Research Article

Received: 16 August 2022 / Accepted: 28 October 2022 / Published: 5 November 2022

Development, Validation and Standardization of Instrument to Measure Statistical Literacy among Universities Undergraduates in Cross River State

Usani Joseph Ofem[1]

Isaac Ubi Ofem[2]

Samuel Akpan[2]

Sylvia Victor Ovat[2]

C. E. Nworgwugwu[2]

Bernard Otu[2]

Cletus Akpo Atah[3]

[1]CEO, De Oracle Research Consultancy, Cross River State, Nigeria
[2]Department of Educational Foundations,University of Calabar, Calabar, Nigeria
[3]University of Calabar, Calabar, Nigeria

DOI: https://doi.org/10.36941/mjss-2022-0047

Abstract

The study developed, validated, and standardised a statistical literacy test for university undergraduates in Cross River State, Nigeria. Five research questions were raised to guide this study. The study adopted an instrumentation design. A stratified sampling technique was used to select a total of 753 respondents from a population of 4705 students. Three instruments were used to collect data: the Statistical Literacy Test (SLT), the Statistical Reasoning Test (SRT), and the Statistical Anxiety Test (SAT). Descriptive statistics (means and simple percentages), correlational analysis, and the factor analytical technique were used to analyse the data collected. The result showed that only 25 items survived the last item assessment. The instrument was found to have strong construct validity as the items in the correlation matrix were all positively correlated. The statistical literacy scale showed convergence with statistical reasoning but diverged with statistical anxiety. The reliability of the instrument was found to be high as the coefficients of the scale were 0.81, 0.77, and 0.78 respectively, and it was adjudged reliable. Based on the findings of the study, it was concluded that the instrument was valid and reliable. Thus, it was recommended that students' skills in statistics should not just be on the computational aspect but on the utilisation of statistical knowledge. Statistics lecturers should emphasise statistical literacy to help students reason and think statistically, and the instrument should be widely distributed for use by researchers and psychometricians.

Keywords: *Development, Validation, Standardization, statistical literacy, statistical reasoning, and statistical anxiety*

1. Introduction

The world has become so technologically advanced that there is a global demand for information. Without a doubt, the power of information is enormous in today's world, especially when obtained through scientific means. Scientific procedures refer to information that is obtained in a systematic manner with little or no subjectivity. The rationale for global acceptance of systematically obtained information stems from the fact that such information can be verified or statistically proven, even if it is relative and can be used for value judgment.

University systems all over the world have the critical function of producing scientific thinkers through research, and graduates are expected to be active members of society with extensive knowledge of critical and analytical information skills that will be applied when confronted with statistical reports. Such graduates are expected to be dynamic thinkers, learners who are enduring and capable of navigating the complexities of an uncertain world. For example, challenging statements and research reports are frequently presented in media reports, and rather than approaching them with emotions, headlines can be critically examined to distinguish between credible and unbelievable information. It is critical that citizens in the twenty-first century have the necessary skills to interpret all information in a meaningful way, especially when it involves simple graphs and tables.

According to Ben-Zvi and Gafield (2004), most of the information presented in research outputs is incorrect, which has the potential to distort knowledge as well as wreak havoc on the educational system. Watson (2004) stated that statistical literacy has become a major goal of many curricula around the world. This is because students are expected to learn how to interpret the results of studies and other reports they present, as well as ask questions about those reports and explain why they are the way they are when necessary. More specifically, learners need to have the capacity to use information gotten from a finding to diagnose the study's quality, appreciate or critically examine the statistical techniques used, and provide justification for the correctness or incorrectness of a particular result. Where this is lacking, it becomes a disaster because anything in the name of assessment can be done, potentially affecting the learner's prospects.

Even though students and teachers are not trained to be professional statisticians, they must have a basic understanding of statistics. Gall (2002) asserted that "for full participation in our increasingly data-driven society, students must be able to interpret and critically evaluate statistical information, data-related arguments that they encounter in a variety of contexts, and, when appropriate, to discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of this information." Thus, having a thorough understanding of statistics in the social sciences can help students accumulate a wide range of difficult concepts, take active part in debates, and hold tenaciously to their rights.

Despite the importance of statistical literacy in a developing economy like Nigeria, Ogbuegu (2011) found that students' statistical literacy in tertiary institutions is extremely low. This manifests itself in the student's inability to think and reason statistically. Most students struggle to understand statistics' computational and procedural aspects, interpret graphs or results, or make sense of what is simply presented to them. In most cases, students find it difficult to make a simple inference from statements involving numbers, which explains why most of them fail in defence in research and statistical courses. This situation has sparked a great deal of concern, particularly regarding what is likely to be the cause of students' low statistics literacy. The causes of the problem are deduced as follows: statistical literacy is still conceived as a new area when compared to other learning areas, students' diverse backgrounds and abilities, negative past experiences with mathematics at the fundamental level, lack of interest in the subject, perceived irrelevance of the subject by most students, and poor teaching. This low level of statistical literacy manifests itself in students' inability to diagnose information presented in printed and non-printed media, inability to participate in societal discourse that may require facts, inability to interpret anything that requires numbers, poor reasoning with mathematical figures, and excessive anxiety when presented with numbers

(Udinwnege, 2011; Ogbuegu, 2011; Sharma, 2017; Kaplan & Thorpe, 2010).

Jacobe, Foti, and Whitaker (2014), as the need for transferring knowledge about statistics increases, so does the demand for instruments to measure learners' conceptual knowledge of statistics. As a result, it is critical that a validated instrument with no or minimal error be made available for the study of this construct. However, this field of study has received little attention in this part of the world. The instruments available for studying this construct are not culturally appropriate (Gaal, 2004, Watson, 2016; and delmas, 2002). Thus, using most of the instruments used in other foreign areas in the Nigerian context, specifically Cross River State, may introduce bias and fail to provide us with the necessary information. In this context, the researcher aimed to create, validate, and standardize an instrument for measuring statistical literacy among undergraduate students in universities in Cross River State, Nigeria.

2. Statement of the Problem

Over the years, it has been observed that attention has been drawn to the relevance of statistical literacy among students in educational settings. The rationale is that, regardless of discipline, there is an urgent need to acquire statistical literacy to critically evaluate, discuss, and make sense of information presented in the media or printed materials. However, with the rising interest in the subject area, there is a need for a validated instrument that can facilitate the activities of researchers. Unfortunately, available instruments, as shown in literature, are not culturally fair. This is because these instruments were developed in societies that have different technological inclinations, scientific orientation, learning environments, cognitive ability, facilities, as well as standards of living, from where the study is situated, among others. It is not to my knowledge whether any instrument to measure statistical literacy is available in Nigeria and particularly in Cross River State. Even where there is a Nigerian-made instrument, it is feared that the time that it was developed and now may differ, and changes in society may make them outdated. It is this presumed unavailability that has motivated the researcher to delve into the development, validation, and standardisation of an instrument to measure statistical literacy that will match the cultural context of the students in Nigeria.

3. Research Questions

The questions raised for the study are as followed:

i. What are the processes that were involved in the development of the scale?
ii. What are the characteristics of the items of the SLT scale?
iii. How does the instrument converge and diverge with other instruments?
iv. What is the reliability of the instrument when spilt half and Cronbach alpha is used?

4. Theoretical Background

4.1 Classical test theory (CTT) by Cronbach (1977)

Charles Spearman is widely regarded as the originator of this test theory. The classical test theory holds that every measurement contains some element of error, that these errors are random variables, and that they can be related Most classical approaches assume that a single individual's raw score (X) is composed of a hypothetical score assumed to be true (T) and an error that is random (R). For this, X = T + E.

By applying this theory to the study, classical test theory provides a foundation for explaining test behaviour and characteristics.. CTT also assists us in determining the examinee's cognitive ability in a specific construct, which can aid us in critically evaluating and synthesising the quality of items used and how to improve them for objectivity. These analyses look at the quality of the items as well as the test in its aggregate.. Such analyses can also be used to revise and improve both individual items and the test. As a result, because CTT is part of the total score concept, we must ensure that the test instrument's validity and reliability are established. Conditions that tend to raise or lower a test taker's score must be fixed in order to get a score that accurately reflects the test taker's skills.

4.2 *Conceptual framework*

According to Sharma (2017), statistical literacy is the ability to understand and deeply evaluate statistical findings that beset us and present themselves daily on different outlets, as well as the competence to identify the contributions that statistical knowledge is able to make to professional and individual decisions. According to Chick, Pfannkuch, and Watson (2005), literacy in mathematics is defined as deeper thinking that utilises the information concerning numbers to make meaning of the world around them as well as make efforts to learn the content of the information presented to deduce further meaning. Gal (2004) defines literacy in statistics as individuals' competence to make use of reports or information that is obtained, ability to interpret and critically evaluate statistical information, make thorough arguments based on data generated, and to make contributions to statistical discussion and communicate same when needs arise,

According to Garfield, delMas, and Chance (2003), statistical thinking entails understanding of why and how statistical activities are carried out and understanding "big ideas" such as the nature of differences and sampling, the use of data analysis methods and the display of clear data, and research methods to claim causality. Also, statistical thinking means knowing how to use the different models that help you understand the problem in the context in which they were used.

Garfield and Ben-(2007, Zvi's 2008) perspective on statistical literacy, reasoning, and thinking is consistent with delMas's first model (2002). Furthermore, they stated that it appears there is a level of literacy between them, with literacy in numbers serving as the background for thinking statistically with deeper thoughts than statistical reasoning. For better understanding of these levels, delMas (2002) assumes that cognitive engagement with the content, rather than the content itself, is a determinant factor in distinguishing these three domains. Furthermore, in Table 1, the author listed the tasks he gathered from the literature related to each domain. Figure 1 shows the connections between these parts, which will be talked about in the next sections.

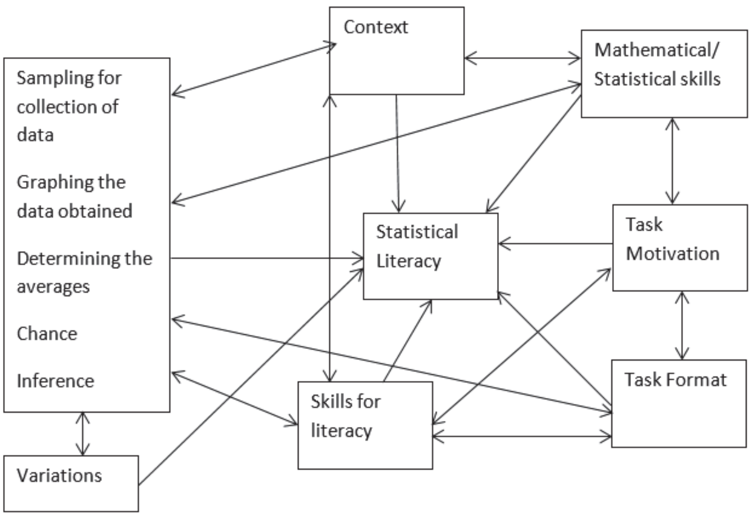

Figure 1: Components of literacy in statistics

4.3 Empirical literature review

Hahi (2009) conducted research on the development of the statistical literacy scale for undergraduate students. The main objective of this study is to focus on developing and validating a reliable scale to determine statistical literacy in university students. a Statistical Content Rating Form (SLCRF) was developed. The content coverage was determined based on the responses of the scholars, and then questions were chosen from among the existing instruments in the literature. The researcher created new questions for topics where the questions in the literature were insufficient. Experts reviewed the suggested questions, and the option questions were chosen and tested in a pilot study with 33 participants. As a result, a 20-option test with 90 participants was used in the study. The number of questions was further reduced to seventeen (17), and the Statistical Literacy Scale (SLS) was finally developed. 476 undergraduate students were given SLS. The construct validity of SLS was investigated using expert item-based opinions and factor analysis results. The SLCRF results ensured content validity. The Cronbach alpha coefficient was calculated using data from 476 participants and was found to be.532. Constructive, content, and curricular validity, to name a few characteristics.

Garfield, DelMas, Chance, Poly, and Ooms (2003) carried out a study on a comprehensive assessment of outcomes for the first course in statistics (CAOS) test. The study was to develop a scale for measuring the conceptual understanding of important statistical ideas by a broader range of students enrolled in first, non-mathematical statistics courses at the undergraduate level. Three rounds of evaluation by content experts for a college-level nonmathematical first course in statistics ensured the content validity of CAOS (DelMas, Garfield, Ooms, and Chance, 2006). When tested in undergraduate student groups, the psychometric properties of this scale were found to be valid and reliable (Cronbach alpha =.82).

Colin and Hay (2009) conducted a similar study on the development and validation of the students' self-efficacy for statistical literacy scale. It was noted that the concept is a relatively new one in mathematics, and while there is some agreement on how it can be conceived, there has been scanty effort made on how the concept is determined in the classroom. The study described the development and validation of a scale to measure students' self-efficacy for statistical literacy. The items were created using a relevant review of literature that is related and then the amount of scale was determined using 366 students. The measure was created using the Rasch measurement methodology, and evidence for its construct validity was provided using the Rasch measurement methodology. The proposed instrument has good reliability and validity, according to the evidence presented in this study.

5. Methodology

5.1 Research design

The research design that was adopted for the study was an instrumentation research design. It is a design used in the development of valid and reliable instruments for data collection.

5.2 Area of study

The area of study is the Cross River State of Nigeria. Cross River State is in the tropical region. The state also has three universities: the University of Calabar, the University of Cross River State and the privately owned Arturas's University, Akpabio .

5.3 Population of the study

The number of subjects used for the study was made up of 4705 year four UTME undergraduates from the universities in Cross River State as at the 2020/2021 academic session. The population distribution of the different universities is shown in Appendix 1,

5.4 *Sampling techniques and sample*

The stratified sampling technique was used for the study. First, the area was divided into two groups according to the number of institutions that are in the area. In each stratum, the researcher divided the area into faculties offering statistics as a compulsory subject for graduation. Thus, a total of seven (7) faculties were selected for the study. A sampling frame was created to assist the researcher in determining which departments will be used for the study. All the names of the 70 departments were written on pieces of paper and were randomly chosen for the study. Thus, a total of 14 departments were chosen, which represents 20% of the total faculties. The names of all the students in each department were written down, and the researcher used a systematic sampling method in selecting the students. Thus, 753 final-year students were selected for the study.

5.5 *Instrumentation*

The researcher created the data collection instrument with the assistance of the supervisors. The instrumentation was carried out in three stages. To determine the suitability and non-suitability of the content of what can be used to describe statistical literacy, the researcher used expert opinion, statistics lecturers' comments, and Ph.D. students from across those faculties that take statistics as a compulsory course for undergraduates in phase 1. For the study, the level of instructional domain suggested by Bloom in his taxonomy was used.'

The next part of the instrumentation was on the items that were generated from the specifications table's direction. The items were created with a 'right or wrong' response format. The data tool was divided into parts. Part A was created to elicit demographic information such as gender, academic discipline, and age, whereas Part B was created to elicit construct information under study. The total number of items remaining after the first administration was forty (40). The study also adapted two instruments, the Statistical Reasoning Scale (SRS) developed by Garfield (2003) and the Statistical Anxiety Scale (SAS) developed by Ogar (2019), whose psychometric properties in terms of validity and reliability were already established.

The validity of the statistical literacy test was done in phases. In phase 1, the draught of the instrument was given to eleven lecturers and Ph.D. students in Measurement and Evaluation to scrutinise the items at face level. Secondly. the content validity was determined after the expert's opinions on what should be considered as the content for a statistical literacy test. Their comments and suggestions were followed to produce content. Content validity was done using a test blueprint with 40 items covering all the six levels of instructional objectives (See Appendix 2 and 3). The initial instrument was made up of 100 items developed by the researcher, even though the target was 40 as contained in the test blueprint. The instrument was first subjected to an expert's analysis to ensure that the items that survived the first assessment will be used for the study. The expert opinion, group discussion with supervisors finally provided forty (40) items. These forty (40) items were trial tested with 50 students to carry out item analysis. At the end of it, 10 items were removed based on the cutoff point as specified by the researcher. The Cronbach alpha reliability technique was was used in determining the reliability of the scale an coefficient of the scale was found to be 0.77, which showed that the instrument is reliable. Data collection was done by the researchers and was analysed using descriptive statistics and factor analysis, and the result is presented appropriately.

5.6 *Presentation of the result*

Research question one:

What are the processes involved in the development and validation of the instrument? First, the construct validity of the instrument was first examined using factor analysis and the results are presented in sub-sections as presented in Tables 1 and 2. The result in Table 1 is the inter-item correlation matrix. The matrix was presented to show the relationship between each of the items and another. It is believed that items that are correlated, i.e., measuring the same thing, will be positively correlated. In this respect, all the items are highly correlated except for items 1 and 7 (.50), item 12 and 2 (.093), item 8 and 5 (.067), item 10 and 5 (.074), item 7 and 12 (.046), and item 12 and 17 (.016). However, these items were still significant.

In principal component factor analysis, the KMO value shown in Table 2 was found to be 0.848, which was acceptable. The null hypothesis states that the variables in the matrix are not correlated and is tested using Bartlett's test of sphericity. The fact that the chi-square value obtained in this test is meaningful indicates that the data is from a multivariate normal distribution. The observed significance level in this study was p 0.001. The conclusion is that the magnitude of the correlation between variables was strong (George and Mallery, 2001).

Using the statistical package for social science (SPSS) version 20.0, the 30 items were subjected to principal component analysis (PCA). Two factors were extracted, and they converged after two iterations.. Factor one was responsible for 94.2% of the total difference, while factor two was responsible for 5.8%. Almost all the items scored well on the first factor, and none scored poorly.. However, all extracted components explain 67.79 percent of the total variance as a whole.

Table 1: Inter item-correlation matrix of Statistical Literacy Test (SLT)

items	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1.00	.427	.450	.254	.270	.210	.050	.130	.112	.197	.140	.174	.164	.185	.172	.201	.208	.176	.179	.169	.204	.182	.192	.241	.236	.024	.211	.310	.222	.012
2	.427	1.000	.459	.251	.213	.167	.077	.160	.108	.131	.120	.093	.128	.118	.126	.183	.216	.199	.207	.180	.150	.154	.151	.175	.168	.410	.011	.013	.314	.100
3	.450	.459	1.000	.518	.418	.173	.082	.026	.160	.204	.256	.214	.359	.265	.308	.306	.277	.212	.204	.169	.152	.225	.246	.270	.252	.102.	.213	.042	.231	.233
4	.254	.251	.518	1.000	.440	.308	.106	.092	.087	.154	.224	.223	.218	.239	.222	.244	.245	.248	.275	.229	.181	.224	.246	.308	.246	.001	.121	.191	.121	.312
5	.270	.213	.418	.440	1.000	.484	.334	.067	.148	.074	.180	.156	.252	.224	.278	.271	.236	.240	.304	.324	.304	.259	.213	.230	.269	.102	.121	.321	.431	.191
6	.210	.167	.173	.308	.484	1.000	.371	.285	.196	.194	.169	.171	.197	.256	.208	.221	.178	.177	.212	.220	.228	.226	.222	.239	.206	.042	.091	.342	.441	.210
7	.050	.077	.082	.106	.334	.371	1.000	.401	.390	.147	.143	.066	.118	.182	.165	.240	.211	.210	.204	.149	.188	.163	.204	.210	.156	.011	.312	.212	.141	.311
8	.130	.160	.026	.092	.067	.285	.401	1.000	.484	.279	.155	.199	.116	.163	.087	.118	.119	.154	.143	.169	.246	.223	.222	.214	.213	.229	.122	.016	.232	.422
9	.112	.108	.160	.087	.148	.196	.390	.484	1.000	.539	.441	.182	.179	.123	.115	.164	.160	.207	.237	.216	.252	.234	.269	.202	.165	.215	.192	.013	.012	.001
10	.197	.131	.204	.154	.074	.194	.147	.279	.539	1.000	.518	.426	.239	.236	.130	.139	.182	.221	.266	.238	.255	.326	.311	.343	.237	.054	.221	.521	.196	.023
11	.140	.120	.256	.224	.180	.169	.143	.155	.441	.518	1.000	.569	.439	.209	.118	.077	.096	.177	.198	.242	.245	.330	.291	.268	.239	.511	.211	.232	.211	.033
12	.174	.093	.214	.223	.156	.171	.066	.199	.182	.426	.569	1.000	.489	.373	.084	.069	.013	.112	.144	.230	.230	.312	.304	.337	.285	.202	.310	.321	.062	.055
13	.164	.128	.359	.218	.252	.197	.118	.116	.179	.239	.439	.489	1.000	.505	.292	.232	.203	.150	.160	.247	.234	.242	.283	.289	.309	.131	.111	.321	.272	.059
14	.185	.118	.265	.239	.224	.256	.182	.163	.123	.236	.209	.373	.505	1.000	.409	.311	.194	.105	.078	.184	.183	.258	.256	.274	.263	.221	.101	.032	.311.	.121
15	.172	.126	.308	.222	.278	.208	.165	.087	.115	.130	.118	.084	.292	.409	1.000	.557	.422	.290	.268	.276	.273	.288	.263	.243	.302	.212	.232	.053	.231	.099
16	.201	.183	.306	.244	.271	.221	.240	.118	.164	.139	.077	.069	.232	.311	.557	1.000	.713	.593	.402	.298	.251	.284	.313	.315	.306	.233	.012	.045	.310	.441.
17	.208	.216	.277	.245	.236	.178	.211	.119	.160	.182	.096	.013	.203	.194	.422	.713	1.000	.736	.574	.305	.295	.265	.275	.257	.265	.316	.214	.041	.076	.124
18	.176	.199	.212	.248	.240	.177	.210	.154	.207	.221	.177	.112	.150	.105	.290	.593	.736	1.000	.699	.505	.332	.310	.269	.294	.307	.076	.123	.045	.171	.082
19	.179	.207	.204	.275	.304	.212	.204	.143	.237	.266	.198	.144	.160	.078	.268	.402	.574	.699	1.000	.569	.460	.332	.323	.291	.299	.213	.222	.061	.056	.167
20	.169	.180	.169	.229	.324	.220	.149	.169	.216	.238	.242	.230	.247	.184	.276	.298	.305	.505	.569	1.000	.583	.514	.344	.300	.250	.042	.217	.187	.102	.066
21	.204	.150	.152	.181	.304	.228	.188	.246	.252	.255	.245	.230	.234	.183	.273	.251	.295	.332	.460	.583	1.000	.645	.559	.402	.230	.321	.191	.042	.167	.121
22	.182	.154	.225	.224	.259	.226	.163	.223	.234	.326	.330	.312	.242	.258	.288	.284	.265	.310	.332	.514	.645	1.000	.691	.554	.292	.062	.065	.103	.192	.231
23	.192	.151	.246	.246	.213	.222	.204	.222	.269	.311	.291	.304	.283	.256	.263	.313	.275	.269	.323	.344	.559	.691	1.000	.699	.366	.212	.072	.056	.162	.234
24	.241	.175	.270	.308	.230	.239	.210	.214	.202	.343	.268	.337	.289	.274	.243	.315	.257	.294	.291	.300	.402	.554	.699	1.000	.462	.111	.052	.142	.217	.211
25	.236	.168	.252	.246	.269	.206	.156	.213	.165	.237	.239	.285	.309	.263	.302	.306	.265	.307	.299	.250	.230	.292	.366	.462	1.000	.243	.192	.075	.048	.102
26	.024	.410	.102	.102	.001	,102	.042	.011	.229	.215	.054	.511	.202	.131	.221	.233	.316	.076	.213	.042	.321	.062	.212	.111	.243	1.00	.142	.421	.391	.062
27	.211	.011	.213	.213	.121	,121	.091	.312	.122	.192	.211	.211	.310	.111	.101	.012	.214	.123	.222	.217	.191	.065	.072	.052	.192	.142	1.00	.084	.192	321
28	.310	.013	.042	.042	.191	.321	.342	.212	.016	.013	.521	.232	.321	.321	.032	.045	.041	.045	.061	.187	.042	.103	.056	.142	.075	.421	.084	1.00	.066	.412
29	.222	.114	.231	.231	.121	.431	.441	.141	.232	.012	.196	.211	.062	.272	.311	.310	.076	.171	.056	.102	.167	.192	.162	.217	.048	.391	.192	.066	1.00	.102
30	.012	.100	.233	.233	.312	.191	.210	.311	.422	.001	.023	.033	.055	.059	.121	.441	.124	.082	.167	.066	.121	.231	.234	.211	.102	.062	321	.412	.102	1.00

Table 2: Kaiser-Mayer-Olkin measures of sampling adequacyResearch question two

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.848
Bartlett's Test of Sphericity	Approx. Chi-Square	8367.478
	df	300
	Sig.	.000

What are the psychometric properties of the items of the statistical literacy test (SLT)? To determine the psychometric properties of the items, descriptive statistics were used, and the results as presented in Table 3 showed that the level of difficulty for all the items ranged from .27-.74. Although, there is no fast rule to the value that determines the difficult nature of an item, for the objectives of this study, the researcher made simple cut off marks which can be used to determine which item to be retained or removed. A p-value of 0.00 shows extreme difficulty, 0.10-0.30 shows moderate difficulty, and.50 shows average difficulty, which those with average ability can attempt. However, where the item p-value ranges from.70-1.00, it shows that the items are too easy and almost 70% of the people in the class can attempt the right answer to the item. Such items that are too easy or too difficult are removed. The items in Table 4 showed the difficulty and discriminating values of the items. According to Joshua (2015), items that have negative discriminating values and those with low and high difficulty indices are bad items and should be removed. However, items with a positive high discriminating index and a difficulty index within the range of 0.40-0.65 could be retained for the study. The result showed that from the thirty items that survived the initial trial, five (5) items were still removed from the study, whose discriminative index was still found to be negative and the difficulty index very low. These were items 3, 6, 11, 12 and 28.

Table 3: Psychometric properties of the items –item difficulty and discrimination of Statistical Literacy Test (SLT)

items	N	P-value	D-value
1	739	.55	0.32
2	739	.54	0.15
3	739	.35*	-0.34
4	739	.66	0.55
5	739	.66	0.43
6	739	.28*	-0.21
7	739	.55	0.54
8	739	.57	0.31
9	739	.57	0.10
10	739	.46	0.16
s11	739	.74*	0.21
12	739	.62	0.32
13	739	.56	0.22
14	739	.56	-0.40
15	739	.59	0.44
16	739	.62	0.13
17	739	.27*	0.27
18	739	.47	0.40
19	739	.57	0.19
20	739	.57	0.12
21	739	.55	0.21
22	739	.58	0.18
23	739	.49	0.33
24	739	.50	0.40
25	739	.52	0.29
26	739	.45	0.22
27	739	.54	0.12
28	739	.33*	-.022
29	739	.44	0.11
30	739	.62	0.21

*=stands for bad items that were removed from the study

Research question three:

How does the instrument converge and diverge with other constructs? To answer this research question, multitrait mono method correlational analysis was carried out to find out the relationship between the variables. Variables that converge highly correlate, while those that are not correlated will naturally discriminate among themselves. In this study, the results as presented in Table 4 showed that the Statistical Literacy Test (SLT) showed a positive correlation with the Statistical Reasoning Test (SRT) (r =. 672) but showed an insignificant negative relationship with the Statistical Anxiety Test (r =-.033). More so, the Statistical Reasoning Test (SRT) showed an insignificant negative relationship with the Statistical Anxiety Test (r =-.002).

Table 5: Multi-trait mono method correlational analysis of SLT, SAT and SRT

Variables	SLT	SRT	SAT
SLT	1.00	.678	-.033
SRT	. 672*	1.00	-.002
SAT	-.033	-.002	1.00

Research question four:

What is the reliability of the instrument when spilt in half and Cronbach's alpha is used? To answer this research question, the Cronbach and Split Half methods of reliability estimation were employed, and the results as shown in Table 6 showed that the reliability coefficient obtained using the Cronbach alpha technique is 0.799 while the coefficient of the statistical literacy scale when the split half technique is applied is 0.711. 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. This is the main reason why Cronbach alpha is mostly preferred when the internal consistency of an instrument is to be ascertained.

Table 6: Reliability coefficient of SLT using split half and Cronbach alpha techniques

Reliability technique		N	X	S.D	α
Split Half	Odd	13	9.05	2.97	0.711
	Even	12	8.41	3.01	
Cronbach		25	17.29	4.75	0.799

6. Discussion of Findings

The psychometric properties of the items in terms of item difficulties and item discrimination were established. The discriminative values range between 1.000 and 1.00. The average item discrimination in a dichotomous response is 0.50, and the researcher used it to set the cut off in judging good items for selection and inclusion in the scale. A total of 30 items were subjected to item analysis, and only 25 items survived the test, whose item difficulties were considered moderate for selection and the discriminative values. Only five items showed discriminative negative values, which showed that they were very bad items.

The reliability coefficient for the split half test and Cronbach's alpha showed a coefficient of 0.78 and 0.81, respectively. According to Nunally in Promise (2008), if the reliability coefficient of an entire scale is within the range of 0.80 to 1.20, the scale is good and there is no need for additional items. More so, in terms of utilisation of a scale, Kaplan and Saccuzzor (2005) noted that where the reliability coefficient of a scale ranges from.70-.80, it is considered excellent. So, the SLT scale's reliability coefficient of.81 showed that the instrument consistently measures what it was meant to measure, making it a good choice for a study.

Research question two attempted to establish the convergent and divergent validity of the instrument. The positive inter-correlation could be used to explain that people have similar characteristics in terms of their literacy of statistics. This is because, if these items were not positively correlated, the correlation matrix would have been replete with items that were negatively correlated. Where items and variables are not measuring the same thing, they tend to move in opposite directions. A further look at the convergent and divergent validity table showed that the scale Statistical literacy was highly correlated with other constructs such as statistical reasoning, but it differed from the statistical anxiety score. The findings are in line with that of Gregory (2012), who posited that items that are highly correlated tend to measure the same construct, and they are always positively related, but items that are not measuring the same thing tend to diverge and often have a negative correlation.

The cut off mean score of SLT was estimated by using the mean of the T-score. The expected highest raw score of a student is supposed to be 25, which is the number of items multiplied by 1, when the student gets them all correct. In reporting the scores in deviation scores, the raw score was first converted to normalised standard scores of SLT, which have a mean of 50 and a standard deviation of 10. This is to facilitate comparison of students' performances. Any person who scores below the mean of 12.5, or 13, is at the middle point of the normal curve distribution of the T-score, which is the average of statistical literacy. Those who score from 14–20 could be regarded as having a plus one (+1) standard deviation above the mean and could be termed as having a moderate statistical level of literacy; those from 20-above could be regarded as having a plus one (+2) standard deviation above the mean and could be termed as having a high statistical level of literacy; and any group that scores 12 below could be regarded as having one standard deviation below the mean and could be termed as having a low statistical level of literacy. A cumulative score based on range showed that those who have a high statistical level of literacy constitute 33.42% of the study population, those with a moderate statistical level of literacy constitute 51.56%, those with an average statistical level of literacy constitute 4.6 high statistical level of literacy, and those with a low statistical level of literacy constitute 10.42%. The set of groups is to be considered for statistical literacy development and improvement. The findings collaborate with that of Ofem, Ovat, Nworgwugwu , and Effiom (2022) that noted that while the mean of the group is very close to the cut of men for the study, it showed that such a group possesses the characteristics that are assessed in that study.

Generally, the results showed a high level of statistical literacy among students based on the percentage analysis. Those with a high level of statistical literacy are more literate than those with a low or average level of statistical literacy. This could explain why more students try to avoid statistics classes, especially those in behavioural science who didn't expect to have to take statistics.

7. Conclusions and Recommendations

Based on the findings of the study, which were that only 25 items survived the last item assessment, the instrument was found to have strong construct validity as the items in the correlation matrix were all positively correlated; the statistical literacy scale showed convergence with statistical reasoning but diverged with statistical anxiety; and the reliability of the instrument was found to be high when assessed with Cronbach alpha and split half techniques as the coefficients of the two scales were greater than 0.70.; it was concluded that the instrument is valid and reliable for use . It was recommended that

1. Students should not just know how to do math in statistics; they should also know how to use statistical knowledge.
2. Statistical literacy should be emphasised by statisticians to help students reason and think about statistics.
3. The instrument should be supported to be widely distributed for usage by researchers and psychometricians

References

Ben-Zvi, D., & Garfield, J. B. (2004). *Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges*. In D. Ben-Zvi & J. Garfield (Eds.), The Challenge of Developing Statistical Literacy, Reasoning and Thinking (pp. 3-15). Netherlands: Kluwer Academic Publishers

Burnham,T.(2003).*Statistical Literacy: Purpose and Definitions* http://www.statlit.org/pdf/2003BurnhamStatLit.pdf,

Callingham, R. (2007). *Assessing Statistical literacy: A question of interpretation*. ICOTS

Callingham, R. (2014) Measuring Statistical Literacy. *Journal of Applied Measurement*, 6(1), 1-29.

Callingham, R. and Watson, J. M. (2005). Measuring statistical literacy. *Journal of Applied Measurement, 6*(1), 29, 19-47

Chance, B. L. (2002).Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education, 10*,3

Chick, H., Pfannkuch, M., & Watson, J. (2005). Transnumerative thinking: Finding and telling stories within data. *Curriculum Matters, 1,* 86-107

Colin and Hay (2009) Development and validation of the students' self-efficacy for statistical literacy scale. *Structural Equation Modeling, 9*, 413-430.

delMas, R. C. (2002). Statistical Literacy, Reasoning, and Learning. *Journal of Statistics Education 10* (3).

Gal, J. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review,* 70(1), 1-51.

Gal, J. (2004). *Statistical literacy: Meanings, components, responsibilities.* In D. Ben-Zvi& J. Garfield (Eds.), The challenge of developing statistical literacy, reasoning and thinking. Netherlands: Kluwer Academic Publishers.

Garfield, J. (2003) Development of instrument for Statistical Reasoning Assessment (SRA). International *Journal of Mathematics Education, Sciences, and Technology, 31,* 679-689

Garfield, J. DelMas, Chance, Poly, Ooms (2006). Comprehensive Assessment of Outcomes for a first course in Statistics (CAOS) test produced.

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372-39

Garfield, J., delMas, R., &Zieffler, A. (2010). Assessing statistical thinking. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.), *Assessment methods in statistical education:* An international perspective. Milton: John Wiley &Sons..

Hahi,j. (2009) Statistical literacy scale development for undergraduate students. Jpurba;of Social Since Research, 23. 567-589

Hayden, R. (2004).*Planning a Statistical Literacy Program at the College*

Jacobe, T., Foti, S., & Whitaker, D.(2014). *Middle school (ages10–13) Understanding of statistics*. In K. Makar, B. de Sousa, & R. Gould (Eds.), Sustainability in statistics education. Proceedings of the ninth international conference on teaching statistics (ICOTS9,July, 2014),Flag staff, Arizona, USA. Voorburg: International Statistical Institute. Retrieved from iase-web.org

Joshua, M. T. (2005). *Fundamentals of test and measurement in education.*Calabar-Nigeria: University of Calabar printing press.

Kaplan, J. and Saccuzzor, U. (2005) Development of statistical literacy in undergraduate students .*Journal of Statistics Education*, 14(2).

Kaplan,J.J.,&Thorpe,J.(2010).Post secondary and adult statistical literacy: Assessing beyond the classroom. Data and context in statistics education: Towards an evidence-based society. *Proceedings of the eighth international conference on teaching statistics. Voorburg*: International Statistical Institute.

Keosayian, S. and Johnson, D. (2014) Development of statistical literacy in undergraduate students .*Journal of Statistics Education*, 14(2).

Marayat , S., Duangkamol, H. and Wannee, J. (2014). Undergraduate Students' Statistical Literacy: A survey study. *Journal of Clinical Epidemiology, 53*, 499-503

Ofem, **U. J.,** Ovat, S.V., Nworgwugwu , C. E. and Effiom, A.P (2022).Development and Validation of Instrument for Measuring Statistical Anxiety of Undergraduates in Universities in Cross River State, Nigeria. *Multi-disciplinary Journal of Research and Development Perspective*, 10(1), 12-24

Ogbuegu, G. (2011) Statistical Literacy in a Glance. *Journal of Statistical Educators* 1(1), 145-167

Pfannkuch, M. & Wild, C. (2004). *Towards an understanding of statistical thinking*. In D. BenZvi and J. Garfield (Eds.), The Challenge of Developing Statistical Literacy, Reasoning and Thinking, (pp. 17-46). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Promise , H.(2008) Development and validation of instrument to measure emotional intelligence among secondary school students in Cross Rive State. Unpublished Ph.D work, University of Calabar.

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education, 10*(3).

Schield, M. (2001) Statistical literacy: A calling IASE/ ISI Satellite, 2001 Available on http:/www.stat.auckland.ac.nz/~iase/publications.php [cited 21 June 2007

Schield, M. (2004). Information Literacy, Statistical Literacy and Data Literacy. IASSIST. *Quarterly 28* (2/3), 6-11.

Sharma, S. (2017) Definitions and models of statistical literacy: a literature review, Open *Review of Educational Research, 4* (1), 118-133,

Tversky, A. & Kahneman, O. (1982) . The effect of a student-designed data collection project on attitudes toward statistics. *Journal of Statistics Education, 16*(1).

Udinwnege, O. (2011) In Enhancing Statistical Literacy: Enriching Our Society. *Journal of the American Statistical Association, 88* (4)1-8.

Utts, J. (2004). *Seeing through statistics*. Belmont, CA: Thomson Brooks/Cole

Watson, J. and Kelly, O. (2003) Development of understanding of sampling for statistical literacy. *Journal of Mathematica1 Behavior, 19*, 109-136.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals Mahwah*, NJ: Lawrence Erlbaum.

Watson, J. M. and Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2*(2), 3–46.

Watson, J.M. (2000). Statistics in context. *Mathematics Teacher, 93*, 54-58

Wild, C.J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry .*International Statistical Review, 67*(3), 223-265.

Appendix 1: Test blue print for Statistical Literacy Test (SLT) showing level of instructional objectives and content area

Content/level of object	Knowledge 20%	Comprehension 20%	Analysis 10%	Application 30%	Synthesis 10%	Evaluative 10%	**Total**
Level of measurement 5%	0	0	0	1	0	1	2
Data organization and presentation 5%	0	0	0	1	0	0	2
Measures of location 10%	1	1	0	1	1	0	4
Measures of spread 10%	1	1	1	1	0	0	4
Hypothesis testing 20%	2	2	1	2	0	1	8
Pearson product moment correlation 10%	1	1	0	1	1	0	4
Measures of relative standing –Z&T score 5%	0	1	0	1	0	0	2
Regressive analysis 10%	1	0	1	2	1	0	4
t-statistic 20%	2	2	1	2	0	1	8
Chi-square 5%	0	0	0	0	1	1	2
TOTAL	8	8	4	12	4	4	40

Appendix 2: Number of faculties, departments, and year 4 students in the University of Calabar

Faculties	Departments	No of yr 4 Students
Agriculture	7	411
Social Sciences	7	644
Management Sciences	4	570
Education	14	1070
Total	47	2695

Appendix 3: Number of faculties, departments and year 4 students in the Cross River University of Technology

Faculties	Departments	No of yr 4 Students
Agriculture and Forestry	5	695
Education	4	895
Management Sciences	3	320
Total	23	1910