European Journal of Education Studies
ISSN: 2501 - 1111
ISSN-L: 2501 - 1111
Available online at: www.oapub.org/edu

04DOI: 10.46827/ejes.v13i8.6960 Volume 13 | Issue 8 | 2026

DETERMINATION OF NORMATIVE SCORES OF THE DEVELOPED STANDARDIZED SENIOR SECONDARY SCHOOL MATHEMATICS ACHIEVEMENT TEST (MAT)

Christogonus Onyemaechi Nweke[1],
Julius O. Ugama[1],
Obeten Okoi Oka[1],
Ikporo Festus Boss[1],
Nwodo Mary Odinaka[1],
Egbe Irene[1],
Igwe Micheal Igwe[1],
Nweke Benjamine Ifeanyi[1],
Kpirigbe Benjamine Arobor[1],
Ukworu Amarachi Favour[1],
Mgbogo Sunday Ogbule[1],
Augustina Nnenna Awoke[2i]
[1]Department of Science Education,
Ebonyi State University,
Abakaliki, Nigeria
[2]Department of Arts and Social Science Education,
Ebonyi State University,
Abakaliki, Nigeria

Abstract:
The roles of norm scores in the interpretation and comparison of test scores in education assessment motivated the researchers to conduct this study. The design of the study was an instrumentation research design. Five hundred and ten (510) SS2 students selected from public and private schools across the three education zones of Ebonyi State were sampled for the study. The population of the study was all the SS2 students in their third term of SS2 class. Three research questions guided the study. The instrument for data collection was the Ebonyi State Ministry of Education-developed MAT achievement test. Data collected were analyzed using the SPSS computer package. The analysis yielded the mean, standard deviation and the transformed standard scores (Z-scores, T-scores, Percentile Ranks and Stanine ranks). Based on the outcome of this study, the Ebonyi State Ministry of Education was advised to integrate these developed norm values into the components of this MATS standardized test; develop a norm-referenced standardized test in all subjects taught in schools in the state; and mandate the State Universal Basic

[i] Correspondence: email christian.aleke2020@gmail.com, christian.aleke@yahoo.com

Education Board to develop norm-reference scores for the unified examination for basic schools to raise the standard of their test instrument.

Keywords: standardized test; norm-test; Z-test; stanine; percentile ranks; and normalized T-test

1. Introduction

One of the foundational subjects taught in basic and senior secondary schools globally is mathematics. The importance of mathematics as a basic subject studied in schools lies in the vital roles it plays in the development of individuals and nations. The most important/vital role is its usefulness and contributions to economic, scientific and technological development. Any country, society or individual that aspires to develop scientifically, economically and technologically cannot toy with the study of mathematics at any level of education, most especially at the formative stage. It is at this stage that a strong foundation and interest in the subject are laid. Acquisition and application of mathematical skills are key in understanding the concepts and language of science and technology. Undoubtedly, strong knowledge and acquisition of mathematical skills by citizenry are the harbinger or precursor for the development of the much-needed technological breakthrough by nations (Denga, 2017). Nigeria, as a developing nation, is conscious of these roles of mathematics and, conceiving knowledge in mathematics as a pathway in societal development, designated mathematics as a core subject in both basic and senior secondary schools. This perhaps spurred the Nigerian federal government to make the study of mathematics compulsory at both levels of education (FRN, 2019). Specifically, mathematics as a subject of study in basic and secondary schools is intended to develop in the learners the ability to recognize problems and solve them with mathematical knowledge and to stimulate and encourage creativity in the minds of the citizenry (Visu, 2019).

To determine the level of attainment of these lofty objectives of government in mathematics, systematic school-level assessments of pupils' learning outcomes are carried out on a periodic basis, through testing with the aid of appropriate test instruments. A test instrument is an assessment tool used for evaluating students' level of performance and level of knowledge in a particular subject (Formplus, 2025). A test instrument is designed in a way that it will detect the specific instructional objectives of interest to be measured as contained in the curriculum. The test itself connotes the method to determine learners' ability to demonstrate mastery of skills or knowledge of content (Toor, 2025). It could equally be viewed as the administration of test items and the use of the results to determine whether the learner has mastered the content of the instruction and can demonstrate the traits outlined in the instructional objectives. Concisely, testing is a process of administering appropriate test instrument/tools with a view to eliciting vital data to guide decision-making on the instructional process in

education. Therefore, testing in the context of this paper is conceptualized as a process of using appropriate test tools that help in the assessment of the extent to which learners have mastered the objectives of teaching mathematics as clearly outlined in the curriculum of senior secondary schools in Nigeria. This type of test is generally referred to as an achievement or ability test (Thorndike & Hagen, 1969). The categorization of the test as an ability test is because it is meant to appraise what a learner is able to learn or do based on planned sequential instruction or experience provided in a formal setting such as school (Annastasi & Urbina, 1997). There are two types of tests for assessment of learning outcomes. There are teacher-made tests and standardized tests. Most assessment processes prioritize the use of standardized tests. Notwithstanding this attached priority on standardized test it does not mean that teacher made test does not serve an important purpose. Each type of test has its strength and weaknesses; they equally serve specific purposes in the assessment of pupils' learning achievement. For instance, a teacher-made test, which is often regarded as deficient in psychometric properties due to its low quality in writing, administration, scoring and absence of normative scores (Nwagu, 2023; Kabir, 2016), is specific and focused on detecting the extent to which outlined instructional objectives in the curriculum desired by a particular teacher or school management in different schools have been attained. Hence, teacher-made tests are narrower in focus but are more specific and concise in detecting the purpose of conduct of the testing and hence more useful to individual teachers and schools. For standardized tests, which is test constructed by experts and rich in quality as it has been subjected to certain statistical analysis to ascertain their psychometric properties (Nworgu, 2017; Osadebe, 2016), and have the tendency to be used across schools and regions, has deficiency in detecting specific challenges of learners that each school and teacher is interested. It is broader in scope, technically superior and costly to develop, especially by non-professionals who lack the skills and the resources. It is acknowledged to be the best test tool for assessment of instructional learning outcomes.

In the conduct of testing in education, education managers have often been greatly challenged or confronted with the ability to develop or procure and use appropriate test tools that are standardized (with the normative scores) and contain the necessary psychometric properties for the testing. It is a case and a matter of serious concern that most test instruments used in the assessment of instructional objectives in schools are teacher-made tests (not standardized with norm scores), and therefore lack necessary psychometrics properties. This gap erodes trust in the quality of data collected with the test instrument and, by extension, challenges the efficacy of the decision made with the data. Research reports have attributed the construction and use of test materials which lack psychometric properties and norm scores to the test maker's lack of skills in the development of standardized tests which professionally are adjudged to be statistically valid and reliable. The development of tests that are completely adequate in psychometric properties is a challenge not only to most test makers but to the entire testing process. This was aptly captured by Asuk (2021), who declared that the

development of tests with adequate psychometric properties is a challenge to non-specialists. Given the low emphasis attached to the regular use of standardized tests with norm scores in testing by the public despite their usefulness, teachers and professional organizations that have the expertise in the development of standardized tests feel reluctant to embark on the process, as it is both rigorous and cost-demanding.

Standardization of tests is most useful and relevant where scores are to be compared directly or indirectly to scores of people of an equivalent group (Kline, 2015). The comparison of a test to people of equivalent group is called test norming (Thorndike *et al.*, 1969). Standardization through norming helps to make tests uniform and helps to adhere strictly to specific standards (Study.com, 2025). Test norming, therefore, is a process of administering standardized tests to a large representative sample of the target population to establish norms, which serve as a frame of reference for interpreting individual scores (Anastasi *et al.*, 1997). Worthy of note is that most standardized tests in schools and education institutions are not norm-referenced and hence incomplete. Any standardized test developed without an accompanying normative referencing is incomplete, as it serves little purpose in the interpretation of test takers' scores. Such test tools are almost equivalent to teacher made test.

Norm-referenced tests help to compare test takers' scores to the scores of another group of individuals called the normative group. Consequently, a norm is the score measure of the normative group on a test. The norm group is always a representative sample of individuals of the same age, same academic/class grade and sometimes the same gender as the test takers (Hegde & Pomaville, 2021). This clearly indicates that standardized test and norming are intrinsically linked and inseparable in the test assessment process and are most useful whenever accurate comparison of the test taker's score is made or interpretation of a test taker's achievement is being made relative to a norm group (national, regional, state, LGA, race, gender or another group). This is because standardization is concerned with the creation of a uniform testing environment. Norming provides the essential framework for the interpretation of individual scores by comparing them to the results of a specific group (Bond, 2015). Norming plays a vital role in the interpretation of test scores, as raw scores, when considered relative to a standardized test, are meaningless without this framework set up through the use of a normative group. Based on this explanation, any developed achievement tests that have no corresponding comparative normative scores serve no purpose. Unfortunately, there are many developed test tools by education researchers/managers scattered around in Nigerian schools and education institutions which are not utilized because of the absence of normative scores which will aid the interpretation of the scores of learners that it is administered.

The theoretical bases for norming standardized tests are directly linked to or firmly rooted in the two major test theories (the Classical Test Theory (CTT) and the Item Response Theory (IRT)). For the classical test theory, which was propounded by Charles Spearman in 1904, it holds that an observed score (O) of the test taker is composed of the

test taker's true score (T) and a measurement error (E). Hence, CTT tries to ensure that tests are both valid by measuring what it purports to measure by reducing error score and reliable by consistently measuring the trait it is intended to measure. Since CTT focuses on ensuring the measurement of the true score, as it has maintained that the observed score is comprised of true and error scores, this shows that efforts must be made at minimizing error score from the observed score through standardization, which ensures standard measurement. For the imperative of norm, the CTT provides tools like the standard deviation, mean, and reliability coefficient that are used to interpret test takers' scores relative to the norm of the group, such as percentile ranks; T-score; and Z-scores. Therefore, CTT helps to ensure that the test is developed, standardized and norm-referenced, so that meaningful, comparable results that aid decision-making are obtained.

For IRT, which was developed by Rasch (1960) and Lord (1955), its main focus is on a measurement framework that models how each item interacts with the individual's ability. Hence, IRT provides item-level information (difficulty, discrimination, guessing) and links them with the ability scale. IRT places both the test taker and items on the latent metric. IRT therefore enhances equating test forms or their administration to a single metric, which is essential for maintaining standardized scores over time and ensuring precision of measurement. Consequently, while standardization is concerned with consistent tests or uniform testing procedure/environment, IRT guarantees a precise, item-level measurement model. Norming, as was earlier described, has the main purpose of indicating where a test taker stands relative to other test takers on that standardized IRT-based scale; this is similar to the functions of percentile ranks, which are one of the scales in norm-referenced scores. This shows that IRT models are properly aligned with the principles of standardization and test norming. The IRT theory gives a clear explanation and deep insight into the importance and functions of norming a standardized test. Without norm scores, it is difficult to properly place a test taker on the IRT –based scale, which are scales for determining test takers' ability along the continuum of test difficulty, and which reflects the test taker's latent ability and its discrimination along the test latent ability. Based on these insights from the two theories, Education managers and parents should advocate for increased use of norm-referenced standardized tests in the assessment of learning in schools. This is because of the broader perspectives in the explanation it conveys, whereas a teacher-made test is structurally deficient. It is also evident that it eliminates localized interpretation of pupils' performance, as it helps to place a learner's performance on a broader frame of reference. Given these important roles of norm-referenced standardized tests in the assessment of learning achievement, the researchers undertook this study and determined the normative values of this developed mathematics achievement test for secondary schools. The norm values enhance the gains of standardized test materials and make the interpretation of learners' achievement easy. Specifically, it provides alternatives scales (Z-scores, Percentile Ranks, T-scores and Stanine ranks) for the interpretation and

understanding of learners' achievement in any subject and enables stakeholders to make comparisons of learners' achievement with those of their peers within the normative group.

2. Statement of the Problem

There are some developed standardized tests in various subjects studied in schools which are not in use, and sometimes their utilization is limited by the absence of norm scores that will aid the users' interpretation of test scores. This renders null and void the efforts and resources invested in developing the test instrument and defeats the purpose of the test developer(s). Most achievement tests have corresponding curricular objectives they measure, and this developed mathematics achievement test is inclusive. This present test was designed to assess learners' proficiency in some topics in the senior secondary school curriculum in Nigeria. Unfortunately, the presumed standardized test did not have norm scores (Percentile Ranks, Standard scores, Normalized T-scores of age /class/gender equivalent) that allow teachers/parents/researchers or school owners) to accurately compare individuals' raw test scores to those of the norm group. This study was carried out to fill the gap and generate data for evidence-based progress monitoring through objective comparison and policy evaluation. The statement of the problem of this study is the development of normative scores for standardized MATs for secondary schools in Nigeria.

2.1 Purpose of the Study

The purpose of this study was to establish reliable and accurate alternative scaled normative scores for the developed MAT for senior secondary schools. Specifically, the study determined the alternative scaled normative scores of the developed MATS senior secondary schools in Nigeria.

2.2 Research Questions

The following research questions were raised to guide the study:

RQ1: What is the mean of the developed standardized MAT instrument?

RQ2: What is the standard deviation of the developed standardized MAT instrument?

RQ3: What is the norm (Percentile-Rank; Standard score; Z-score; normalized T-score) of the developed standardized MAT instrument?

3. Research Methodology

This research study adopted an instrumentation research design. Instrumentation research design involves, among others, the determination of the norms of test instruments in order to make it a standard reference and easy-to-use tool. The choice of

this design was that the focus of the study was on establishing the baseline for comparison of the scores of the test takers. The instrument before now, when used, yielded raw data (scores) which were ambiguous in interpretation, as there were no norm-referenced scores for comparison. Interpretation of test takers' relative standing among the peers for which the test was designed to serve was difficult, as there was no benchmark (reference scores) for the test instrument. This study was carried out to bridge the gap and make the tool more effective and useful. The population of this study comprised all the Senior Secondary School two (SS2) students in Nigeria who sit for the third term examination in all private and public schools and are taught mathematics with the Nigeria Education Research and Development Center (NERDC) developed mathematics curriculum.

The MAT instrument was developed and used by officials of Ebonyi State Ministry of Education to assess students' readiness for entry to final year. The researchers obtained permission from the Ministry officials before they adopted the tool as an instrument for data collection. The instrument contains topics in statistics, number and numeration, probability, trigonometry and geometry. Since the instrument was already developed and validated, what the researchers did was adopt and use it. The researchers sampled a total of five hundred and ten (510) three educational zones of Ebonyi State.

The instrument was administered to the students by the class teachers of the sampled schools. The researchers only monitored the exercise to ensure that standard procedure was followed as contained in the developed test manual. The class teachers were sensitized using the test manual. The study was carried out towards the end of the third term before the beginning of the third term examination. The teachers marked the scripts, and the scores of the students were collated and analyzed using percentage and standard deviation. The SPSS (Statistical Package for Social Sciences) computer application was used to determine the mean, standard deviation and norm of the raw scores (Percentile Ranks, Z-score, T-score and Stanine). The SPSS package was also used to check the accuracy of the percentile ranks, normalized T-score and stanine outputs by determining the cumulative frequency, mean, and standard deviation of each output, as may be applicable.

4. Results

The results of this study are presented in accordance with the research questions that guided the study.

RQ1: What is the mean score achievement of students in the MAT instrument?

To answer this research question, the researchers relied on the information provided in Table 1.

Table 1: The output of data analysis of the scores of SS2 students who took the MAT test

Number	Mean (%}	S.D (%)	S.E (%)
510	51.9	20.5	0.91

From Table 1, the output of the analysis shows that the mean score value = 51.9% while the Standard Error (S.E) is 0.91. Based on these results, and consequent upon the size of the SE, which is 0.91, it is evident that there is close proximity of the sample mean to the population mean. This low SE of 0.91 is indicative of high reliability of the instrument which produced the calculated mean.

RQ2: What is the Standard Deviation (SD) of students' scores in the MAT test?

To answer the research question 2, the researchers further consulted the output of data analysis contained in Table 1.

Based on the data in Table 1, the standard deviation is 20.5. Consequent upon this SD of 20.5 and the mean value of 51.9, the instrument is able to separate the high and low achievers. In fact, the mean of the sample scores of the instrument is not placed very closely to the scores; rather, there is a moderate variability of approximately 40%. The scores are not clustered around the mean scores. Hence, the spread demonstrated a balanced in the scores of those who performed highly and those with low performance.

RQ3: What are the norm (Z-score; Percentile _ranks; T-score; and Stanine score) values of the MAT scores?

To answer this research question, the researchers relied on the SPSS computer output of the data analysis on the scores of 510 SS2 students who took the test. The output is contained in Appendix 1. However, an extract from a further breakdown of the outputs of the analysis is present in Table 3.

Table 3: Extract from the output of the analysis of
the normed values of the students' test scores from MATS

Variable (X)	Z-score	T-score (%)	Percentile Rank (%)	Raw score (%)	Stanine (%)
Mean	0.0	50	51.9	51.9	4.9
SD	1.0	10		20.5	1.9
SE		0.443		0.91	0.88
First Quantile	-0.8235	41.76	36	35	
Second Quantile	0.1029	51.02	54	54	
Third Quantile	0.68819	56.89s	66	66	

The output of the data analysis displayed in Table 3 indicates that the mean of the raw scores is 51.9, with an S.D of 20.5, while the SE is 0.91. The above figures are equivalent to a Z-score mean of 0.0 and an S.D of 1.0. The Z-score 25th quartile rank is -0.82350 away from the mean of the Z-score (0.0). The 50th (median) quartile is 0.102904 away from the mean, while the 75th quartile is 0.688192 away from the Z-score mean. Based on the

distances of those quartile rank values, it is shown that the 25th quartile rank is below the mean of the distribution, while the 50th and the 75th quartile ranks are above the mean of the normal (Z) distribution.

For the Percentile ranks, Table 3 shows that the mean is 51.9, the median P (50) is 54%; P (25) is 35%; while the P (75) is 66%. This shows that 66% of the normed group scored less than or equal to 88%. This number occupies the 75th percentile, while the P (50) of 54% showed that less than or equal to 54% of the students (the norm group) are in the median group and obtained 54%. Finally, the top 25th of the norm group obtained scores greater than or equal to 66%.

For the T-score equivalent, Table 3 shows that the mean of the norm group is 50% while the S.D is 10. The S.E is 0.443; the maximum value of the distribution is 71.51% while the minimum value is 27.13%. The median score is 51,02%, the 25th quartile rank is 41.76%; the 75th quartile rank is 56.89. The T=60 is approximately 84.1%. The T=60 value shows that a scorer with such a rank has approximately done better than 84% of the norm group.

For the Stanine distribution, which is a ranking with a scale of 1-9, the calculated mean is 4.9, which is approximately 5.0; the standard error is 0.88, while the standard deviation is 1.99, which is approximately 2. The outputs show that the stanine 1=4%, stanine 2=9% stanine 3=13%, stanine 4=17%, stanine 5=17%, stanine 6=20%, stanine 7=11%, stanine 8=5%, while stanine 9=5%. The slight variations in the output percentages of the stanine ranks and the standard stanine percentage distribution may have been caused by the distribution of the raw scores.

5. Summary of the Findings

The followings are the key findings of this study on the determination of the normative values or scores of the MATS test instrument:

1) The mean score of the norm group is 51.9, with the Standard Error (SE) of 0.91 indicating high precision and close proximity of the sample norm group to the population mean. This further indicates that the instrument is reliable.
2) The Standard Deviation (SD) of the norm group is 20.5. This shows that the scores are not clustered around the mean, which has a value of 51.9.
3) The derived values of the transformed scores from the norm group did not deviate from standardized scores. For instance, the mean Z-score of the norm group is 0.0 while the SD is 1.0. All the other transformed or derived scores produced values that are equivalent to values from the standardized instrument.

6. Discussion

The discussion of the findings of this study was done in line with the research questions that guided the study.

The findings of the study showed that the mean score of the study was 51.9%, with a mean standard error of 0.91. This indicates the proximity of the sample mean to the population mean and further demonstrates the high reliability of the instrument. Even though the sample mean is not synonymous with, nor one of the norm values (Alabama Parental Education Centre APEC, 2026), the mean (one of the measures of central tendency) plays an important role in data norming, as it is the statistic that presents data for test norming and describes the distribution (Anastacia *et al.*, 2010). The mean score is a foundational statistic for transforming raw data to derived scores, and without the mean transformation of raw scores cannot take place (Cronbach, 1990). Transformed scores make scores easy to understand, interpret and, most importantly, use for unbiased comparisons of individuals in the norm group's performance. Hence, mean impacts highly the quality of the transformed scores, irrespective of the quality of the instrument. Conversely, an inaccurate mean score leads to spurious derived values. It is the mean that gives the locus of standing (data centering and scaling) of any derived score, as it helps to measure how many units a score is from the mean in a consistent unit (Hans, 2022; Bishop, 2006; 2K Schools, 2026). Consequently, the roles of mean scores are central to the effective transformation of scores. Based on those realizations, test developers must ensure that the mean scores of data for transformation are correctly determined, as the quality of derived scores generated depends on it. Finally, without the mean, test developers cannot transform raw scores into scaled scores that make them an important tool for learning assessment in education.

The findings of this study showed that the standard deviation of the norm group is 20.5. This represents moderate variability of approximately 40%. This shows that the variability of the norm group separated the low and high achievers. The S.D as a statistic tool, is another important variable that occupies a central position in test transformation from raw to derived scores. Consequently, it is pertinent that test developers recognize the critical roles and significant impacts S.D plays in test norming. Undoubtedly, S.D is indispensable in test norming, as it is a very important foundational statistic just like mean for the transformation of raw scores it describes the distributions, with these calculated values of the S.D of 20.5 and mean value of 51.9%, it is now possible and easy to transform the raw scores of the norm group of this MATS test to any derived score of interest, which can serve as a baseline score in comparing scores of other test takers of this instrument who are in this norm group. The comparison serves as a yardstick for assessing the performance of the test taker.

In fact, transformed score operations depend wholly on standard deviation (Anasticia *et al.*, 2010; Cronbach, 1990). Without S.D, raw scores will not be changed to derived scores, which makes test scores useful. Its importance in test development is very obvious, as the S.D of a test shows the spread of the scores around the mean and helps in the interpretation of individual scores (APEC, 2026), as it shows the relative positioning of scores to the mean of the group. It helps to compare individuals' performances since it shows the variability of scores around the mean. It also quantifies the spread of the scores

from the mean. For instance, the Z-score is the measure of the number of S.D above or below the mean (Stafford, 2026). S.D provides the basis for the expression of an individual's scores on different tests in terms of the norms (Wordpress.com, 2013). Based on these values (S.D and mean) of this instrument, it is obvious that this instrument's derived scores are worthy of being used as a tool for comparison of students who fall within this norm group achievement score in the MATS test.

The findings of this study revealed that values of the derived scores from students' achievement in the MATS are valid and reliable. The values have a similar distribution to scores obtained in standardized test instruments; this signifies that the output can be relied on as a baseline for objective assessment of the test takers who are within the norm group's performance. The Z-score had a mean of 0.0 and an S.D of 1.0; the percentile rank, on the other hand, had a derived score with a mean of 51.9%; this shows that approximately 52% of the scores fall below the mean, while the stanine had a mean score of approximately 5.0 and a standard deviation of approximately 2,0. It is pertinent to acknowledge that the data contained in Table 3 are slightly lower and, in some cases, larger than expected standard-derived scores of equivalent groups, but that does not make them incorrect. For instance, in the case of Stanine, Study.com (2026) and Stanford (2018) stated that stanine scores (1-9) are broad categories (coarse grouping) rather than precise measurements. Slight deviations are normal in real-world data, especially with smaller sample sizes. According to them, Slight difference between the percentage of students in a specific stanine and the expected normal score does not make norming wrong.

Generally, norm scores play vital roles in the interpretation and understanding of students' performance in a widely used test instrument. The norm values of such tests provide the framework for objective comparison of an individual's achievement against their peer (norm) group on a standard basis. It provides context above raw scores or numbers to place students' ability in rank order. This guides education managers in students' class placement; it assists in the identification of students' weaknesses as well as strengths. In a national and statewide normed test, it helps schools to determine their performance and relative standing. Parents see it as a valuable tool in the interpretation of their children/wards' performances as they use it to compare the scores of their children/wards with those of their peers who formed the norm group. It is against these numerous benefits of the norm scores that the Ebonyi State Ministry of Education should strive to develop and use normative values of this nature in all subjects taught in all levels of basic and senior secondary school in the state to ease the interpretation and comparison of students' test achievement by stakeholders.

7. Conclusion

This study on the development of the normative scaled scores of the developed standardized MATs test instrument was carried out to develop the normative scaled

values for the MATS test instrument in order to bridge the gap of lack of a tool for comparison and interpretation of test takers' raw scores which had hitherto existed, and had as such hampered the efficient utilization of the instrument. Although most schools use this instrument in assessing the achievement of SS2 students in mathematics, there has been the challenge of a lack of a baseline tool for comparison and accurate interpretation of the scores of the students who take the test because of the unavailability of normative scores.

It is obvious that students' raw scores provide limited information on students' achievement, which does not aid accurate decision-making. The limited information from raw scores drastically affects the quality of decisions taken on students' academic progress; it jeopardizes the assessment of the rate of achievement of the instructional objectives outlined in the curriculum. To curb those challenges, the researchers undertook and developed this norm-scaled instrument. The instrument is in alternative scales to enhance its understanding, interpretation and utilization. The instrument will help parents and stakeholders to make comparisons and draw valid conclusions on the academic efficiency of education institution (schools) and the academic progress of the learners individually.

The researchers were motivated to conduct this study by the realization that any standardized instrument without a corresponding norm value in a scale that is known to the users cannot be adjudged to be truly a standardized tool. With the development of these standardized norm-scaled values for the MATS instrument, it has become a truly standardized instrument for educational assessment. It can now be used both in Ebonyi State and beyond to compare the scores of SS2 students who are administered the test.

8. Recommendations

In line with the important functions which standardized scaled-norm test instruments serve in educational instructional assessment, it is hereby recommended that:

1) Ebonyi State Ministry of Education should integrate these developed scaled norm values to be part of the MATS test instrument.
2) Ebonyi State Ministry of Education should develop a standardized test instrument for all the subjects taught in basic and secondary schools in the state for the assessment of students' achievement in the curriculum instructional objectives.
3) Ebonyi State Ministry of Education should mandate the Ebonyi State Universal Basic Education Board (SUBEB) to use this standardized test with the scaled norm values for the end-of-term unified mathematics examinations in the state.

Conflict of Interest Statement
The authors declare no conflicts of interest.

About the Author(s)
Dr. Christogonus Onyemaechi Nweke is a University Lecturer in Department of Science Education Ebonyi State University Abakaliki. His research focus is on Measurement and Evaluation with research interest in Foundational Learning.
ORCID: http://orcid.org/0009-0008-6196-9751
Dr. Julius O. Ugama is a senior Lecturer in Department of Science Education, Ebonyi State University Abakaliki. His research focus is on Mathematics Education.
Dr. Obeten Okoi Oka is a Lecturer in the Department of Science Education, Ebonyi State University Abakaliki, Nigeria. His area of specialization is on Biology Education.
Ikporo Festus Boss is a lecturer in the Department of Science Education, Ebonyi State University Abakaliki. His area of specialization is on Biology Education.
Nwodo Mary Odinaka is a lecturer in the Department of Science Education, Ebonyi State University Abakaliki. Her area of specialization is Physics Education.
Dr. Egbe Irene is a senior Lecturer in Department of Science Education, Ebonyi State University Abakaliki. Her research focus is Mathematics Education.
Dr. Igwe Micheal Igwe is a lecturer in the Department of Science Education, Ebonyi State University Abakaliki. Her area of specialization is Mathematics Education.
Nweke Benjamine Ifeanyi is a PG student in the Department of Science Education, Ebonyi State University Abakaliki. Her area of specialization is Mathematics Education.
Kpirigbe Benjamine Arobor is a PG student in the Department of Science Education, Ebonyi State University Abakaliki. Her area of specialization is Mathematics Education.
Ukworu Amarachi Favour is a PG student in the Department of Science Education, Ebonyi State University Abakaliki. Her area of specialization is His research focus is on Measurement and Evaluation with research interest in Foundational Learning.
Mgbogo Sunday Ogbule is a PG student in the Department of Science Education, Ebonyi State University Abakaliki. Her area of specialization is Science Education.
Dr. Augustina Nnenna Awoke is a lecturer in the Department of Arts and Social Science Education, Ebonyi State University, Abakaliki, Nigeria**.** Her area of specialization is Economics Education.

References

21K Schools (2026). *Norm-referenced test: meaning, types, examples and benefits.* Retrieved from https://www.21kschools.com

Alabama Parental Education Center APEC (2026). Test norming, standard scores and psychometric standards.

Anastesi, A. & Urbina, S. (2010). *Psychological Testing.* Pearson. Retrieved from https://books.google.ro/books/about/Psychological_Testing.html?id=lfFGAAAAMAAJ&redir_esc=y

Asuk, U. W. (2021). The challenges of test assessment and evaluation in the 21st century Nigeria education system. *Journal of Education and Society.* 11(3) 1755-1767. Retrieved from https://foeiaue.ng/publications/Volume11_Issue3/10.pdf

Bond, L A. (2015). *Norm and Creation-Reference Testing*. SCRIBD. https://doi.org/10.7275/dy7r-2x18

Bishop, C. (2026). *Pattern recognition and machine learning.* Springer. Retrieved from https://books.google.ro/books/about/Pattern_Recognition_and_Machine_Learning.html?hl=cs&id=kOXDtAEACAAJ&redir_esc=y

Denga, D. I. (2017). Science and Mathematics Education as Key to Nigeria's Technological Boom in the 21st Century. Akamkpa, *Journal of Science and Mathematics Education, 1(1),* 98-100.

Cronbach, L. J. (1990). *Essentials of Psychological Testing Education*, Inc., Upper Saddle River, New Jersey. Retrieved from https://books.google.ro/books/about/Essentials_of_Psychological_Testing.html?id=l_8MAAAAIAAJ&redir_esc=y

FRN (2017). *National policy on education (NPE), 5th edition. Nigeria.* Research and Development Council, Press, Lagos.

Han, J. (2022). *Data mining concepts and techniques*. Morgan Kaufman. Retrieved from https://books.google.ro/books/about/Data_Mining_Concepts_and_Techniques.html?id=pQws07tdpjoC&redir_esc=y

Hedge, M. N., & Pomaville, F. (2021). *Assessment of communication disorder in children. Resources and Protocols. 4th Edition.* Plural Publishing. Retrieved from https://books.google.ro/books/about/Assessment_of_Communication_Disorders_in.html?id=oEYgEAAAQBAJ&redir_esc=y

Kline, P. (2015). *A Handbook of Test Construction. Introduction to Psychometric Design.* Routledge. Retrieved from https://www.routledge.com/A-Handbook-of-Test-Construction-Introduction-to-Psychometric-Design/Kline/p/book/9781138905122

Kubi, J.A. (2021). *Assessing Content Validity of West African Senior Secondary School Certificate Examination Core Mathematics Question*. University of Cape Coast. Institutional Repository. Retrieved from https://ir.ucc.edu.gh/server/api/core/bitstreams/dccc131f-156d-453a-87b6-fd0027599721/content

Nwagu, E. K, N. (2023). *Development and standardization of social studies achievement test (SSATS).* Unpublished PhD thesis. University of Nigeria, Nsukka.

Nworgu, B. G. (2017). *Educational research: basic issues and methodology. 2nd edition*. Nsukka: University Trust Publishers.

Osadebe, P. U. (2016). *Construction and standardization of an economic achievement test for Senior Secondary School students.* Unpublished PhD Thesis. University of Port Harcourt. Rivers State. Retrieved from https://www.researchgate.net/publication/269836798_Construction_of_Economics_Achievement_Test_for_Assessment_of_Students

Quora (2026). Is Standard deviation useful in Psychology? Retrieved from https://www.quora.com/Is-standard-deviation-and-correlation-useful-in-psychology-Why

Stafford, M. E. (2026). *Standardized scores, what do you measure up?* SAGE Research Methods. Retrieved from https://methods.sagepub.com/book/mono/testing-and-measurement/chpt/standardized-scores-do-you-measure-up#_

Standford Information (2018). *University survey and assessment services (USAS).* University of Minnesota, USA.

Study.com (2025). Standardization & norms in psychological tests. Retrieved from https://study.com/academy/lesson/standardization-and-norms-of-psychological-tests.html

Study.com (2026). Stanine score /definition, calculation and uses. Retrieved from https://study.com/learn/lesson/stanine-score-uses-examples.html

Formplus (2026). Assessment tools: types, examples, & importance. Retrieved from https://www.formpl.us/blog/asssessment-tools

Thorndike, R. & Hagen E. (1969). *Measurement and evaluation in education and psychology. Third Edition.* John Wiley & Son Inc. Retrieved from https://books.google.ro/books/about/Measurement_and_Evaluation_in_Psychology.html?id=QsKdAAAAMAAJ&redir_esc=y

Toor, G. (2005). Meaning, purpose and characteristics of a test. GHG Khalsa College of Education.

Vasu, D. J. (2019). Pedagogical of mathematics. Retrieved from www.pedagogybyvasu.blogspot.com

WordPress.com (2026). Why standard deviation. Retrieved from www.scheung10.wordpress.com