Brooks/Cple Publishing Company.
- Patel, C. L. (2010). Approaches to extra-curricular statistics support for non-statistics undergraduate and post graduate: Facilitating the transition to higher education. *Journal International Statistical Institute*, 3(6), 12-19.
- Riaz, M. N (2008) *Test construction development and standardization of psychologized test in Pakistan Islamabad HEC* Rumsey, D. (2002). *Statistical literacy as a goal for introductory statistics courses*. Retrieved From www.amstat.org/publications/jse/v1On3/rumsey2.html.
- Schild, M. (2008). *Information literacy, statistical literacy and data literacy. I. Quarterly*, 28 (3), 6-11
- Schumacker, R. E. (2010'), *Classical test analysis*. Retrieved from <http://appliedmeasurementassociates.com/ama/assets/File>.
- Sidhu, K.S. (2007). *New approaches to measurement and evaluation*, New Delhi, India. Sterling Publication Pvt. Ltd.
- Singh, A.S. (2017). Common procedures for development, validity and reliability of a questionnaire. *International Journal of Economics, Commerce and Management*, 5(5), 790-801
- Singh, A.S., and Masuku, M. B. (2012). Understanding and applications of test characteristics and basic inferential statistics in hypothesis testing, *European Journal of Applied Sciences*, 4(2), 90-97. .
- Sundre, L. A., (2011). *Quantitative reasoning quotient scale*. Center for assessment and research studies. Harrisonburg, VA: James Madison University.
- Ukar, Y. (2012) *making data meaningful*, A guide to improving statistical literacy, <http://worrydream.com/MagicInk/>, downloaded 1th June.
- Usani, J. O., Isaac, V. O., Samuel, A., Sylvia, V. O., Nworgwugwu, O., Out, B., and Atah, C. A.(2022). Development, validation and standardization of instrument to measure statistical literacy among universities undergraduates in Cross River State. *Mediterranean Journal of Social Sciences*, 13(6}, 45-56.
- Watson, J. (1997). Assessing statistical thinking using the media. In I. Gal & J.Garfield (Eds.), *Handbook of Assessment in Statistics Education (pp. 107-121)*. Amsterdam: IOS Press and the International Statistical Institute.
- Watson, J. (2003). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior*, 19, 109-136.
- Watson, J. M. (2005), 'Is statistical literacy relevant for middle school students?' *Vinculum*, 42(1}, 66 -81.
- Watson, J. M. (2006). *Statistical literacy at school: Growth and goals* Mahwah, NJ: Lawrence Erlbaum.



- Watson, J., and Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3-46.
- Zwick, M. (2013). *European masters in official statistics*, http://www.croportal.ei^sites/default/files//NTTS2013fullPaper_231%20zwick.pdf, downloaded 15th April 2014.
- Zwick, R. (2013). *Development and validation of mathematics economics test journal of science education*, 7(3) &1-79.

Appendix

PG Students' Statistical Literacy Scale [PGSSLS]

Name (optional) -----
Sex-----
School-----
Type of programme-----
Reg No-----

Time: 1hr

Instruction

1. The value of a variable is called
 - A. Estimate
 - B. Statistic
 - C. Variate
 - D. Mean
2. Ordinal data can be analysed using the
 - A. Mean
 - B. Mode
 - C. Median
 - D. Standard deviation
3. The standard deviation is equal to:
 - a. The variance
 - b. The square root of the variance
 - c. The variance squared
 - d. The variance divided by the number of scores
4. Which of the following descriptive statistics is least affected by adding an outlier to a data set?
 - a. the mean
 - b. the median



- c. *the range*
- d. *the standard deviation .*
- 5. *If the p-value is greater than alpha in a two-tail test, what conclusion should you draw?*
 - A. *The null hypothesis should not be rejected.*
 - B. *The null hypothesis should be rejected*
 - C. *A one-tail test should be used*
 - D. *Alpha level is appropriate*
- 6. *A repeated t-test can be used to test which of the following*
 - A. *Goodness of fit*
 - B. *Relationship between two groups of interval data set*
 - C. *Correlated sample data*
 - D. *Relationship between two ratio data set*
- 7. *Which of the following statistics are important when interpreting an independent sample t-test*
 - A. *Mean difference*
- 8. *In a study without a pretest, subjects are randomly assigned to one of three groups: control, experimental A, or experimental B. After treatment, the mean scores for the three groups are compared. The appropriate statistical test for comparing these means is:*
 - a. *the correlation coefficient*
 - b. *chi square*
 - c. *the t-test*
 - d. *the analysis of variance*
- 9. *You carried out an ANOVA on a preliminary sample of data. You then collected additional data from the same groups; the difference being that the sample sizes for each group were increased by a factor of 10, and the within-group variability has decreased substantially. Which of the following statements is NOT correct?*
 - a. *The degrees of freedom associated with the error term has increased*
 - b. *The degrees of freedom associated with the treatment term has increased*
 - c. *SSE has decreased*
 - d. *F calculated has changed*
- 10. *ANCOVA is based on*
 - A. *T-test*
 - B. *Z-test*
 - C. *F-test*
 - D. *Chi-square*
- 11. *Which of following assumption may not be considered when carrying out Ancova*
 - A. *Homogeneity of variance*



- B. Independence of observation
 - C. Normality
 - D. Discrepancy of observation
12. ANCOVA is more appropriate when the covariate is
- A. linearly related to the dependent variables
 - B. Linearly related to the independent variables
 - C. Linearly related to moderator variable
 - D. Asymmetrical distributed
13. The variance in Analysis of covariance (ANCOVA) that is being analyzed must be
- A. Dependent
 - B. Interdependent
 - C. independent
 - D. Mutually exclusive
14. In Analysis of covariance (ANCOVA), the different types of the independent variables are assumed to be drawn from the normal population having a mean of
- A. 2
 - B. 0
 - C. 4
 - D. 3
15. Ancova is more appropriate in a
- A. Post test research design
 - B. Time series research design
 - C. Correlational research design
 - D. Pre and post research design
16. The estimated marginal means section of the output in ANCOVA gives the
- A. Adjusted means
 - B. Effect mean
 - C. Pre-test mean
 - D. Interaction mean
17. what must a levene's test be in order to use an ANCOVA
- A. significant at $p < 0.001$
 - B. non-significant at or above $p > 0.05$
 - C. non-significant at $p < 0.001$
 - D. significant at $p < 0.01$

Use the statement below to answer Question 18 relate to the following:

Dr Okafor is analysing scores on a memory test for four groups of people who have taken different amounts of alcohol. He has an idea that memory is related to IQ, so he decides to control for that and chooses ANCOVA for his inferential test.

18. Which is the covariate?
- a. Scores on the memory test



- b. The amounts of alcohol
 - c. IQ
 - d. The control group
19. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study: $y = -0.0127 + 0.0180x$. The above equation implies that:
- a. each beer consumed increases blood alcohol by 1.27%
 - b. on average it takes 1.8 beers to increase blood alcohol content by 1%
 - c. each beer consumed increases blood alcohol by an average of amount of 1.8%
 - d. each beer consumed increases blood alcohol by exactly 0.018
20. In regression analysis, the variable that is being predicted is the
- a. response, or dependent, variable
 - b. independent variable
 - c. intervening variable
 - d. moderating variable
21. A regression analysis between sales (in N1000) and price (in Naira) resulted in the following equation: $Y = 50,000 - 8X$. The above equation implies that an
- a. increase of N1 in price is associated with a decrease of N8 in sales
 - b. increase of N8 in price is associated with an increase of N8,000 in sales
 - c. increase of N1 in price is associated with a decrease of N42,000 in sales
 - d. increase of N1 in price is associated with a decrease of N8000 in sales
22. If two variables, x and y , have a very strong linear relationship, then
- a. there is evidence that x causes a change in y
 - b. there is evidence that y causes a change in x
 - c. there might not be any causal relationship between x and y
 - d. there is no moderating relationship between X and y .
23. There is a very strong correlation between two variables then the correlation coefficient must be
- a. any value larger than 1
 - b. much smaller than 0, if the correlation is negative
 - c. much larger than 0, regardless of whether the correlation is negative or positive
 - d. None of these alternatives is correct.
24. If the coefficient of determination is equal to 1, then the correlation coefficient
- a. must also be equal to 1
 - b. can be either -1 or +1
 - c. can be any value between -1 to +1
 - d. must be -1
25. By which other name is the chi-square goodness of fit test known?
- A. Wilcoxon



- B. Two sample chi-square
 - C. Directional chi-square
 - D. One-sample chi-square
26. How many cases need to appear in one category for Chi-square?
- A. You have to redo your experiment
 - B. You can exclude outliers
 - C. You can transform your data
 - D. You can increase your sample size or combine categories
27. By which other name is chi-square contingency table analysis known?
- A. Chi-square test for independence
 - B. Independent samples chi-square test
 - C. Directional chi-square test
 - D. One sample chi-square
28. Which tests test could be used if the expected cases in chi-square analysis are few
- A. Cronbach's alpha or content analysis
 - B. Person or spearman's correlations
 - C. Yates' continuity correction
 - D. Paired or independent t-test
29. How is Spearman's Rho calculated?
- A. Spearman Rho is calculated by transforming original data into categories
 - B. spearman's Rho is calculated by transforming original scores into ranks
 - C. spearman's Rho is calculated by transforming original scores into deviations
 - D. spearman's Rho uses the original scores, no transformations are needed
30. The -----test is calculated by ranking all of the participants scores from lowest to highest and adding up ranks separately for each condition
- A. Spearman's Rho
 - B. Mann Whitney
 - C. Friedman's ANOVA
 - D. Pearson's r
31. Under what circumstances would you use a non-parametric test
- A. When you think your sample size is too big
 - B. When your data does not meet the assumptions for a parametric test
 - C. When you do not really understand a parametric test
 - D. When your data are symmetrically distributed
32. What does ASL stand for
- A. Asymmetrical significance level
 - B. Asymptotic significance level
 - C. Asymbolic significance level
 - D. Assigned significance level
33. What is the formular for Kruskal- Wallies based upon?
- A. Ranks



- B. Categories
 - C. Deviations
 - D. Means
34. What statistic is used to check the significance of the Kruskal-wallies test
- A. Mean rank
 - B. T
 - C. Chi-squared
 - D. Partial n^2
35. Which of the following tests would you use to assess whether there is a significant difference between the mean ranks of two conditions?
- A. Friedman
 - B. Wilcoxon
 - C. Kruskal-Wallis
 - D. Spearman's Rho
36. If your data met the assumption of homogeneity of variance. Which of the following post hoc is more appropriate
- A. Honesty significance difference
 - B. Turkey method
 - C. Scheffe
 - D. Games Howell test
37. In post hoc analysis simultaneous confidence interval applies to
- A. Some groups of comparison
 - B. Entire family of comparison
 - C. Groups that are not significant
 - D. Groups with low mean
38. The more groups you have in a post hoc the -----the statistical power of comparison
- A. Lower
 - B. Higher
 - C. Stringent
 - D. Accurate
39. Which of the following post hoc analysis uses harmonic mean of the cell size of the comparison groups of means
- A. Honesty significance difference
 - B. Turkey method
 - C. Cramer V
 - D. Turkey Kramer test
40. Which of the following post hoc test analysis is based on F-distribution
- A. Scheffe's method
 - B. Cramer V method
 - C. Turkey method
 - D. Dunnet method



Scoring Guide for Final Scale

1C	11A	21A	31A
2C	12A	22C	32A
3B	13B	23D	33C
4C	14D	24B	34D
5A	15A	25D	35C
6C	16B	26D	36A
7D	17B	27C	37B
8B	18A	28C	38A
9C	19C	29C	39B
10D	20A	30A	40A