WWW.IJRPR.COM

**International Journal of Research Publication and Reviews**

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Meta-Analysis of Empirical Studies on the Psychometric Properties of Standardized Examination Items in Science Subjects in Nigeria

***Adoga Ocheme Emmanuel[1], Agi Christiana Ikpoyi (Ph.D)[2], Baba Otukpa Patrick [3], Abdulahi Abdulkarim [4]***

Department of Guidance Counselling Education, Joseph Sarwuan Tarka University, Makurdi-Nigeria
Department of Guidance Counselling Education, Joseph Sarwuan Tarka University, Makurdi-Nigeria
Department of Guidance Counselling Education, Joseph Sarwuan Tarka University, Makurdi-Nigeria
Department of Guidance Counselling Education, Joseph Sarwuan Tarka University, Makurdi-Nigeria

**A B S T R A C T :**

The study is on Meta-Analysis of Empirical Studies on the Psychometric Properties of Standardized Examination Items in Science Subjects in Nigeria. Three research questions were raised while Ex post Facto Survey research design was adopted for the study. The area of the study is Nigeria and the population is all empirical studies on the psychometric properties of standardized examination items in science subjects in Nigeria, which are numerically infinite. Sample of 70 primary studies that considered at least one of the moderator variables (Difficulty, Discrimination, Distracter indices) were purposively drawn from the population using the Cochran's statistic for infinite or unknown population. The instrument used for data collection was a pro forma titled "coding sheet", aimed at eliciting the required information for the study. Comprehensive Meta-Analysis (CMA), effect sizes, Z-transformations and descriptive summaries were used to answer research questions. The findings revealed that the effect sizes of the psychometric properties (Difficulty, Discrimination and Distracter indices) of standardized examination items in science subjects in Nigeria are significantly positive. Based on the obtained results, it was concluded that the psychometric properties of standardized examination items in science subjects in Nigeria are in conformity with the set standard. Therefore, the quality of psychometric properties of standardized examination items in science subjects in Nigeria should be sustained, concerted efforts should be made to improve upon the relatively good psychometric properties of standardized examination items in science subjects in Nigeria toward achieving excellence, and team work amongst educational stakeholders should be fostered and encouraged to avoid decline in quality.

**Keywords:**
Test
Difficulty
Discrimination
Distracter indices
Meta-Analysis

## Introduction

The effectiveness of educational processes can only be determined through appropriate assessment mechanisms. Consequently, tests and examinations play a central role in evaluating the extent to which learning objectives have been achieved (Adom, Mensah & Dake, 2020). According to Olajide (2018), a test is a systematic assessment used to measure a student's knowledge, skills, or abilities in a specific area, which can be in various forms, such as: quizzes or standardized examination, and aims to provide data on student's ability, knowledge, performance and achievement while Testing specifically in education is a process that is administered to measure learner's ability to perform in a specified learning area in a defined time limit with some specific goals which generally act as a gate-pass to successful outcome. The necessity of testing in education cannot be overstressed. This is in tandem with the submissions of Emaikwu (2014) that testing has become one of the most important parameters by which a society adjudges the product of her educational system, as it has always been a critical part of the school structure that even the habitual absentees normally turn up to school and present themselves for testing on examination days. Emaikwu (2014) further submitted that testing has been fully accepted in most modern societies as the most objective method of decision making in schools, industries and government establishments, commonly used for admission, recruitment, promotion, placement, evaluation, guidance, research and teaching purpose, with the aim of revealing the latent ability of the examinee. Because of the importance and far reaching application of testing, it is very important that common standards are set to ensure objective evaluation, consistency and accountability. Assad (2022) simply referred to the observation of these benchmarks as test standardization.

Test standardization or standardized examinations, as it could be interchangeably used in this study provides every candidate with equal opportunities to demonstrate competence using the same set of carefully developed assessment items (Lai, 2023). Test standardization is a method of assessment built on the principle of consistency without which measurement will only be an informal process that varies from examiner to examiner. There are finite basic steps to be followed when preparing a standardized test in any subject area. These include: preparation of content outline, specification of the

objectives, preparation of test blue print or table of specification, determination of test forms, construction of test items, face validation of the items, trial testing, empirical item analysis, item selection, determination of test length, determination of reliability coefficient, determination of Cut-off score, preparation of scoring key and periodic revision. Assad (2022) asserted that because of the procedures followed in the standardization of an examination or test, the items are usually of high technical quality and objectivity, and one of the critical stage in the process of standardizing an examination or test is item analysis. Baidoo-Anu (2019) defined item analysis as a statistical technique which involves collecting, summarizing and exploiting information from students' responses to make decisions about each assessment items, thereby aiding instructors to ascertain the effectiveness of test items for retention, revision or removal with the primary aim of developing a quality assessment.

The primary aim of item analysis is to improve items of a test for future use based on indices of item difficulty, item discrimination as well as the extent of effective functioning of the distractors, from which the dimensionality of the items and relative properties of the test like its reliability and validity could be subsequently established.

Item difficulty parameter, location or item threshold as it could be interchangeably used and a variable of this study, is the likelihood of an examinee to answer an item correctly. Assad, (2022) simply stated that item difficulty is the number of students who answered an item correctly, and that in classical test theory (CTT), difficulty index ranges from 0.00 to 1.00, whereby the higher the value, the easier the question and the lower the value, the difficult the items. Baidoo-Anu (2019) generally asserted that the range of items difficulty can be categorized into difficult, moderate, and easy, underlining that easy and difficult items are reported to have very little discrimination power. Therefore, items should have acceptable values of difficulty since very difficult or very easy items contribute little to the discriminating power of a test.

Item discrimination or item slope index, as a variable of this study, and one of the psychometric properties of standardized examinations is the feature of an item to discriminate or differentiate between examinees who are high achievers and the low achievers, which is simply defined as the slope of item characteristics curve at the point of inflection. Discrimination reflects the extent to which an item differentiates learners based on their abilities and is often represented by the slope of the item characteristic curve at its point of inflection, usually ranging between -1 to +1. Positive values indicate that high-achieving students are more likely to answer the item correctly than low-achieving students, while negative values suggest potential flaws in the item. The closer the index is to +1, the more effectively the item distinguishes between the two groups of students, which is desirable as it indicates that high achievers answer the item correctly more than the low achievers, and the closer the index to -1, the less effectively the item distinguishes between the two groups as this indicates that low achiever examinees answer the item more correctly, while zero discrimination suggests equal numbers of students in the upper and lower groups (Baidoo-Anu, 2019). Discrimination indices are positively associated with difficulty index and distractor efficiency as the discrimination power of an item is reduced by the increased number of non-functional distractors, being a variable of interest in this current study and a psychometric property of standardized examination items.

Distractor efficiency is the ability of incorrect answers to distract the examinee. Assad (2022) disclosed that a distractor could be functional or non-functional. Functional distractors are clearly incorrect but plausible options to an examinee who is not sufficiently knowledgeable in the content area while non-functional distractors are non-plausible options with no contribution to the measurement of the test. A distractor selected more frequently than the key by a higher-scoring examinee may indicate faulty item, therefore, it should be selected by a greater proportion of the lower scorers than the top group. Also, for a distractor to be functional it should attract at least one examinee and if no one selects a distractor it is important to revise the option and attempt to make it a more plausible choice as item quality essentially depends relatively on the effective functioning of the selected distractors.

The prevalent investigation of the psychometric properties of standardized examinations in science subjects in Nigeria with contradicting results has brought about the worry to use Meta-analysis in an attempt to finding a common ground for these studies. Meta-analysis, which is the crux of this study, is the statistical combination of information from different individual studies which provides a consensus estimate of effect size statistic, allowing the researcher to exploit associated attributes (Hannah, 2014). According to Demirel and Dagyar (2016), Meta-analysis is a quantitative statistical method for aggregating and statistically evaluating reported descriptive from several relevant published and unpublished research studies that discusses and test the same conceptual research questions and hypotheses, used to test competing theoretical assumptions against each other and to identify important moderators where the results of different primary inquiries differ from each other, in order to draw evidence-based recommendations for practice and policy making. Ugwuanyi (2014), further clarified that Meta-analysis is a statistical technique used in blending data from more than one study that addresses the same issue in the hope of identifying places of diversity and conflictions between study results that will help in taking decision which can lead to a generalizable scientific conclusion.

The contribution of science and science education to the immediate society and national development has given it prominence that learners are being taught across levels of learning and are examined to ascertain the extent of knowledge gained in different science subject areas such as Mathematics, Further Mathematics, Chemistry, Physics, Biology, Agricultural Science and geography among others. Significantly, Standardized examination bodies has made provision for examinees to be assessed in these subject areas and graded based on performance for subsequent decision making

### *Research Questions*

Three research questions were raised to guide the study

1 What is the mean effect size of difficulty indices of empirical studies on standardized examination items in science subjects in Nigeria?
2 What is the mean effect size of discrimination indices of empirical studies on standardized examination items in science subjects in Nigeria?
3 What is the mean effect size of distracter efficiency indices of empirical studies on standardized examination items in science subjects in Nigeria?

## Literature Review

Testing is as old as learning and the correlation between the two is quite inseparable. Conversely, one may say that the household popularity widely gained by testing in the society today can be credited to the sweeping force of the very common standardized examination bodies such as the West African Examinations Council (WAEC), National Examinations Council (NECO), National Business and Technical Examinations Board (NABTEB), the Joint Admissions and Matriculations Board (JAMB) among others on which qualification for formal learning at the upper level is mostly centered upon. These tests are deemed standardized because they require have met some very specific and unique features of the term like generalizability and uniformity, having undergone the processes of standardization. Despite the perceived consonance in standardization of these tests, researchers in education and psychology have continued to investigate and flood the body of literatures of research studies on these examinations, their psychometric properties and examinees' performances, year in and out, with diverse and conflicting findings. From literatures available, researchers such as Omirin and Adekunle (2021) conducted a Meta-Analysis of Studies on Psychometric Properties of WAEC and NECO Examinations in Nigeria on WAEC and NECO Examinations since the inception of NECO in the year 2000 and Findings from the study established a significant effect sizes for the psychometric variables under study. Aborisade and Fajobi, (2020), Thomas, Uchegbue and Ugbe (2011) conducted studies on psychometric properties of standardized examinations independently with disparity in findings.

If the implication of standard is as a guide towards established benchmark then these literatures are pointers that the psychometric properties of some of the examined test items deviate from set-standards and from one test to another, which could in turn question the quality of examination items, capable of exerting influence over the performance of examinees, assuming performance is a function of ability, and ability in part, dependents on the standard of examination items.

## Methodology

Ex Post Facto Survey research design with triangulation of data, design and investigators was adopted for the study and the population consisted of all published and unpublished previous empirical studies on the psychometric properties of standardized examination items in science subjects in Nigeria. Seventy primary studies were sampled as the sample size for the study using the Cochran's statistic for infinite or unknown population at 90% confidence level, 10% margin of error and 50% population proportion. These includes published and unpublished research reports (in master degree project reports, doctoral dissertations, journal articles and reports, conference and seminar papers) on standardized examination items in science subjects in Nigeria, that have considered at least one of the psychometric properties of test were selected using purposive and snowballing sampling technique in line with the inclusion criteria viz: specified examination body, subject area, sample size and psychometric indices. Comprehensive Meta-Analysis (CMA), effect sizes, Z-transformations and descriptive summaries were used to answer research questions.

## Findings

**4.1. Research Question 1:** What is the mean effect size of difficulty indices of empirical studies on standardized examination items in science subjects in Nigeria?

**Table 1: weighted mean, r-value of weighted mean, Cohen's d value of r, interpretation of obtained Cohen's d value of r and percentage variance of all the Studies Investigated on difficulty indices of empirical studies on standardized examination items in science subjects in Nigeria**

| Studies on diff. index | $\sum(N-3)$ | $\sum$ Weighted Z | Average Zr | r-value average Zr | d-value of r | (%) Variance | Cohen Result |
|---|---|---|---|---|---|---|---|
| 58 | 59829 | 17052 | 0.29 | 0.28 | 0.590 | 7.84 | Large |

d<0.2= small effect size, 0.2≤ d ≥ 0.49 = medium effect size, d ≥ 0.5 = high effect size

Table 1 shows the mean effect size estimate of difficulty indices of all studies investigated on empirical studies on standardized examination items in science subjects in Nigeria. The result shows a Cohen's effect size of 0.590 which indicates a large and positive effect with a percentage variance of 7.84 %.

**4.2 Research Question 2:** What is the mean effect size of discrimination indices of empirical studies on standardized examination items in science subjects in Nigeria?

**Table 2: weighted mean, r-value of weighted mean, Cohen's d value of r, interpretation of obtained Cohen's d value of r and percentage variance of all the studies investigated on discrimination indices of empirical studies on standardized examination items in science subjects in Nigeria**

| Studies on disc. Index | $\sum(N-3)$ | $\sum$ Weighted Z | Average Zr | r-value average Zr | d-value of r | (%) Variance | Cohen Result |
|---|---|---|---|---|---|---|---|
| 58 | 59829 | 15369 | 0.26 | 0.25 | 0.530 | 6.25 | Large |

d<0.2= small effect size, 0.2≤ d ≥ 0.49 = medium effect size, d ≥ 0.5 = high effect size

Table 2 shows the mean effect size estimate of discrimination indices of all studies investigated on empirical studies on standardized examination items

in science subjects in Nigeria. The result shows a Cohen's effect size of 0.530 which indicates a large and positive effect, with a percentage variance of 6.25 %.

**4.3 Research Question 3:** What is the mean effect size of distracter efficiency indices of empirical studies on standardized examination items in science subjects in Nigeria?

**Table 3: weighted mean, r-value of weighted mean, Cohen's d value of r, interpretation of obtained Cohen's d value of r and percentage variance of all the studies investigated on distracter efficiency indices of empirical studies on standardized examination items in science subjects in Nigeria**

| Studies on dist. index | $\sum (N-3)$ | $\sum$ Weighted Z | Average Zr | r-value average Zr | d-value of r | (%) Variance | Cohen Result |
|---|---|---|---|---|---|---|---|
| 40 | 42838 | 9409 | 0.22 | 0.22 | 0.440 | 4.84 | Medium |

d<0.2= small effect size, 0.2≤ d ≥ 0.49 = medium effect size, d ≥ 0.5 = high effect size

Table 3 shows the mean effect size estimate of distracter efficiency indices of all studies investigated on empirical studies on standardized examination items in science subjects in Nigeria. The result shows a Cohen's effect size of 0.440 which indicates medium and positive effect, with a percentage variance of 4.84 %.

## Conclusion and Recommendations

### *5.1. Conclusion*

Based on the findings of the study, it was concluded that the psychometric properties of standardized examination items in science subjects in Nigeria are in conformity with the set standard.

### *Recommendations*

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

1. The quality of psychometric properties (Difficulty, Discrimination, Distracter, Reliability indices and dimensionality) of standardized examination items in science subjects in Nigeria should be sustained.
2. Continuous and concerted efforts should be made by test experts to continue to improve upon the relatively good psychometric properties of standardized examination items in science subjects in Nigeria toward achieving excellence.

**REFERENCES**


- Aamodt, M. G., & McShane, T. D. (1992). A meta-analytic investigation of various test item characteristics on test scores and test completion times. *Public Personnel Management*. 21 (2) 151-160.
- Aborisade, O.J., & Fajobi, O.O (2020). Comparative analysis of Psychometric Properties of mathematics items construction by Waec and Neco in Nigeria using item response theory approach. *Educational Research and Review, 15(1), 1-7*
- Adelakun, N. O. (2021). Present state of tertiary institutions in Nigeria, Medeley data, V1, doi: 10.17632/4c8vv5h42j.
- Adeyemo, E.O. (2012). A meta-analysis of empirical studies on the validity of university matriculation examinations in Nigeria. 18. 45-46
- Adom, D., Mensah, J. A., & Dake, D. A. (2020). Test, measurement and evaluation: Understanding and use of the concepts in education. *International Journal of Evaluation and Research in Education (IJERE)* Vol. 9, No. 1, March 2020, pp. 109 – 119 ISSN: 2252-8822, DOI: 10.11591/ijere.v9il.20457
- Assad, A. R. (2022). Item Analysis: Concept and Application. Department of Basic Medical Sciences (Unit of Anatomy), College of Medicine, University of Bisha, Saudi Arabia. *DOI: http://dx.doi.org/10.5772/intechopen.100138*
- Baidoo-Anu, D. (2019). Item Analysis of Multiple Choice Items Queen's University Technical Report · October 2019 DOI: 10.13140/RG.2.2.16627.89124
- Benzer, J. & Horner, M. (2014). A Meta-analytic Integration and Test of Psychological Climate Dimensionality. *Human Resource Management* 54(3) 457-482 Doi.org/10.1002/hrm.21645
- Demirel, M. & Dagyar, M. (2016). Effects of Problem-Based Learning on Attitude: A Meta- analysis Study. *Eurasia Journal of Mathematics, Science & Technology*, *12*(8), 2115–2137. https://doi.org/10.12973/eurasia.2016.1293a

- Emaikwu, S.O., (2014). Aptitude testing in the 21st century in Nigeria: The basic issues to rivet. Research Journal in Organizational Psychology and Educational Studies 3(3) 14

- Fausiyah, A. Z., & Abdul-Wahab, I. (2015) A Meta-Analytic Assessment of Standard Fixing at Senior Secondary School Certificate Examination by West African Examinations Council, Nigeria.

- Hannah, R. (2014). Basic aspects of meta-analysis. *Baruch College-City University of New York.*

- Knowles, S.L.,& Welch, C.A. (1992). A Meta-Analytic Review of Item Discrimination and Difficulty in Multiple-Choice Items Using "None-Of-The-Above." *Educational and Psychological measurement*, 52(3), 571-577

- Lai, Y., (2023). The Double Effect of Standardized Testing on Students and Environment. February 2023., Journal of Education Humanities and Social Sciences 8:1615-1620 DOI: 10.54097/ehss.v8i.4533

- Olajide, S. B. (2018). Test as Tool for Effective Research in the Humanities. *Bulgarian Journal of Science and Education Policy (BJSEP),* Volume 12, Number 2, 2018

- Omirin, M.S. & Adekunle, A. (2021). Meta-Analysis of Studies on Psychometric Properties Of Waec and Neco Examinations in Nigeria. *Universe International Journal of Interdisciplinary Research* DOI: http://www.doi-ds.org/doilink/02.2021-22842258/UIJIR

- Rios, J.A., Deng, J & Samuel D. Ihlenfeldt, S.D (2022). *To What Degree Does Rapid Guessing Distort Aggregated Test Scores? A Meta-analytic Investigation*. (2022). https://doi.org/10.31234/osf.io/8jnsu 10.1080/10627197.2022.2110465

- Thomas, E.S., Uchegbue, H., & Ugbe, E.B. (2016). Comparative item analysis of students' Waec and Neco Mathematics objectives test item scores. *Educational Assessment and Research Network in Africa*

- Ugwuanyi, C.C. (2014). Effectiveness of Cooperative Learning Method in Mathematics in Nigeria: A Meta- Analysis. *International Journal for Cross-Disciplinary Subjects in Education (IJCDSE),* 6 (2), 2150 - 2160.

- Zhao, H., & Aryadoust , V. (2024). A meta-analysis of the reliability of second language reading comprehension assessment tools *Cambridge University Press.*