UNIJERPS
Unizik Journal of Educational Research and Policy Studies VOL.22 (1);
July-December, 2026 **https://unijerps.org**
2023 Impact Factor: **6.834;** **2022** ***Impact Factor*****: 5.156;** ***2021 Impact Factor: 5.641;***
2020 Impact Factor: 5.355. Indexing and Impact Factor URL: ***http://sjifactor.com/passport.php?id=21363***

STANDARDIZATION AND FURTHER VALIDATION OF PG STUDENT'S STATISTICAL LITERACY SCALE

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

ABSTRACT

Statistical literacy is the ability to read and use tools such as percentage, ratio, measures of relationship, central tendency and variability, as well as tables, graphs and maps. It is the ability to analyze, interpret and evaluate statistical information. Statistical literacy is the ability to organize data, construct and display graphs and tables and work with different representations of data. It requires being able to extract, understand and explain data that is presented in a variety of ways such as words, tables, graphs and increasingly in maps. To be statistically literate, one must understand that how data are organized can contribute to how they are interpreted. Despite the growing importance of statistical literacy, the measurement of this construct remains inconsistent because of the use of diverse assessment instruments lacking standardization. (PGST) This article examines the concept of standardization of instruments for assessing statistical literacy, highlighting the procedures, importance, and challenges involved in developing valid and reliable measurement tools. The article also discusses the theoretical foundations of statistical literacy, principles of instrument standardization, methods of establishing validity and reliability utilizing both Classical Test Theory (CTT) and Item Response Theory (IRT), and the stages involved in constructing standardized instruments. The study concludes that standardized instruments are necessary for ensuring accurate measurement, comparability of results, and enhancement of educational and research outcomes in statistical literacy assessment.

Keywords: Statistical literacy, instrument development, validation, standardization, reliability, validity, psychometric properties.

Introduction

We live in an era dominated by empirical data, where decision-making across healthcare, economics, public policy, and daily life heavily relies on the interpretation of quantitative findings (Sabbag et al., 2025). Consequently, higher educational institutions have shifted their educational paradigms to emphasize statistical literacy as a critical learning outcome (Sabbag et al., 2025). Statistical literacy is broadly conceptualized as the cognitive ability to read, understand, critically evaluate, and

communicate statistical information encountered in academic or real-world contexts (Sabbag et al., 2025). Statistical literacy refers to the ability to understand, interpret, critically evaluate, and communicate statistical information and data-based arguments (Gal, 2002). In modern society, citizens encounter statistics in newspapers, television broadcasts, social media algorithmic feeds, academic reports, and public policy discussions. Consequently, statistical literacy has become a fundamental component of education, digital literacy, and informed citizenship.

Garfield and Ben-Zvi (2008) explained that statistical literacy enables learners to reason with data and make evidence-based decisions. Educational institutions worldwide have therefore integrated statistics into school curricula to equip learners with the skills necessary for data interpretation, risk assessment, and critical thinking. However, assessing statistical literacy effectively remains a major concern among educators, psychometricians, and researchers.

Various instruments have been developed to measure statistical literacy, but differences in content, structure, scoring procedures, and psychometric properties often result in inconsistent findings. According to Watson (2006), many assessment tools fail to adequately capture the multiple cognitive and dispositional dimensions of statistical literacy. Standardization of assessment instruments is therefore essential to ensure that tools used in measuring statistical literacy produce reliable, valid, objective, and comparable results across diverse contexts (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). The performances of (mean scores) of MEd and PhD were compared and PhD students performed better than the masters students. The instrument has been already developed by Okoye R.O and Okonkwo U.P, and it is necessary to standardize and further validate it

Purpose of the Study

The main purpose of this study is to standardize and further validate an instrument for assessing statistical literacy.

Specifically, the study seeks to:

1. determine the content validity of the statistical literacy instrument;
2. examine the construct validity of the instrument;
3. determine the reliability of the instrument;
4. examine the performance of individual items of the instrument;
5. determine whether the items adequately represent the dimensions of statistical literacy;
6. establish appropriate procedures for administering and scoring the instrument; and
7. develop appropriate norms or standards for interpreting scores obtained from 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 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.

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 instrument consists 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. Thereafter 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 have the desired qualities were assembled. The 40 items constitute the final instrument.

To ascertain both convergent and divergent validity of the final instrument, the instrument along with two others instruments namely: Statistical Reasoning Test (SRT) by Liu (1998) also, administered to the standardization sample of 600 PG students. See Appendix H, Page 156 and Appendix I 156 for details. The scores obtained from TAS and SRT were correlated with scores from the final instrument using Pearson Product Moment Correlation. Furthermore, the students' scores in Advanced Educational Statistics were also obtained from the appropriate Departments of Educational Foundations of the various sampled universities. The students' scores in Advanced Educational Statistics were correlated with scores obtained from the final instrument using Pearson Product Moment Correlation.

The reliability of the final instrument was established using Kuder-Richarson 21 because the items were dichotomously score, that is each was awarded 1 for correct answer and 0 for wrong answer. The mean and standard deviation scores for Masters and Ph.D. students on the final instrument were also computed. The mean and standard deviation scores are to serve as norms for interpreting future scores obtained with the instrument.

Area of the Study

The study was conducted in the South East geopolitical zone of Nigeria. South-East bordered by North Central Geopolitical Zone, in the West and South by South south Geopolitical Zone and the East with Cameroon Republic. The South Eastern Nigeria has five states which are: Abia, Anambra, Ebonyi, Enugu and Imo states.

Population of the Study

The population of the study comprised 1102 postgraduate students MEd and Phd 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 via: 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, 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 has been done, while the second sample was used for final testing of the instrument to establish validity, reliability and norms of the final scale. The first sample consist of 400 PG students in faculties of

education from four universities in the study area. The researcher made use of multi-stage sampling procedure. Ali (2006) opined that multi-stage sampling procedure required several stages of sampling the elements of the population using different sampling techniques. At the first stage, simple random sampling technique was used to obtain four universities in the study area. At the second stage, one department from each of the four faculties was obtained through simple random sampling. At the third stage the students were stratified in each department according to gender and then 50% of each group was obtained. The exercise gave rise to 400 PG students and 224 master students. The second sample consisted of 600 PG students in four universities in the study area. Any university previously selected for trial testing was excluded. The researcher made use of multi-stage sampling procedure. At the first stage, simple random sampling technique was used to obtain four Universities in the study area. At the second stage, one department from each of the 4 faculties was obtained through simple random sampling. At the third stage the students were stratified in each department according to gender and 50% of each group was obtained. The exercise gave rise to 600 PG students made of 224 Ph.D students and 336 Masters students.

Method of Data Analysis

1. Research question 6 was answered using Kuder Richardson 21 reliability approach
2. Research question 7 was answered using standard deviation.
3. Hypothesis 1 and 2 were tested at 0.5 level of significance using t-test.

MEAN AND STANDARD DEVIATION for masters and Ph.D. students as measured by the instrument

Masters	345	28.74	5.95
Ph.D.	255	33.59	5.21

The above table shows that the mean achievements scores for Ph.D. students and that of masters students is 28.74, in other words, Ph.D. students performed better than masters students in the PG statistical literacy scale.

Result and Discussion

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 if the split half technique, the error may come as a result as the variances may get wider. However, Ruder 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. Okeke (2008), found the reliability index for mathematics achievement test to be 0.80. Kazeni (2005) reported a study on development and validation of test in integrated science process skills for further education and training learners. The reliability index of the 467 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 postgraduate students? Statistical literacy scale very valuable.

Standardization

Standardization ensures that the scores generated by the test carry a consistent meaning across different test-taking cohorts. This was achieved by fitting the item response matrix to a Unidimensional Item Response Theory (IRT) model—specifically, the Two-Parameter Logistic (2PL) model.

The 2PL model calculates the probability (Pi) of a student with a given latent statistical literacy ability (θ) answering an item (i) correctly.

Procedures for Standardizing Instruments for Assessing Statistical Literacy

1. Identification of Objectives
2. Review of Literature
3. Development of Test Blueprint
4. Item Construction
5. Expert Validation
6. Pilot Testing
7. Item Analysis
8. Reliability Testing
9. Establishment of Validity
10. Norm Development
11. Final Revision and Standardization

Ways of Improving Standardization

1. Through theoretical frameworks.
2. By collaboration with statisticians, educators, and psychometricians.
3. By reviewing and updating of instruments.
4. Through Use of technology-assisted assessment methods.
5. Inclusion of culturally relevant examples and contexts.
6. Always train test developers and administrators.

Conclusion

This study has further validated and standardized PG Statistical Literacy Scale for Use in Faculties of education in Public Universities in South East Nigeria. The test is

highly good. The study focused on the standardization and further validation of an instrument for assessing statistical literacy. The validation process emphasized content validity, construct validity, item analysis and reliability, while the standardization process focused on uniform administration, scoring and interpretation. A statistically sound instrument is important for accurately assessing learners' statistical literacy. The availability of an appropriately validated and standardized instrument can support researchers, teachers and educational institutions in identifying strengths and weaknesses in statistical literacy and in evaluating educational interventions. The final version of the Statistical Literacy Scale should therefore be used in accordance with the administration, scoring and interpretation procedures established through the standardization study.

Recommendations

Based on the purpose of the study, the following recommendations are made:

1. Researchers assessing statistical literacy should use instruments with adequate evidence of validity and reliability.
2. The Statistical Literacy Scale should be administered under standardized conditions to ensure comparability of scores.
3. Researchers using the instrument with populations that differ substantially from the standardization sample should conduct additional validation studies.
4. Teachers and educational practitioners may use validated statistical literacy assessment instruments to identify areas in which learners require additional instructional support.
5. Future researchers should examine the instrument using larger and more diverse samples to provide additional evidence concerning its psychometric properties.
6. Further studies should investigate whether the instrument functions similarly across relevant demographic and educational groups.
7. Norms should be periodically reviewed and updated when substantial changes occur in the target population or educational context.

REFERENCES

Adonu, B. (2016). *Development and preliminary validation of an instrument for assessment of psychomotor skills in physics*.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anikweze, C. M. (2010). *Measurement and evaluation for educational research. Enugu*: Snaap Press Ltd.

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.

Chime, N. O. (2012). *Instrumentation Research in Education. Enugu*: Academic Publishing Company.

Effiong, (2006). *Study on instrument reliability using Kuder Richardson formula*.

Eze, D. N. (2003). *Instrumentation research and educational measurement*.

Ezeh, D. N. (2003). *Educational measurement and evaluation.* Enugu: Computer Edge Publishers. Reliability and validation of tests. In B.G. Nworgu (Ed.), educational measurement and evaluation: Theory and practice (pp.123-135). Nsukka: University Trust Publishers.

Gal, I. (2002). *Adults' statistical literacy:* Meanings, components, responsibilities. International Statistical Review, *70(1), 1–25*. Education research.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: connecting Research and Teaching Practice. Dordrecht*: Springer.

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.

Isere 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 alanlarmd akitez cahsmalarinin istatistiksel a9idan incelenmesi", Pamukkale University, E gitim 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, 0. (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_lB3_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,* 7S(1),1-11.

Man, P. O (2016) *Education in Nigeria a futuristic perspective.*

Retrieved from . csd. uwa. edu. an/icedu2002

Meredith, S.,Joyce,N. and Walter,0. (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.

Onah, D. U. (2014). *Development and Standardization of Agricultural Science Achievement Test for Senior Secondary School Students.*

Onunkwo, G. I. N. (2002). *Fundamentals of Educational Measurement and Evaluation.*

Sabbag, A., Zieffler, A., & Ng, C. (2025). Can we distinguish statistical literacy and statistical reasoning? Statistics Education Research Journal,

Singh, A. S., &Masuku, M. B. (2012). *Sampling techniques and determination of sample size in applied statistics research.* International Journal of Economics, Commerce and Management.

Watson, J. M. (2005), 'Is statistical literacy relevant for middle school students?' *Vinculum,* Watson, J. M. (2006). *Statistical Literacy and Assessment Issues.*

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

Zwick, R. (2013). *Development and validation of mathematics economics test journal of science education, 7(3) &1-79.*

DeVellis, R. F., & Thorpe, C. T. (2021). Scale development: Theory and applications (5th ed.). SAGE Publications.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. International Statistical Review, 70(1), 1–25.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis (8th ed.). Cengage.

Kline, R. B. (2016). Principles and practice of structural equation modeling (4th ed.). Guilford Press.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). McGraw-Hill.

OECD. (2019). PISA 2021 mathematics framework: Draft. Organisation for Economic Co-operation and Development.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International Journal of Medical Education, 2, 53–55.

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

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