ISSN 2736-0091
E-ISSN 2737-0107

INFLUENCE OF TEST ITEM ARRANGEMENT AND DEMOGRAPHIC VARIABLES ON THE DISCRIMINATION INDICES OF PHYSICS MULTIPLE-CHOICE ITEMS IN NIGERIAN SECONDARY SCHOOLS

BY
Oyemomi Mope Esan: Department of Guidance and Counselling, Federal University, Oye-Ekiti, Nigeria;
E-mail: oyemomi.esan@fuoye.edu.ng

Samuel Oye Bandele: Institute of Education, Ekiti State University, Ado-Ekiti, Ekiti State, Nigeria;
E-mail: professorsamueloyebandele@gmail.com

Femi Emmanuel Babalola: Department of Science Education, Federal University, Oye-Ekiti, Nigeria;
E-mail: femi.babalola@fuoye.edu.ng
&
Audu Godwin: Department Science Education, Federal University Oye-Ekiti;
E-mail: godwinaudu610@gmail.com

Abstract
Assessment is essential for evaluating students' learning and guiding instruction, particularly in science education, where Physics is a gateway to STEM careers. A critical psychometric property is the discrimination index, which indicates how well test items differentiate between high- and low-achieving students. However, limited research has explored the impact of test item arrangement and demographic factors on discrimination indices in Nigerian Physics assessments. This study examined the discriminating power of 50-item multiple-choice Physics tests arranged in ascending, descending, and non-directional (random) formats, while also assessing the influence of gender and state of origin. Using a quasi-experimental ex post facto design, 675 senior secondary students from Ekiti, Lagos, and Osun states were selected for the study. Three parallel forms of the test were administered according to different arrangements. Data were analyzed using descriptive statistics, one-way and two-way ANOVA, the Kruskal-Wallis test, and Scheffé post hoc comparisons at a significance level of 0.05. Results revealed that item arrangement significantly affected discrimination indices, with ascending order yielding the highest differentiation among students of varying ability. Gender and state of origin had no significant effect and did not interact with arrangement. The study concludes that the test structure, rather than demographic factors, determines the fairness and validity of Physics assessments. It recommends adopting ascending item sequencing and integrating psychometric analysis into policy and test development to ensure equitable and valid evaluation in science education.

Keywords: Physics education, Test item arrangement, Discrimination index, Gender and State of origin

Introduction
Assessment is a cornerstone of the educational process, serving as both a measure of learning outcomes and a guide for instructional improvement (Levy-Feldman, 2025). In science education, particularly Physics, assessment plays a crucial role in evaluating students' knowledge, identifying learning gaps, and informing instructional strategies (Murphy et al., 2023). Multiple-choice tests are widely employed for their objectivity, reliability, and efficiency in large-scale evaluation (Butler, 2018). Beyond measuring content knowledge, test construction must prioritize fairness, validity, and reliability (Kemper, 2020; Ayanwoye et al.,2025). A key psychometric property in this regard is the discrimination index, which measures how effectively test items distinguish between high and low-performing students (Kumar et al., 2021; Esan & Babalola, 2024). High discrimination values indicate well-designed items that accurately reflect mastery of the content, while low values undermine validity and compromise the interpretation of student performance (Rush et al., 2016; Chai et al., 2019).

In Physics, where abstract reasoning and problem-solving are emphasized, the design and arrangement of test items can substantially influence both student outcomes and the reliability of assessment interpretations (Odden et al., 2021). Physics assessments are critical because they serve as a gateway to STEM careers such as engineering, medicine, and technology (Tan et al., 2023). Researchers have increasingly focused on test item arrangement,

ISSN 2736-0091
E-ISSN 2737-0107

ascending, descending, or non-directional (random), as a factor influencing discrimination indices (Robie et al., 2022). Cognitive psychology suggests that item sequencing affects motivation, anxiety, and cognitive load (Otero & Alonso, 2023). Ascending arrangements, which begin with easier items and progress to more challenging ones, may help reduce anxiety and build confidence (Anaya et al., 2022). In contrast, descending sequences may discourage less prepared students by presenting difficult items first (Tobia et al., 2022). Non-directional arrangements eliminate predictability, potentially offering a more authentic measure of ability, but can introduce inconsistency and unpredictability in outcomes (Gu, 2010). Despite these theoretical insights, limited empirical research has examined the influence of item arrangement on discrimination indices in Nigerian Physics education (Rezigalla et al., 2024).

Fairness in assessment also requires attention to demographic variables such as gender and regional background (Rezai et al., 2022; Adaramoye et al., 2025; Olayiwola-Adedoja et al., 2025). Persistent gender disparities in science performance have been attributed to differences in socialization, access to resources, and confidence levels (Smith et al., 2023), as well as to test design, including item sequencing, which may inadvertently contribute to these disparities (Bowman et al., 2022). Similarly, regional factors, including students' state of origin, affect access to quality instruction, exposure to resources, and preparedness for standardized assessments (Silverman et al., 2022). These considerations underscore the importance of integrating psychometric evaluation with equity considerations to ensure fair and valid assessments (Muskens et al., 2024).

This study addresses these gaps by investigating the discrimination indices of multiple-choice Physics test items arranged in ascending, descending, and non-directional formats, while examining the moderating effects of gender and state of origin. Grounded in psychometric principles, particularly Classical Test Theory, the study acknowledges that observed test scores comprise both accurate scores and measurement error, with item difficulty and discrimination playing central roles in minimizing error and enhancing reliability (Cappellert et al., 2014; Allen & Yen, 2002; McNeish, 2024). Cognitive Load Theory complements this perspective by emphasizing how working memory limitations interact with item sequencing, affecting cognitive load, performance, and potentially discrimination indices (Sweller, 2011; Sepp et al., 2019; Olayiwola-Adedoja et al., 2024). Ascending arrangements may reduce extraneous cognitive load, whereas descending or non-directional sequences may increase cognitive demands, particularly for weaker learners (Fowler, 2022; Langerock et al., 2025; Gorgun et al., 2025). By integrating psychometric theory and cognitive learning principles, this study posits that the arrangement of test items significantly influences their discrimination power. At the same time, demographic variables such as gender and state of origin may moderate or interact with these effects. Understanding these dynamics is critical for developing fair, valid, and reliable Physics assessments that accurately distinguish between high- and low-achieving students, promote equity, and inform instructional and policy decisions in science education.

Research Objectives

The study seeks to:

1. Examine the differences in item discriminating indices of Physics test items arranged in ascending, descending, and non-directional order.
2. Determine whether gender significantly influences the discriminating indices of Physics test items across different arrangement formats.
3. Investigate whether the state of origin significantly influences the discriminating indices of Physics test items across different arrangement formats.

Research Hypotheses

1. There is no significant difference in item discriminating indices of test items arranged in ascending, descending, and non-directional order
2. The Gender of students will not significantly influence the item discriminating indices of Physics test items arranged in ascending, descending, and non-directional order.
3. State of origin will not significantly influence item discriminating indices of Physics test items arranged in ascending, descending, and non-directional order.

ISSN 2736-0091
E-ISSN 2737-0107

Methodology

This study employed a quasi-experimental, ex post facto research design to examine the discriminating indices of Physics test items across three arrangement formats: ascending, descending, and non-directional (random). This design was suitable because the study aimed to analyze existing test outcomes, compare the discriminatory power of items across different arrangements, and assess the influence of demographic variables such as gender and state of origin. The independent variables were test arrangement, gender, and state of origin, while the dependent variable was the discrimination index of the Physics test items. The population consisted of senior secondary school students who offered Physics in selected states in southwestern Nigeria. A sample of 675 students was drawn from Ekiti, Lagos, and Osun states using stratified random sampling to ensure representation across gender and region. Within each state, schools were purposively selected to include both male and female students from diverse backgrounds, providing a robust dataset for analyzing the effects of item arrangement and demographic factors on discrimination indices.

Data were collected using a 50-item multiple-choice Physics achievement test developed by the researcher and validated by subject-matter experts in Physics education and measurement. The test spanned core concepts, including mechanics, heat, waves, electricity, and modern Physics. Each item had four options, with one correct answer and three distractors. To examine the effects of item arrangement, three parallel forms of the test were constructed: ascending order, arranged from simplest to most complex based on Bloom's taxonomy and expert judgment; descending order, arranged from most complex to simplest; and non-directional, with items randomized. The test was pretested for clarity, difficulty, and discrimination indices, and reliability was established using the Kuder-Richardson Formula 20 (KR-20), yielding a coefficient above 0.70. The study was conducted in multiple stages. Schools granted permission, and students were informed about the purpose and assured of confidentiality. Students were randomly assigned to one of the three test arrangements, ensuring comparable group sizes and demographic composition. Tests were administered under standardized conditions to minimize extraneous variables, and scripts were scored objectively: correct answers earned one mark, and incorrect responses earned 0. Item analysis was then conducted to compute discrimination indices for each item across the three arrangements.

Data analysis involved both descriptive and inferential statistics. Means and standard deviations were summarized to provide discrimination indices across arrangements. One-Way Analysis of Variance (ANOVA) was used to test for significant differences in discrimination indices among the three arrangements, with the Scheffé post hoc test and the Kruskal-Wallis test used for pairwise comparisons. Two-way ANOVA was employed to examine the influence of gender and arrangement, as well as the interaction between these factors, and similarly for state of origin and arrangement. All analyses were conducted using SPSS version 25, with results interpreted using p-values, effect sizes, and post hoc comparisons to provide a comprehensive understanding of how item arrangement and demographic factors influenced discrimination indices. Ethical standards were rigorously observed. Participation was voluntary, and informed consent was obtained from both school authorities and students. Confidentiality was strictly maintained. The study was designed to avoid bias and harm, ensuring fairness and objectivity throughout the research process. All hypotheses were tested at the 0.05 level of significance, providing robust evidence on the effects of test arrangement, gender, and state of origin on Physics item discrimination.

Results

To answer the formulated hypothesis, data on the discriminating indices for the three arrangements (Ascending, descending, and non-directional order) were used.

Research Hypotheses

Hypothesis One: There is no significant difference in item discriminating indices of test items arranged in ascending, descending, and non-directional order.

To test this hypothesis, a one-way ANOVA was employed (Table 3a), with the Scheffé post hoc test (Table 3b) for pairwise comparisons. A nonparametric statistical test, the Kruskal-Wallis test, was used. This is equivalent to one-way analysis of variance (ANOVA). Table 3a presents the results as follows:

Table 1a: Kruskal-Wallis test for item discrimination indices of physics test item arrangement

Statistics	Discrimination (a)

ISSN 2736-0091
E-ISSN 2737-0107

Kruskal-Wallis H	16.24
N	150
Df	2
Asymp. Sig.	0.00

a. Kruskal-Wallis Test

Table 1b: Grouping Variable: Test Item Arrangement
Analysis of Variance for item discrimination of physics test item arrangement

	Sum of squares	Df	Mean square	F	Sig.
Between Groups	1.147	2	574	6.334	002
Within Group	13.312	147	091		
Total	14.460	149			

Table 1a remarked that there was a significant difference in the discrimination statistics of the three forms of test arrangement of Physics multiple-choice with (Kruskal-Wallis H Statistic = 16.24, $p < 0.05$). Also, pair-wise comparison (post-hoc analysis) was conducted to identify where the difference lies.

Table 1c: Pairwise Comparisons of Discrimination Indices across Physics Test Arrangement

Sample 1-Sample 2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.a
Descending Order-non-directional Order	5.02	8.69	0.58	0.56	1.00
Descending Order-Ascending Order	32.51	8.69	3.74	0.00	0.00
Topic Order-Ascending Order	-27.49	8.69	-3.16	0.00	0.01

*Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.

*Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

Table 1a showed that descending-order discrimination indices were significantly different from ascending-order discrimination indices (Standardized Test statistic = 3.74, $p < 0.05$). Moreover, the table showed that non-directional order discrimination indices were significantly different from those of ascending order (Standardized Test statistics = -3.16, $p < 0.05$). However, descending-order items' discrimination indices were not significantly different from those in non-directional order (Standardized Test statistic = 0.58, $p > 0.05$). This was further corroborated by Figure 2, which shows that test items presented in ascending order (95.50) discriminated better among test-takers than those presented in non-directional order (68.01) or descending order (62.99).

Hypothesis Two: The Gender of students will not significantly influence item discriminating indices of Physics test items arranged in ascending, descending, and non-directional order

Table 2a: Summary of Two-Analysis of Variance of Gender and Test Arrangement on Item Discriminating Indices

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	326.912a	5	65.382	514.686	0.000
Intercept	1066.611	1	1066.611	8396.287	0.000
Gender	0.017	1	0.017	0.135	0.713
Physics Test Arrangement	290.697	2	145.348	1144.173	0.000
Gender * Physics Test Arrangement	0.056	2	0.028	0.221	0.802
Error	84.986	669	0.127		
Total	1650.181	675			

ISSN 2736-0091
E-ISSN 2737-0107

Corrected Total	411.898	674			

a R Squared = .794 (Adjusted R Squared = .792)
b Dependent Variable: Item Discrimination Indices

Table 2a presents a two-way ANOVA examining the effects of gender and physics test arrangement on item discrimination indices. Table 4a results showed that there was no significant interaction between gender and test arrangement on item discrimination indices, $F_{(2, 669)} = 0.221$, p = 0.802. Additionally, no statistically significant difference in item discrimination indices was observed between genders (p = 0.713); however, a statistically significant difference was found between test arrangements ($p < 0.05$). This was further confirmed using estimated marginal means descriptive statistics, as presented in the table below.

Table 2b: Estimated Marginal Means for Gender and Test Arrangement on item discrimination indices.

Gender	Physics Test Arrangement	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Male = 1	Non-directional Order	2.332	0.04	2.253	2.41
	Ascending Order	0.795	0.028	0.74	0.851
	Descending Order	0.945	0.029	0.889	1.001
Female = 2	Non-directional Order	2.337	0.029	2.279	2.395
	Ascending Order	0.795	0.044	0.71	0.881
	Descending Order	0.906	0.043	0.822	0.99

Table 4b showed that there was no mean difference between male and female item discrimination indices when various test arrangement format was administered to them, with (means for non-directional order: Male = 2.332, female= 2.337; means for ascending order: Male = 0.795, female= 0.795; and means for descending order: male=0.945, female = 0.906) respectively.

Hypothesis Three: State of origin will not significantly influence item discriminating indices of Physics test items arranged in ascending, descending, and non-directional order.

Table 3a presents a two-way ANOVA summary as follows

Table 3a: Summary of Two-Analysis of Variance of State of Origin and Test Arrangement on Item Discriminating Indices

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	327.172a	8	40.897	321.473	0.000
Intercept	1238.283	1	1238.283	9733.71	0.000
State Origin	0.111	2	0.055	0.435	0.648
Physics Test Arrangement	326.84	2	163.42	1284.589	0.000
State Origin * Physics Test Arrangement	0.221	4	0.055	0.435	0.783
Error	84.726	666	0.127		
Total	1650.181	675			
Corrected Total	411.898	674			

a R Squared = .794 (Adjusted R Squared = .792)
b Dependent Variable: Item Discrimination indices

Table 3a presents a two-way ANOVA examining the effects of State of origin and physics test arrangement on item discriminating indices. Table 5a remarked that there was no significant interaction between the effects of State of origin and physics test arrangement on item discriminating indices, $F_{(2, 666)} = 0.435$, p = 0.783. Also, the result showed that there was no statistically significant difference in the item discriminating indices between State of Origin ($p = 0.648$), but there were statistically significant differences between discriminating indices ($p < 0.05$).

ISSN 2736-0091
E-ISSN 2737-0107

Table 3b: Estimated Marginal Means for State of Origin and Test Arrangement on Item Discriminating Indices

State Origin	Physics Test Arrangement	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Ekiti	Non-directional Order	2.345	0.041	2.264	2.426
	Ascending Order	0.795	0.041	0.714	0.876
	Descending Order	0.933	0.041	0.852	1.014
Lagos	Non-directional Order	2.284	0.041	2.203	2.365
	Ascending Order	0.795	0.041	0.714	0.876
	Descending Order	0.933	0.041	0.852	1.014
Osun	Non-directional Order	2.377	0.041	2.296	2.458
	Ascending Order	0.795	0.041	0.714	0.876
	Descending Order	0.933	0.041	0.852	1.014

Table 3b showed that there was no mean difference between Ekiti, Lagos and Osun item discriminating indices when various test arrangement format was administered to them with (means for non-directional order: Ekiti = 2.345 Lagos = 2.284, Osun = 2.377; means for ascending order: Ekiti = 0.795, Lagos = 0.795, Osun = 0.795 and means for descending order: Ekiti = 0.933, Lagos = 0.933, Osun = 0.933) respectively.

Discussion of the Findings

Hypothesis One: There is no significant difference in item discriminating indices of test items arranged in ascending, descending, and non-directional order

The results of the Kruskal–Wallis and ANOVA tests rejected this null hypothesis, revealing a significant effect of item arrangement on discrimination indices. The ascending order of item presentation yielded the highest discriminating power (mean = 95.50), followed by non-directional (68.01) and descending orders (62.99). This finding suggests that item sequencing has a significant impact on the psychometric quality of Physics test items, particularly in their ability to differentiate between high and low-performing students. This outcome aligns with previous studies that emphasize the influence of test structure on performance and validity (Quaigrain & Arhin, 2017; Abdellatif & Al-Shahrani, 2019). According to Cognitive Load Theory (Sweller, 2011; Sepp et al., 2019), presenting items in ascending order minimizes cognitive overload by allowing students to build confidence gradually, whereas presenting items in descending order can heighten anxiety and impede performance (Anaya et al., 2022; Tobia et al., 2022). The result corroborates Haasnoot et al. (2025), who reported that forward sequencing enhances discrimination by fostering smoother cognitive engagement. Conversely, descending and random sequencing may induce disorientation, lowering discriminative precision (Gu, 2010; Otero & Alonso, 2023). The study therefore affirms that item arrangement constitutes a vital determinant of test fairness and validity, consistent with psychometric standards that emphasize the optimization of item discrimination for accurate assessment (Kumar et al., 2021; Rezigalla et al., 2024). This also reinforces the position of Cappelleri et al. (2014) and Hughes (2018) that well-structured assessments must consider both cognitive processing demands and psychometric robustness to ensure valid interpretation of student ability.

Hypothesis Two: Gender will not significantly influence the item discriminating indices of Physics test items arranged in ascending, descending, and non-directional order

The analysis revealed no significant gender differences in item discrimination indices ($p = 0.713$) and no interaction between gender and test arrangement ($p = 0.802$). This finding suggests that the test items' discriminative power was not biased toward male or female students, indicating gender fairness in the assessment. This is consistent with recent evidence suggesting a diminishing gender gap in science assessment outcomes when test content and structure are equitable (Afolabi & Olajuyigbe, 2018; Burkholder & Salehi, 2022; Gil, 2024). The result contrasts with earlier concerns that test sequencing or item framing might favor male students in STEM

ISSN 2736-0091
E-ISSN 2737-0107

fields (Bowman et al., 2022; Cimpian et al., 2020). Instead, the current finding reflects improved balance in item construction, as recommended by Khasawneh and Khasawneh (2023), who argued that eliminating structural bias enhances assessment fairness and reliability. By confirming that gender does not significantly influence discrimination indices, the study reinforces the notion that well-designed Physics assessments can provide equitable measures of competence when cognitive and psychometric variables are properly controlled (Wrigley-Asante et al., 2023; Tseer et al., 2025). Thus, any observed performance differences in Physics are more likely attributable to instructional and environmental factors rather than test design biases (Rezai et al., 2022; Smith et al., 2023)

Hypothesis Three: State of origin will not significantly influence item discriminating indices of Physics test items arranged in ascending, descending, and non-directional order

The results also showed that the state of origin did not significantly affect the discrimination indices ($p = 0.648$) or interact significantly with the test arrangement ($p = 0.783$). This suggests that the regional background, whether Ekiti, Lagos, or Osun, did not significantly influence how well the test differentiated students. This outcome is consistent with Rezai et al. (2022) and Eden et al. (2024), who found that equitable assessment practices can neutralize the effects of demographic background when standardized conditions are maintained. Although differences in educational resources and exposure often vary across regions (Silverman et al., 2022), the uniform psychometric structure and standardized administration in this study appear to have minimized such disparities. This supports Apelehin et al. (2025) and Oyeronke and Adeoye (2024), who advocate for standardized item analysis and test validation procedures to ensure fairness and comparability across diverse populations. Overall, the finding reinforces the argument of Muskens et al. (2024) and Levy-Feldman (2025) that fairness in assessment derives from scientifically grounded test design rather than demographic uniformity. Hence, the quality of physics assessments depends primarily on structural validity, particularly item arrangement, rather than extraneous demographic factors.

Conclusion

The study concludes that the arrangement of test items significantly influences the discrimination indices of Physics multiple-choice questions in Nigerian secondary schools, with ascending order proving most effective in differentiating between high- and low-performing students. The Kruskal-Wallis test revealed a statistically significant difference in discrimination indices across the three arrangements ($H = 16.24$, $p < 0.05$), with post hoc analyses confirming that the ascending order format outperformed both the descending and non-directional formats. This aligns with Cognitive Load Theory, which posits that beginning with easier items reduces extraneous cognitive load, builds student confidence, and thereby enhances performance differentiation. In contrast, neither gender nor state of origin exerted a significant influence on item discrimination ($p = 0.713$ and $p = 0.648$, respectively), nor did they interact significantly with the test arrangement ($p = 0.802$ and $p = 0.783$, respectively). These findings suggest that structural aspects of test design are more critical to assessment fairness and validity than demographic variables. The absence of gender or regional bias indicates that a well-constructed, standardized Physics test can function equitably across diverse student populations, reinforcing the importance of psychometric rigor in test development.

Recommendations

Based on the findings of this study, the following recommendations are proposed for various stakeholders in education to enhance the fairness, validity, and psychometric quality of Physics assessments in Nigerian secondary schools:

1. **To Test Developers and Examination Bodies:** Test developers at the West African Examinations Council (WAEC), National Examinations Council (NECO), and school-based examination committees should adopt ascending item sequencing when designing Physics and other science tests. This approach enhances item discrimination, alleviates test anxiety, and promotes gradual cognitive engagement, thereby improving overall test effectiveness, measurement accuracy, and fairness.
2. **To Curriculum Planners and Policy Makers**: The Federal Ministry of Education and State Ministries of Education should integrate psychometric principles and cognitive load theory into curriculum policy documents. Specifically, assessment guidelines should mandate that tests undergo item analysis, including difficulty and discrimination indices, before being approved for use in schools.

ISSN 2736-0091
E-ISSN 2737-0107

3. **To Teacher Education Institutions**: Colleges of education and faculties of education in universities should embed training on modern test construction and psychometric evaluation into teacher education programmes. Pre-service teachers should learn to design, analyze, and interpret test items based on Classical Test Theory and cognitive sequencing principles to ensure equitable assessment practices.
4. **To Physics Teachers and School Assessment Committees**: Secondary school Physics teachers should routinely analyze test items after administration to evaluate discrimination and difficulty indices. This feedback loop will help teachers refine future assessments and improve their understanding of how item structure influences student performance and test reliability.
5. **To Students and Learner Support Units:** Guidance counselors and academic support units should teach students about test-taking strategies and cognitive sequencing, helping them understand how to approach assessments with increasing difficulty. This can boost self-efficacy, lessen anxiety, and enhance engagement during exams.

References

Abdellatif, H., & Al-Shahrani, A. M. (2019). Effect of blueprinting methods on test difficulty, discrimination, and reliability indices: Cross-sectional study in an integrated learning program. *Advances in Medical Education and Practice, 10*, 23–30. https://doi.org/10.2147/AMEP.S190827

Adaramoye, T.O., Omotoso, G.T., Olayiwola-Adedoja, T.O**.,** Dikko, B.O., & Iretor-Oscar, B.O. (2025). Prevalence, Perceived Causes and Consequences of Pre-marital Sex among Secondary School Students in Ekiti State. African Journal of Social and Behavioural Sciences, Vol. 15, No. 5.

Afolabi, F., & Olajuyigbe, O. A. (2018). Bridging gender gap in physics classrooms: Do teaching approaches matter? *KIU Journal of Social Sciences*, *4*(2), 69–80. https://ir.unilag.edu.ng/bitstreams/2db9a1a2-d775-4e22-8625-44a590eba146/download

Anaya, L., Iriberri, N., Rey-Biel, P., & Zamarro, G. (2022). Understanding performance in test taking: The role of question difficulty order. *Economics of Education Review*, *90*, 102293.

Apelehin, A. A., Imohiosen, C. E., Ajuluchukwu, P., Abutu, D. E., Udeh, C. A., & Iguma, D. R. (2025). Assessment and Evaluation for Social Improvement in Education: Strategies for Equity and Fairness. *International Journal of Social Science Exceptional Research. 2025b*, *4*(1), 119-125.

Ayanwoye, O. K., Falebita, O. S., Salami, O. O., Ogunjobi, A. O., Ajewole, P. I., Oyegbami, G. A., ... & Ogundare, A. A. (2025). Students' and Lecturers' Awareness, Knowledge, and Usage of Extended Reality Technology in Nigerian Universities. *International Journal of Technologies in Learning*, *32*(2), 141-167.

Bowman, N. A., Logel, C., LaCosse, J., Jarratt, L., Canning, E. A., Emerson, K. T., & Murphy, M. C. (2022). Gender representation and academic achievement among STEM-interested students in college STEM courses. *Journal of research in science teaching*, *59*(10), 1876-1900.

Burkholder, E., & Salehi, S. (2022). Exploring the pre-instruction gender gap in physics. *PloS one*, *17*(7), e0271184. https://doi.org/10.1371/journal.pone.0271184

Butler, A. C. (2018). Multiple-choice testing in education: Are the best practices for assessment also good for learning?. *Journal of Applied Research in Memory and Cognition*, *7*(3), 323-331.

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical therapeutics*, *36*(5), 648–662.

Chae, Y. M., Park, S. G., & Park, I. (2019). The relationship between classical item characteristics and item response time on computer-based testing. *Korean Journal of Medical Education, 31*(1), 1–9. https://doi.org/10.3946/kjme.2019.113

Cimpian, J. R., Kim, T. H., & McDermott, Z. T. (2020). Understanding persistent gender gaps in STEM. *Science*, *368*(6497), 1317-1319.

Eden, C. A., Onyebuchi, N. C., & Adeniyi, I. S. (2024). Cultural competence in education: Strategies for fostering inclusivity and diversity awareness. *International Journal of Applied Research in Social Sciences, 6*(3), 383–392. https://doi.org/10.51594/ijarss.v6i3.895

Esan, O. M., & Babalola, F. E. (2024). Influence of Personality Development on the Academic Performance of Secondary School Science Students in Ekiti State, Nigeria. *FUDMA Journal of Science Education (FUDJSE)*, *1*(1), 38-45.

ISSN 2736-0091
E-ISSN 2737-0107

FOWLER, A. (2022). *Views from the students' desks: How students experience and comprehend demand and difficulty in GCSE mathematics examination questions* (Doctoral dissertation, Durham University).

Gil, S. V. (2024). The influence of gender on academic performance and psychological resilience, and the relationship between both: Understanding the differences through gender stereotypes. *Trends in Psychology*. https://doi.org/10.1007/s43076-024-00370-7

Gorgun, G., Cai, M., & Demmans Epp, C. (2025,). Modeling Intrinsic Load: Can Item Difficulty Be a Reasonable Proxy?. In *International Conference on Artificial Intelligence in Education* (pp. 260-267). Cham: Springer Nature Switzerland

Gu, G. (2010). Nondirectional hypotheses. In *Encyclopedia of research design* (Vol. 0, pp. 910-910). SAGE Publications, Inc., https://doi.org/10.4135/9781412961288.n270

Haasnoot, C. W., van Krevel, C., & de Vaal, A. (2025). Ordered or scrambled: how the forward sequencing of multiple choice questions affects test item scores. *Assessment & Evaluation in Higher Education*, 1–16. https://doi.org/10.1080/02602938.2025.2568187

Hughes, D. J. (2018). Psychometric validity: Establishing the accuracy and appropriateness of psychometric measures. *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*, 751-779.

Kemper, C.J. (2020). Content Validity. In: Zeigler-Hill, V., Shackelford, T.K. (eds) Encyclopedia of Personality and Individual Differences. Springer, Cham. https://doi.org/10.1007/978-3-319-24612-3_1289

Khasawneh, Y., & Khasawneh, N. A. (2023). Achieving assessment equity and fairness: Identifying and eliminating bias in assessment tools and practices. https://www.researchgate.net/publication/373757288

Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India, 77*(Suppl 1), S85–S89. https://doi.org/10.1016/j.mjafi.2020.11.007

Langerock, N., Oberauer, K., Throm, E., & Vergauwe, E. (2025). The cognitive load effect in working memory: Refreshing the empirical landscape, removing outdated explanations. *Journal of Memory and Language*, *140*, 104558.

Levy-Feldman, I. (2025). The Role of Assessment in Improving Education and Promoting Educational Equity. *Education Sciences*, *15*(2), 224. https://doi.org/10.3390/educsci15020224

McNeish, D. (2024). Practical implications of sum scores being psychometrics' greatest accomplishment. *psychometrika*, *89*(4), 1148-1169.

Murphy, D. H., Little, J. L., & Bjork, E. L. (2023). The value of using tests in education as tools for learning, not just for assessment. *Educational Psychology Review*, *35*(3), 89.

Muskens, M., Frankenhuis, W. E., & Borghans, L. (2024). Math items about real-world content lower test-scores of students from families with low socioeconomic status. *NPJ Science of Learning, 9*(1), 19. https://doi.org/10.1038/s41539-024-00228-8

Odden, T. O. B., Marin, A., & Rudolph, J. L. (2021). How has science education changed over the last 100 years? An analysis using natural language processing. *Science Education*. https://doi.org/10.1002/sce.21623

Olayiwola-Adedoja, T. O., Chinedu, J. H., & Abidemi, O. (2024). Mindfulness-Based Cognitive Therapy: An Effective Counselling Tool for Enhancing Adolescents' Psychological Well-Being. *RUJMASS*, Vol. 10, 2

Olayiwola-Adedoja, T.O**.,** Ayodele, J.O., Olosunde, A.F., Akeredolu, I.F. (2025). Perceived causes and consequences of drug abuse among secondary school students in Ekiti State. African Journal of Educational Management, Teaching & Entrepreneurship Studies, Vol. 15, No. 1

Otero, I., & Alonso, P. (2023). Cognitive reflection test: The effects of the items sequence on scores and response time. *PLoS ONE, 18*(1), e0279982. https://doi.org/10.1371/journal.pone.0279982

Oyeronke, P., & Adeoye, M. A. (2024). Reassessing standardized tests: Evaluating their effectiveness in school performance measurement. *Curricula: Journal of Curriculum Development, 3*(2), 217–234. https://doi.org/10.17509/curricula.v3i2.74535

Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, *4*(1). https://doi.org/10.1080/2331186X.2017.1301013

Rezai, A., Namaziandost, E., Miri, M., & Kumar, T. (2022). Demographic biases and assessment fairness in classroom: Insights from Iranian university teachers. *Language Testing in Asia, 12*(1), 8. https://doi.org/10.1186/s40468-022-00157-6

ISSN 2736-0091
E-ISSN 2737-0107

Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhussein, A. B., Alfaifi, J., ALGhamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O. A., & Adam, M. I. E. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*, *24*(1), 445. https://doi.org/10.1186/s12909-024-05433-y

Robie, C., Meade, A. W., Risavy, S. D., & Rasheed, S. (2022). Effects of response option order on Likert-type psychometric properties and reactions. *Educational and Psychological Measurement, 82*(6), 1107–1129. https://doi.org/10.1177/00131644211069406

Rush, B. R., Rankin, D. C., & White, B. J. (2016). The impact of item-writing flaws and item complexity on examination item difficulty and discrimination value. *BMC Medical Education, 16*(1), 250. https://doi.org/10.1186/s12909-016-0773-3

Sepp, S., Howard, S. J., Tindall-Ford, S., Agostinho, S., & Paas, F. (2019). Cognitive load theory and human movement: Towards an integrated model of working memory. *Educational Psychology Review*, *31*(2), 293-317.

Silverman, D. M., Hernandez, I. A., & Destin, M. (2022). Educators' Beliefs About Students' Socioeconomic Backgrounds as a Pathway for Supporting Motivation. *Personality and Social Psychology Bulletin*, *49*(2), 215-232. https://doi.org/10.1177/01461672211061945 (Original work published 2023)

Silverman, D. M., Hernandez, I. A., & Destin, M. (2022). Educators' Beliefs About Students' Socioeconomic Backgrounds as a Pathway for Supporting Motivation. *Personality and Social Psychology Bulletin*, *49*(2), 215-232. https://doi.org/10.1177/01461672211061945 (Original work published 2023)

Smith, J., Meyer, F., & McClure, H. (2023). The persistence of bias in education: A call for research to move policy and practice from aspiration to results. *Policy Futures in Education*, *0*(0). https://doi.org/10.1177/14782103231180423

Tan, S. H., & Pang, J. S. (2023). Test anxiety: An integration of the test anxiety and achievement motivation research traditions. *Educational Psychology Review*, *35*(1), 13.

Tobia, V., Sacchi, S., Cerina, V., Manca, S., & Fornara, F. (2022). The influence of classroom seating arrangement on children's cognitive processes in primary school: The role of individual variables. *Current Psychology, 41*(9), 6522–6533. https://doi.org/10.1007/s12144-020-01154-9

Tseer, T., Ngmenkpieo, F., & Damwah, A. K. (2025). Gendered discriminatory norms and academic performance disparity among senior high school students in Northwestern Ghana. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2469401

Wrigley-Asante, C., Ackah, C. G., & Frimpong, L. K. (2023). Gender differences in academic performance of students studying Science, Technology, Engineering and Mathematics (STEM) subjects at the University of Ghana. *SN Social Sciences, 3*(1), Article 12. https://doi.org/10.1007/s43545-023-00608-8