AJEMATES
African Journal of Educational Management, Teaching and Entrepreneurship Studies
VOL.17 (2), January-April, 2026; https://ajemates.org

INFLUENCE OF TEST ITEM ARRANGEMENT ON DISTRACTOR EFFICIENCY IN MULTIPLE-CHOICE PHYSICS ASSESSMENTS AMONG SENIOR SECONDARY SCHOOL STUDENTS

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

ABSTRACT

*This study investigated the influence of test item arrangement on the distractive indices of multiple-choice Physics test items among Senior Secondary School students in Southwest Nigeria. A comparative research design was adopted for the study. The population comprised all Senior Secondary School Two (SSS II) students offering Physics in public secondary schools, from which a sample of 625 students was selected using a multistage sampling procedure. Three researcher-developed instruments—Ascending Order Performance Test in Physics (AOPTP), Descending Order Performance Test in Physics (DOPTP), and Non-directional Order Performance Test in Physics (NOPTP)—were used for data collection. Each instrument consisted of 50 multiple-choice items designed to assess various cognitive levels. The instruments were validated by experts in Tests and Measurement and Physics Education, while reliability was established using the split-half method. Data were analyzed using one-way Analysis of Variance (ANOVA), Kruskal-Wallis test, and two-way ANOVA at a 0.05 level of significance. The findings revealed that there was no significant difference in the distractive indices of Physics test items arranged in ascending, descending, and non-directional formats. Gender had no significant influence on distractive indices, and no significant interaction effect was found between gender and test item arrangement. However, test item arrangement had a significant main effect in some analyses, and a significant interaction effect was observed between state of origin and test item arrangement. The study concluded that while test item arrangement may not substantially affect distractor efficiency overall, contextual factors may influence outcomes. It was recommended that test developers focus more on item quality than sequencing.***

Keywords: Test Item Arrangement, Distractor Efficiency, Multiple-Choice Test, Physics Education, Secondary School Students

Introduction

Assessment plays a central role in the teaching and learning process, particularly in science education where the evaluation of students' cognitive abilities, conceptual understanding, and problem-solving skills is essential (De Luca, 2025). Among various assessment formats, multiple-choice tests

(MCTs) remain one of the most widely used tools due to their objectivity, ease of scoring, and ability to cover a broad range of content within a limited time (Oc&Hassan,2025). However, the effectiveness of multiple-choice tests depends not only on the quality of the questions but also on the functionality of their components, especially the distractors (Quansah et al.,2025). Distractors, which are the incorrect options provided alongside the correct answer, serve as critical indicators of item quality and play a significant role in differentiating between students of varying ability levels (Chauhan et al.,2025).

A well-constructed distractor should be plausible enough to attract students who lack sufficient understanding while being avoided by those who possess adequate knowledge (Bitew et al.,2025). The effectiveness of distractors is often measured using distractor indices, which reflect the extent to which incorrect options contribute to the discriminatory power of test items (Rezigalla et al.,2024). Poorly functioning distractors can lead to increased guessing, thereby reducing the reliability and validity of the test (Ludewig et al.,2023). Consequently, examining factors that influence distractor efficiency is essential for improving the quality of assessment instruments (Jin et al.,2023).

One such factor that has received limited attention in educational measurement is the arrangement of test items (Young et al.2025). Traditionally, test items in multiple-choice assessments may be presented in various formats, including ascending order of difficulty (from easiest to hardest), descending order (from hardest to easiest), or non-directional (random arrangement) (Rudolph et al.,2019). While these formats are often chosen based on practical considerations or examiner preference, emerging concerns suggest that item arrangement may influence students' test-taking behaviour, including their engagement, motivation, and susceptibility to guessing (Fink et al.,2019)

Theoretically, arranging items in ascending order of difficulty may help to build students' confidence as they begin with simpler questions, potentially reducing anxiety and encouraging persistence throughout the test (Otermans et al.,2025). In contrast, a descending order format may initially challenge students with more difficult questions, which could either stimulate deeper cognitive engagement or, conversely, lead to frustration and increased random guessing (Shcheglova et al.,2025). Non-directional arrangements, on the other hand, present a mixed sequence of item difficulty, which may prevent predictability but could also disrupt cognitive flow during test-taking (Rezigalla et al.,2024). These variations in item sequencing raise important questions about their potential impact on distractor functioning and overall test effectiveness (Gebremichael et al.,2025).

In the context of Physics education, where abstract concepts and problem-solving skills

are fundamental, the design of assessment instruments becomes even more critical (Nurjanah et al.,2025). Secondary school students often experience difficulties in mastering Physics concepts, and poorly structured assessments may further compound these challenges (Musengimana et al.,2025). Given that distractors are designed to reflect common misconceptions in Physics, any factor that alters how students interact with these distractors such as item arrangement warrants empirical investigation (Rezigalla et al.,2024).

Furthermore, student characteristics such as gender and state of origin may also interact with test design variables to influence assessment outcomes (Kumar,2025). Although prior research has explored gender differences in academic performance and test-taking strategies, findings remain inconclusive, particularly in science subjects (Kaedabi-Donkor et al.,2025). Similarly, contextual factors related to students' educational environments, such as geographical location, may affect exposure to instructional resources and learning experiences, thereby influencing how students respond to test items and their associated distractors (Palikara et al.,2025)

Despite the importance of these issues, there is a paucity of empirical studies examining the combined effects of test item arrangement and student characteristics on distractor efficiency in multiple-choice Physics assessments. Most existing research has focused on item difficulty, discrimination indices, and overall test reliability, with limited attention given to how the sequencing of items may subtly shape students' response patterns (Lamgarraj et al.,2024). Addressing this gap is crucial for advancing assessment practices and ensuring that test scores accurately reflect students' true abilities.

Therefore, this study seeks to investigate the influence of different test item arrangements; ascending, descending, and non-directional on the distractive indices of multiple-choice Physics test items among Senior Secondary School students. Additionally, the study examines whether gender and state of origin moderate this relationship. By providing empirical evidence on how item sequencing affects distractor performance, the study aims to contribute to the development of more valid, reliable, and fair assessment practices in Physics education. Ultimately, improving the quality of multiple-choice tests requires a comprehensive understanding of both item construction and test design factors. Insights from this study may guide educators, test developers, and curriculum planners in optimizing assessment strategies to enhance student evaluation and learning outcomes in Physics and other science disciplines.

Statement of the Problem

The effectiveness of multiple-choice tests in assessing students' academic performance largely depends on the quality of test items and the functionality of their distractors (Zoucha et al.,2025). In Physics education, where conceptual understanding and problem-solving are essential, poorly

functioning distractors can compromise the validity and reliability of assessment outcomes by increasing the likelihood of guessing and reducing the discriminatory power of test items (Lichtenberger et al.,2025). Despite the emphasis on item construction in educational measurement, less attention has been given to how the arrangement of test items may influence distractor efficiency and students' response behaviour(Quansah et al.,2025).

These arrangements are often adopted without empirical justification, even though they may affect students' motivation, confidence, and cognitive engagement during test-taking (Murphy et al.,2023). For instance, beginning with difficult items may discourage students and promote random guessing, while starting with easier items may enhance confidence but potentially influence response patterns in later questions (Debrum,2023). Such variations raise concerns about whether item arrangement could systematically affect distractive indices and, consequently, the fairness and accuracy of test scores (Holden&Tanenbaum,2023).

Furthermore, differences in student characteristics, such as gender and state of origin, may interact with test design features to influence how distractors function (Sterberg et al.2025). However, there is limited empirical evidence addressing these interactions within the context of Physics assessments. Therefore, the problem of this study is to determine whether test item arrangement significantly influences distractor efficiency and whether this effect varies across student characteristics, thereby ensuring more valid and equitable assessment practices.

Research Objectives

The main objective of this study is to examine the influence of test item arrangement on the distractive indices of multiple-choice Physics test items among Senior Secondary School students.

The specific objectives are to:

1. Determine whether there is a significant difference in the distractive indices of Physics test items when arranged in ascending, descending, and non-directional order.
2. Examine the influence of gender on the distractive indices of Physics test items arranged in ascending, descending, and non-directional formats.
3. Investigate the interaction effect of gender and test item arrangement on the distractive indices of Physics test items.
4. Determine whether state of origin influences the distractive indices of Physics test items arranged in ascending, descending, and non-directional formats.
5. Examine the interaction effect of state of origin and test item arrangement on the distractive indices of Physics test items.

Research Hypotheses

The following null hypotheses were formulated to guide the study and tested at 0.05 level of significance:

1. There is no significant difference in the distractive indices of Physics test items arranged in ascending, descending, and non-directional order.
2. Gender has no significant influence on the distractive indices of Physics test items arranged in ascending, descending, and non-directional formats.
3. There is no significant interaction effect between gender and test item arrangement on the distractive indices of Physics test items.
4. State of origin has no significant influence on the distractive indices of Physics test items arranged in ascending, descending, and non-directional formats.
5. There is no significant interaction effect between state of origin and test item arrangement on the distractive indices of Physics test items.

Conceptual and Theoretical Framework

The conceptual framework for this study is anchored on the relationship between test item arrangement and distractor efficiency in multiple-choice Physics assessments (Rezigalla et al.,2024). Test item arrangement operationalized as ascending, descending, and non-directional sequencing of items serves as the independent variable, while distractive indices represent the dependent variable (Esan et al.,2025). Distractors are designed to reflect common misconceptions and to differentiate between high- and low-performing students; thus, their effectiveness is central to test validity (Ozdemir &Toker,2025). The framework assumes that the order in which test items are presented may influence students' cognitive engagement, confidence, and test-taking strategies, which in turn affect how they interact with distractors (Polack&Miller,2022). Additionally, student characteristics such as gender and state of origin are treated as moderating variables that may shape response patterns and influence the overall effectiveness of distractors (Cedillo et al.,2025).

The study is theoretically underpinned by Classical Test Theory (CTT), which emphasizes the importance of item-level statistics such as difficulty index, discrimination index, and distractor efficiency in determining the quality of a test (Tutz,2025). According to CTT, each test item contributes to the overall reliability and validity of the assessment, and poorly functioning distractors can increase measurement error by encouraging guessing (Faria&Mariranda,2026). Within this framework, distractive indices provide insight into how well incorrect options perform in attracting less knowledgeable students (Deng et al.,2024). The sequencing of test items may indirectly affect these indices by altering students' psychological states such as fatigue, anxiety, or confidence thereby influencing their likelihood of

selecting certain distractors (Haasnvot et al.,2026). Hence, CTT provides a useful basis for understanding how variations in test design, including item arrangement, can impact measurement precision (Meguellati et al.,2024).

In addition, the study draws on cognitive load theory and test-taking behaviour models to explain how item arrangement may affect students' performance (De Bruin et al.,2025). Cognitive load theory posits that learners have limited working memory capacity, and the organization of tasks can either facilitate or hinder information processing (Gkintoni et al.,2025). An ascending arrangement may reduce cognitive load at the beginning of the test, allowing students to build momentum, whereas a descending arrangement may impose a higher initial cognitive demand, potentially leading to disengagement or increased guessing (Ouwehand et al.,2025). Test-taking behaviour models further suggest that students adapt their strategies based on perceived task difficulty and test structure, which may influence how they respond to distractors (Hatti&leary,2025). Together, these theoretical perspectives provide a comprehensive framework for examining how test item arrangement interacts with individual differences to affect distractor efficiency in Physics assessments.

Review of Related Literature

Multiple-choice testing has remained a dominant assessment format in science education due to its efficiency in measuring a wide range of cognitive abilities (Clark et al.,2025). However, the quality of such assessments depends heavily on the construction and performance of test items, particularly the effectiveness of distractors (Skedsmo &Huber,2025). Empirical studies in educational measurement consistently emphasize that functional distractors should be plausible and capable of attracting students with partial knowledge while being rejected by more knowledgeable learners (Levy-Feldman,2025). Research on item analysis has shown that poorly constructed distractors often fail to contribute meaningfully to item discrimination, thereby increasing the chances of guessing and reducing test validity (Quaigrain&Arhin,2017). In Physics education, where misconceptions are common, well-designed distractors are especially important as they help reveal students' conceptual misunderstandings and enhance the diagnostic value of assessments (Guera-Reyes et al.,2024).

Beyond item construction, recent studies have begun to explore how test design features, such as item sequencing, may influence students' performance and response behaviour(Lions et al.,2024). Although limited, existing research suggests that the order in which test items are presented can affect students' confidence, motivation, and cognitive engagement (Vu et al.,2022). For instance, arranging items in ascending order of difficulty has been associated with improved test persistence and reduced anxiety, as students are able to build confidence by initially responding to simpler

questions (Jerrim,2023). Conversely, presenting difficult items at the beginning of a test may lead to frustration and increased reliance on guessing strategies (Bieleke et al.,2023). Non-directional arrangements, which mix items of varying difficulty, have been argued to reduce predictability but may also disrupt the cognitive flow of test-taking (Wong et al.,2023). Despite these insights, empirical evidence on how item arrangement specifically affects distractor efficiency remains scarce.

Furthermore, studies examining the role of learner characteristics such as gender and geographical background in test performance have produced mixed findings. (Mpiti et al.,2025) Some research reports no significant gender differences in response patterns to multiple-choice items, while others suggest that contextual and socio-cultural factors may influence test-taking strategies (Vera GL,2024). Similarly, variations in educational resources and instructional quality across regions may shape students' familiarity with test formats and their interaction with distractors (Goyibova et al.,2025). However, few studies have integrated these variables with test design factors such as item arrangement (Suleiman et al.,2024). This gap highlights the need for comprehensive investigations that consider both item-level characteristics and learner variables in understanding assessment outcomes. The present study therefore contributes to the literature by examining how test item arrangement, alongside gender and state of origin, influences distractor efficiency in multiple-choice Physics assessments.

Methodology

This study adopted a comparative research design to investigate the influence of test item arrangement on the distractive indices of multiple-choice Physics test items among Senior Secondary School students. The comparative design was considered appropriate because it enables the researcher to examine differences among groups under naturally existing conditions without manipulation of variables. It also allows for meaningful comparison of outcomes across different test formats, thereby providing empirical evidence on how item arrangement influences distractor efficiency.

The population of the study comprised all Senior Secondary School Two (SSS II) science students offering Physics in public secondary schools in Southwest Nigeria. SSS II students were selected because they are assumed to have covered a substantial portion of the Physics curriculum and possess sufficient foundational knowledge required for the study. From this population, a sample of 625 students was drawn from nine public secondary schools across three states in the region. A multistage sampling procedure was employed. In the first stage, three states were selected using simple random sampling. In the second stage, three public secondary schools were purposively selected from each state based on the criterion that each school had at least 100 SSS II students offering Physics. Subsequently, a general Physics test

consisting of 25 items was administered to ensure homogeneity of the sample. Based on performance, 75 students who met the predetermined cut-off mark were selected from each school. Finally, stratified random sampling was used to assign the selected schools into three groups corresponding to the three test item arrangements: ascending, descending, and non-directional formats.

Three instruments were developed and used for data collection: the Ascending Order Performance Test in Physics (AOPTP), the Descending Order Performance Test in Physics (DOPTP), and the Non-directional Order Performance Test in Physics (NOPTP). Each instrument consisted of two sections: Section A captured respondents' demographic information such as school and gender, while Section B contained 50 multiple-choice items with five response options each. The items covered various cognitive levels based on Bloom's taxonomy, including knowledge, comprehension, application, analysis, synthesis, and evaluation. The AOPTP arranged items from the simplest to the most difficult, the DOPTP presented items from the most difficult to the simplest, and the NOPTP followed no specific order. The arrangement of items in the AOPTP and DOPTP was guided by item difficulty indices obtained through prior item analysis, while the NOPTP followed the sequence of the syllabus.

The validity of the instruments was ensured through expert review. Face and content validity were established by specialists in Tests and Measurement, Physics Education, and the researcher's supervisor, who evaluated the clarity, relevance, and coverage of the test items. Reliability of the instruments was determined using the split-half method. A pilot study was conducted with 30 students outside the sampled schools, and the data obtained were analyzed to determine the internal consistency of the instruments, ensuring that they were suitable for the main study. The administration of the instruments was carried out by the researcher with the assistance of trained research assistants in each of the selected schools. The assistants were briefed on the procedures to ensure uniformity in test administration and data collection. The instruments were administered under standardized conditions, and completed scripts were collected immediately after the exercise to prevent data loss.

Data collected were analyzed using both descriptive and inferential statistics. The hypotheses were tested at a 0.05 level of significance. Hypothesis one was analyzed using one-way Analysis of Variance (ANOVA), complemented by the non-parametric Kruskal-Wallis test to account for any violations of normality assumptions. Hypotheses two and three were tested using two-way ANOVA to examine the main and interaction effects of gender and state of origin with test item arrangement on distractive indices. These statistical techniques were considered appropriate for determining significant differences and

AJEMATES
African Journal of Educational Management, Teaching and Entrepreneurship Studies
VOL.17 (2), January-April, 2026; https://ajemates.org

interactions among the variables under investigation.

Results

All the hypotheses were tested using Analysis of Variance (ANOVA). Hypotheses 1 was tested using One-way Analysis of Variance, hypotheses 2 and 3 were tested using Two-way Analysis of Variance (2-WAY ANOVA). All the hypotheses were tested at 0.05 level of significance.

Results and Interpretation

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

To test this hypothesis, the One-way ANOVA was employed (Table 2) with Scheffe Post-hoc test (Table 3) for Pairwise comparison. A non-parametric tool known as Kruskal Wallis test, which is equivalent to one-way analysis of variance (ANOVA) was performed. Table 6a presented the result as follow

Table 1. Kruskal Wallis statistics for item distractive indices of physics test item arrangement

Statistics	Guessing (c)
Kruskal-Wallis H	2.79
N	150
Df	2
Asymp. Sig.	0.25

Table 2: Analysis of Variance for item distraction indices of physics test item arrangement

	Sum of squares	Df	Mean square	F	Sig.
Between Groups	.003	1	.003	.664	.418
Within Group	.3396	.78	.005		
Total	.399	.79			

Table 2 showed the non-parametric statistic comparison of item distractive indices of the test arrangement (non-directional order, ascending order and descending order) of multiple-choice physics test. It was remarked that there was no significant difference among the vulnerability to guessing/distractive of the items of various test arrangement of physics multiple-choice with (Kruskal-Wallis H Statistic = 2.79, $p > 0.05$). This implies that distractive power among different test arrangement was the same.

Hypothesis 2: Gender of students will not significantly influence item distractive indices of Physics test items arranged in ascending, descending, and non- directional order of format.

To test this hypothesis, values for gender, physics test arrangement and item distractive

AJEMATES
African Journal of Educational Management, Teaching and Entrepreneurship Studies
VOL.17 (2), January-April, 2026; https://ajemates.org

indices for different test form were subjected to two-way analysis of variance (ANOVA). Table 9a presented the statistics.

Table 3: Summary of Two-Analysis of Variance of Gender and Test Arrangement on Item distractive indices

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	42.365a	5	8.473	1791.668	0.000
Intercept	80.902	1	80.902	17107.171	0.000
Gender	12.358	1	0.095	0.047	0.829
Physics Test Arrangement	37.560	2	18.780	3971.171	0.000
Gender * Physics Test Arrangement	11.001	2	0.061	0.133	0.876
Error	3.164	669	0.005		
Total	138.86	675			
Corrected Total	45.529	674			

a R Squared = .931 (Adjusted R Squared = .930)
b Dependent Variable: Item distractive indices

Table 3 presented a two-way ANOVA conducted to examine the effect of gender and physics test arrangement on item distractive indices. Table 4 remarked that there was no significant interaction between the effects of gender and physics test arrangement on item distractive indices, $F_{(2, 669)} = 0.133$, $p = 0.876$. Also, the result showed that was no statistically significant difference in the item distractive indices between gender ($p = 0.829$), but there were statistically significant differences between test arrangement ($p < 0.05$). This was also verified using estimated marginal means descriptive statistics, as presented in Table 5.

Table 4: Estimated Marginal Means for Gender and Test Arrangement on Item Distractive Indices.

Gender	Physics Test Arrangement	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	Non-directional Order	0.727	0.008	0.712	0.742
	Ascending Order	0.196	0.005	0.185	0.207
	Descending Order	0.192	0.006	0.181	0.203
2	Non-directional Order	0.726	0.006	0.714	0.737
	Ascending Order	0.196	0.008	0.18	0.212
	Descending Order	0.197	0.008	0.181	0.213

AJEMATES
African Journal of Educational Management, Teaching and Entrepreneurship Studies
VOL.17 (2), January-April, 2026; https://ajemates.org

Table 4: showed that there was no mean difference between male and female item distractive indices when various test arrangement format was administered to them with (means for non-directional order: Male = 0.727, female= 0.726; means for ascending order: Male = 0.196, female= 0.196; and means for descending order: male=0.192; female = 0.197) respectively.

Hypothesis 3: State of Origin of students will not significantly influence item distractive indices of Physics test items arranged in ascending, descending, and non- directional order.

To test this hypothesis, values for State of origin, physics test arrangement and item distractive indices for different test form were subjected to two-way analysis of variance (ANOVA). Table 6 presented the statistics.

Table 5: Summary of Two-Analysis of Variance of State of origin and Test Arrangement on Item distractive indices

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	42.435a	8	5.304	1141.869	0.000
Intercept	93.331	1	93.331	20091.19	0.000
State Origin	0.024	2	0.012	2.563	0.078
Physics Test Arrangement	42.364	2	21.182	4559.789	0.000
State Origin * Physics Test Arrangement	0.048	4	0.012	2.563	0.037
Error	3.094	666	0.005		
Total	138.86	675			
Corrected Total	45.529	674			

a R Squared = .932 (Adjusted R Squared = .931)
b Dependent Variable: Item distractive indices

Table 5 presented a two-way ANOVA conducted to examine the effect of State of origin and physics test arrangement on item distractive indices. Table 12a remarked that there was significant interaction between the effects of State of origin of student and physics test arrangement on item distractive indices, $F_{(2, 666)} = 2.563$, $p < 0.05$. This implies that vulnerability to guessing the items correctly varies across the States when different test form is attempted. Also, the result showed that was no statistically significant difference in the item distractive indices between State of origin ($p = 0.078$), but there were statistically significant differences between test arrangement ($p < 0.05$). This was also verified using estimated marginal means descriptive statistics, as presented in Table 12b. Table 6: Estimated

AJEMATES
African Journal of Educational Management, Teaching and Entrepreneurship Studies
VOL.17 (2), January-April, 2026; https://ajemates.org

Marginal Means for State of origin and Test Arrangement on Item Distractive Indices.

State Origin	Physics Test Arrangement	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Ekiti	Non-directional Order	0.723	0.008	0.708	0.739
	Ascending Order	0.196	0.008	0.181	0.211
	Descending Order	0.193	0.008	0.178	0.209
Lagos	Non-directional Order	0.749	0.008	0.734	0.765
	Ascending Order	0.196	0.008	0.181	0.211
	Descending Order	0.193	0.008	0.178	0.209
Osun	Non-directional Order	0.706	0.008	0.69	0.721
	Ascending Order	0.196	0.008	0.181	0.211
	Descending Order	0.193	0.008	0.178	0.209

Table 6: showed that there was no mean difference between Ekiti, Lagos and Osun item distractive indices when various test arrangement format was administered to them with (means for non-directional order: Ekiti = 0.723, Lagos = 0.749, Osun = 0.706; means for ascending order: Ekiti = 0.196, Lagos = 0.196, Osun = 0.196 and means for descending order: Ekiti = 0.193, Lagos = 0.193, Osun = 0.193) respectively. X

Discussions

This study examined the influence of test item arrangement on the distractive indices of multiple-choice Physics test items among senior secondary school students. The findings provide important insights into the role of test design in educational measurement, particularly regarding distractor efficiency, which is a critical indicator of item quality within the framework of Classical Test Theory (Rezigalla et al., 2024; Tutz, 2025).

The first major finding revealed that there was no significant difference in the distractive indices of test items arranged in ascending, descending, and non-directional formats. This suggests that item sequencing does not substantially influence how distractors function in multiple-choice assessments. This finding aligns with prior research emphasizing that distractor effectiveness is primarily determined by the quality, plausibility, and conceptual relevance of the distractors rather than the structural arrangement of test items (Chauhan et al., 2023; Ludewig et al., 2023; Bitew et al., 2025). Well-constructed distractors are designed to reflect common misconceptions and attract less knowledgeable students regardless of their position within a test

(Özdemir & Toker, 2025; Olayiwola-Adedoja et al.,2025).

Furthermore, this result supports the assertion that distractor efficiency is more closely tied to item-level characteristics such as difficulty and discrimination indices than to test-level design features (Quaigrain & Arhin, 2017; Olayiwola-Adedoja et al.,2024; Gebremichael et al., 2025). Although theoretical perspectives such as cognitive load theory suggest that item sequencing may influence students' cognitive engagement and test-taking behaviour (De Bruin et al., 2025; Ouwehand et al., 2025), the present findings indicate that such effects may not extend to the functioning of distractors. This may be because distractor selection is more directly related to students' conceptual understanding than to their momentary psychological state during test-taking.

However, the finding contrasts with some studies that suggest item order can influence students' emotional responses, confidence, and performance (Bieleke et al., 2023; Shcheglova et al., 2025; Adaramoye et al.,2025a). A plausible explanation for this discrepancy is that while item arrangement may affect overall performance or test experience, it does not necessarily alter the relative attractiveness of distractors within individual items. In other words, even if students feel more confident or anxious depending on item order, their misconceptions and hence their likelihood of choosing specific distractors remain relatively stable.

The second major finding indicated that gender had no significant influence on distractive indices, and there was no interaction effect between gender and test item arrangement. This suggests that both male and female students responded similarly to distractors across all test formats. This finding supports previous studies that report minimal or no gender differences in test-taking strategies and item response behaviour in science education (Adaramoye et al.,2025b; Kaedabi-Donkor et al., 2025; Mpiti et al., 2025). It also reinforces the notion that well-constructed multiple-choice tests can provide equitable assessment conditions when bias is minimized at the item level (Holden & Tanenbaum, 2023; Sanni et al.,2024).

The absence of gender differences in distractor functioning further implies that misconceptions in Physics; which distractors are designed to capture are not significantly gender specific. This aligns with research suggesting that learning challenges in science subjects are more strongly influenced by instructional quality and prior knowledge than by gender (Esan&Babalola,2024; Shaibu et al.,2024; Musengimana et al., 2025). Consequently, the findings provide empirical support for the fairness of multiple-choice Physics assessments when items are properly constructed.

In contrast, the study found a significant interaction effect between state of origin and test item arrangement, although the main effect of state of origin was not significant.

This indicates that while students from different states did not differ significantly in overall distractor performance, the influence of item arrangement varied across contexts. This finding highlights the importance of contextual and environmental factors in shaping students' test-taking behaviour (Palikara et al., 2025; Suleiman et al., 2024). Differences in instructional practices, exposure to assessment formats, and educational resources across states may interact with test design features to influence how students engage with test items.

This result is consistent with literature suggesting that educational outcomes are influenced by a complex interplay of individual and contextual variables (Kumar, 2025; Cedillo et al., 2025; Akomolafe et al.,2025). It also supports the argument that assessment practices cannot be entirely separated from the socio-educational environments in which they are implemented (Skedsmo & Huber, 2025; Ojobola et al.,2025; Ayanwoye et al.,2025). The significant interaction effect implies that although item arrangement may not have a universal impact, its effects may be context-dependent, thereby warranting careful consideration in diverse educational settings.

Additionally, the findings contribute to the broader discourse on assessment design by reinforcing the central role of item construction over test formatting. Studies on distractor efficiency consistently emphasize that poorly functioning distractors increase guessing and reduce test validity (Jin et al., 2022; Rezigalla et al., 2024). The current study extends this understanding by demonstrating that even when item arrangement varies, the fundamental determinant of distractor performance remains the quality of the distractors themselves.

From a theoretical perspective, the findings provide empirical support for Classical Test Theory, which posits that item characteristics are the primary determinants of test reliability and validity (Faria & Miranda, 2026; Meguellati et al., 2024). While cognitive load theory and test-taking behaviour models offer valuable insights into how students interact with assessments, their influence appears to be more pronounced at the level of overall performance rather than at the level of distractor functioning.

Moreover, the study underscores the importance of focusing on higher-order assessment practices that promote meaningful learning. Research has shown that testing can serve as a powerful tool for enhancing learning when items are designed to challenge students' understanding and reveal misconceptions (Murphy et al., 2023; Polack & Miller, 2022). Effective distractors play a key role in this process by prompting students to critically evaluate alternatives, thereby supporting deeper cognitive engagement.

In summary, the findings of this study suggest that test item arrangement has limited influence on distractor efficiency, while

contextual factors may moderate its effects. The results reinforce the importance of high-quality item construction and provide evidence for the robustness and fairness of multiple-choice assessments when properly designed. These insights contribute to the ongoing effort to improve assessment practices in Physics education and beyond, ensuring that tests accurately measure students' knowledge and support meaningful learning outcomes.

Conclusion

The study investigated the influence of test item arrangement on the distractive indices of multiple-choice Physics test items among Senior Secondary School students. The findings revealed that item arrangement—whether ascending, descending, or non-directional—did not significantly affect distractor efficiency. This suggests that the functionality of distractors is largely dependent on the quality of item construction rather than the sequence in which items are presented. Furthermore, gender was found to have no significant influence on distractive indices, indicating that multiple-choice Physics assessments are relatively fair and unbiased across male and female students. Although state of origin did not show a significant main effect, its interaction with test item arrangement was significant, highlighting the role of contextual factors in shaping students' response behaviour.

Overall, the study concludes that effective distractor construction remains the most critical factor in ensuring the validity and reliability of multiple-choice assessments. While test item arrangement may have minimal direct impact, its interaction with contextual variables suggests that it should not be entirely overlooked in test design. These findings contribute to the field of educational measurement by providing empirical evidence that supports the robustness of distractor functioning across different test formats. Consequently, educators and test developers should prioritize the development of high-quality test items while maintaining awareness of contextual influences on student performance.

Recommendations

Based on the findings of this study, several recommendations are proposed to enhance the quality of multiple-choice assessments in Physics. First, test developers and educators should prioritize the construction of high-quality distractors that are plausible and based on common student misconceptions. Since distractor efficiency was found to be independent of item arrangement, greater emphasis should be placed on item writing skills and rigorous item analysis procedures.

Second, although item arrangement did not significantly influence distractor indices, it is advisable for educators to adopt test formats that promote positive test-taking experiences. For instance, arranging items in a manner that builds student confidence, such as starting with moderately easy questions, may help reduce anxiety and improve engagement.

Third, educational stakeholders should provide regular training for teachers in assessment design, particularly in the development and evaluation of multiple-choice items. Such training will ensure that teachers are equipped with the necessary skills to construct valid and reliable test instruments.

Finally, policymakers and curriculum planners should ensure uniformity in assessment practices across schools to minimize contextual disparities that may interact with test design variables. By implementing these recommendations, the overall quality, fairness, and effectiveness of Physics assessments can be significantly improved.

Limitations of the Study

Despite the contributions of this study, certain limitations should be acknowledged. First, the study was limited to Senior Secondary School Two students in public secondary schools in Southwest Nigeria. As such, the findings may not be generalizable to students in other regions, private schools, or different educational levels. Differences in curriculum implementation, instructional quality, and student characteristics across regions may influence the applicability of the results.

Second, the study focused solely on multiple-choice Physics test items, which may limit the generalizability of the findings to other subject areas or assessment formats such as essay or practical-based tests. The nature of Physics as a subject, with its emphasis on problem-solving and conceptual understanding, may also influence how students interact with distractors differently compared to other disciplines.

Third, the study relied on researcher-developed instruments, which, although validated and tested for reliability, may still be subject to inherent biases or limitations in item construction. Additionally, the use of a single method of reliability estimation (split-half) may not fully capture the consistency of the instruments.

Finally, other potential variables such as students' prior knowledge, test anxiety, and motivation were not controlled for, yet they may have influenced students' responses to distractors and overall test performance.

Implications for Counselling

The findings of this study have important implications for counselling, particularly in the educational context. School counsellors play a crucial role in supporting students' academic performance and test-taking behaviours. The result that test item arrangement does not significantly influence distractor efficiency suggests that students' performance in multiple-choice tests is less about the structure of the test and more about their level of preparation and understanding. Therefore, counsellors should focus on helping students develop effective study habits, conceptual understanding, and test-taking strategies rather than attributing poor performance to test format.

Additionally, the absence of gender differences in distractor indices indicates that both male and female students have similar capacities in handling multiple-choice assessments. Counsellors can use this insight to challenge stereotypes and promote equal academic expectations among students, particularly in subjects like Physics where gender biases may exist.

The significant interaction between state of origin and test item arrangement highlights the influence of contextual factors on students' learning experiences. Counsellors should therefore adopt a holistic approach that considers students' backgrounds, including their educational environment and exposure to assessment practices. By providing tailored guidance and interventions, counsellors can help students overcome contextual challenges and improve their academic outcomes.

REFERENCES

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.

Adaramoye, T. O., Adewoye, S. E., & Olayiwola-Adedoja, T. O. (2025). Parental influence on resilience against premarital sex and pornography engagement among teenagers in secondary schools. *EUREKA: Social and Humanities*, (4), 41-51

Akomolafe O.D, Ambode A.B, & Esan O.M (2025). Predicting Engagement in Computer Mediated Professional Development Using Asynchronous Online Discussion: A Case Study of Postgraduate Students at Federal University Oye Ekiti. Unizik Journal of Educational Research and Policy Studies 19(4) 177-189

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

Bieleke, M., Goetz, T., Krannich, M., Roos, A. L., & Yanagida, T. (2023). Starting tests with easy versus difficult tasks: Effects on appraisals and emotions. *The Journal of Experimental Education*, *91*(2), 317–335. https://doi.org/10.1080/00220973.2021.1947764

Bitew, S. K., Deleu, J., Develder, C., & Demeester, T. (2025). Distractor generation for multiple-choice questions with predictive prompting

and large language models. In R. Meo & F. Silvestri (Eds.), *Machine learning and principles and practice of knowledge discovery in databases (ECML PKDD 2023)* (Communications in Computer and Information Science, Vol. 2134). Springer. https://doi.org/10.1007/978-3-031-74627-7_4

Cedillo, Y. E., Davis, E., Durham, L., Kelly, T., Kennedy, R., Smith, D. L., Jr., & Fernández, J. R. (2025). The effect of sociodemographic characteristics, academic factors, and individual health behaviors on psychological well-being among college students. *Journal of American College Health*, *73*(8), 3000–3010. https://doi.org/10.1080/07448481.2024.2355160

Chauhan, G. R., Chauhan, B. R., Vaza, J. V., & Chauhan, P. R. (2023). Relations of the number of functioning distractors with the item difficulty index and the item discrimination power in multiple-choice questions. *Cureus*, *15*(7), e42492. https://doi.org/10.7759/cureus.42492

Clark, S. A., Rivers, M. L., & Overono, A. L. (2025). The role of answer justification in multiple-choice testing: Effects on performance and metacognitive accuracy. *Behavioral Sciences*, *15*(4), 477. https://doi.org/10.3390/bs15040477

De Bruin, A. B. H. (2023). Dealing with desirable difficulties: Supporting students to accept, reduce, or silence effort. *Medical Science Educator*, *33*(Suppl 1), 5–9. https://doi.org/10.1007/s40670-023-01911-y

De Bruin, A. B. H., Janssen, E. M., Waldeyer, J., et al. (2025). Cognitive load and challenges in self-regulation: An introduction and reflection on the topical collection. *Educational Psychology Review*, *37*, 65. https://doi.org/10.1007/s10648-025-10042-2

DeLuca, C. (2025). Assessment in education: Looking back, looking forward. *Assessment in Education: Principles, Policy & Practice*, *32*(1), 1–4. https://doi.org/10.1080/0969594X.2025.2485702

Deng, L., Zhou, Y., & Broadbent, J. (2024). Distraction, multitasking and self-regulation inside university classroom. *Education and Information Technologies*, *29*, 23957–23979. https://doi.org/10.1007/s10639-024-12786-w

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. 1(1) 38-45

Esan, O. M., Bandele, S. O., Babalola, F. E., & Godwin, A. (2025). Influence of test item arrangement and demographic variables on the discrimination indices of physics

multiple-choice items in Nigerian secondary schools. *Kogi Journal of Education and Pedagogy, 4*(2), 54–63.

Faria, A., & Miranda, G. L. (2026). Operationalising CTT and IRT in spreadsheets: A methodological demonstration for classroom assessment. *Analytics, 5*(1), 12. https://doi.org/10.3390/analytics5010012

Fink, M. C., Kaltefleiter, L. J., Reis, I., & Ertl, B. (2025). Format-specific expectations explain preferences and perceived fairness: A study on four common examination formats. *Assessment & Evaluation in Higher Education, 50*(7), 1066–1086. https://doi.org/10.1080/02602938.2025.2501689

Gebremichael, M. W., Baraki, B., Mehari, M. A., et al. (2025). Item analysis of multiple-choice questions from assessment of health sciences students, Tigray, Ethiopia. *BMC Medical Education, 25*, 441. https://doi.org/10.1186/s12909-025-06904-6

Gkintoni, E., Antonopoulou, H., Sortwell, A., & Halkiopoulos, C. (2025). Challenging cognitive load theory: The role of educational neuroscience and artificial intelligence in redefining learning efficacy. *Brain Sciences, 15*(2), 203. https://doi.org/10.3390/brainsci15020203

Goyibova, N., Muslimov, N., Sabirova, G., Kadirova, N., & Samatova, B. (2025). Differentiation approach in education: Tailoring instruction for diverse learner needs. *MethodsX, 14*, 103163. https://doi.org/10.1016/j.mex.2025.103163

Guerra-Reyes, F., Guerra-Dávila, E., Naranjo-Toro, M., Basantes-Andrade, A., & Guevara-Betancourt, S. (2024). Misconceptions in the learning of natural sciences: A systematic review. *Education Sciences, 14*(5), 497. https://doi.org/10.3390/educsci14050497

Haasnoot, C. W., van Krevel, C., & de Vaal, A. (2026). Ordered or scrambled: How the forward sequencing of multiple-choice questions affects test item scores. *Assessment & Evaluation in Higher Education, 51*(2), 247–262. https://doi.org/10.1080/02602938.2025.2568187

Hattie, J., & O'Leary, T. (2025). Learning styles, preferences, or strategies? An explanation for the resurgence of styles across many meta-analyses. *Educational Psychology Review, 37*, 31. https://doi.org/10.1007/s10648-025-10002-w

Holden, L. R., & Tanenbaum, G. J. (2023). Modern assessments of intelligence must be fair and equitable. *Journal of Intelligence, 11*(6), 126. https://doi.org/10.3390/jintelligence11060126

Jerrim, J. (2023). Test anxiety: Is it associated with performance in high-stakes examinations? *Oxford Review of Education, 49*(3), 321–341. https://doi.org/10.1080/03054985.2022.2079616

Jin, K.-Y., Siu, W.-L., & Huang, X. (2022). Exploring the impact of random guessing in distractor analysis. *Journal of Educational Measurement, 59*(1), 43–61. https://doi.org/10.1111/jedm.12310

Kaedabi-Donkor, R., Owusu Amponsah, M., Yusuf Dramanu, B., & Britwum, F. (2025). Gender and age variations in academic self-concept and academic achievement of distance learners. *Cogent Education, 12*(1). https://doi.org/10.1080/2331186X.2025.2493013

Kumar, A. (2025). Predicting student performance through demographic attributes. In S. Mahajan, Á. Rocha, A. K. Pandit, & P. Chawla (Eds.), *Smart systems: Engineering and managing information for future success* (Vol. 22). Springer. https://doi.org/10.1007/978-3-031-76152-2_13

Lamgarraj, M., Joiron, C., Parant, A., & Dequen, G. (2024). Exploring item difficulty prediction: A data-driven approach. In A. Sifaleras & F. Lin (Eds.), *Generative intelligence and intelligent tutoring systems* (Lecture Notes in Computer Science, Vol. 14798). Springer. https://doi.org/10.1007/978-3-031-63028-6_36

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

Lichtenberger, A., Hofer, S. I., Stern, E., et al. (2025). Enhanced conceptual understanding through formative assessment: Results of a randomized controlled intervention study in physics classes. *Educational Assessment, Evaluation and Accountability, 37*, 5–33. https://doi.org/10.1007/s11092-024-09445-6

Lions, S., Blanco, M. P., Dartnell, P., Monsalve, C., Ortega, G., & Lemarié, J. (2024). Item-writing guidelines on response option placement: A systematic review. *Applied Measurement in Education, 37*(4), 289–304. https://doi.org/10.1080/08957347.2024.2424551

Ludewig, U., Schwerter, J., & McElvany, N. (2023). The features of plausible but incorrect options: Distractor plausibility in synonym-based vocabulary tests. *Journal of Psychoeducational Assessment, 41*(7), 711–731.

Meguellati, S., Samia, A., Ferhat, A., & Djelloul, A. (2024). A critical analysis of the use of classical test theory in psychological testing: A

comparison with item response theory. *Pakistan Journal of Life & Social Sciences*, *22*(2).

Mpiti, V. S., Ncanywa, T., & Asaleye, A. J. (2025). Do educators' demographic characteristics drive learner academic performance? Examining the role of gender, qualifications, and experience. *Education Sciences*, *15*(4), 487. https://doi.org/10.3390/educsci15040487

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*, 89. https://doi.org/10.1007/s10648-023-09808-3

Musengimana, T., Yadav, L. L., Uwamahoro, J., et al. (2025). Instructional strategies for enhancing students' problem-solving skills in physics: A systematic review. *Discover Education*, *4*, 380. https://doi.org/10.1007/s44217-025-00733-x

Nurjanah, S., Sultan, J., Arriza, L., et al. (2025). Assessment in physics education research: Trends, patterns, and future directions. *Review of Education*, *13*, e70043. https://doi.org/10.1002/rev3.70043

Oc, Y., & Hassen, H. (2025). Comparing the effectiveness of multiple-answer and single-answer multiple-choice questions in assessing student learning. *Marketing Education Review*, *35*(1), 44–57. https://doi.org/10.1080/10528008.2024.2417106

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

Ojobola F.B, Esan O.M, Bada A. & Babalola F.E (2025) The Potential Impact of Entrepreneurial Skills on University Pre-Service Chemistry Education Teachers' Achievement. Competitive-Journal of Education 4(3) 289-304

Otermans, P. C. J., Kaur, H., Patrao, A., et al. (2025). Building confidence in university students from year 1 to year 2. *Cogent Education*, *12*(1). https://doi.org/10.1080/2331186X.2025.2462269

Ouwehand, K., Lespiau, F., Tricot, A., & Paas, F. (2025). Cognitive load theory: Emerging trends and innovations. *Education Sciences*, *15*(4), 458.

https://doi.org/10.3390/educsci15040458

Özdemir, A. Z., & Toker, Z. (2025). Analysis of distractors in mathematics questions and their potential to lead misconceptions. *Thinking Skills and Creativity*, *56*, 101730.

Palikara, O., Bonneville-Roussy, A., & Allen, K. A. (2025). Individual and contextual factors determining school belonging of adolescents in the UK. *School Mental Health*, *17*, 598–613. https://doi.org/10.1007/s12310-024-09725-y

Polack, C. W., & Miller, R. R. (2022). Testing improves performance as well as assesses learning: A review of the testing effect. *Journal of Experimental Psychology: Animal Learning and Cognition*, *48*(3), 222–241. https://doi.org/10.1037/xan0000323

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

Quansah, N., Quansah, F., & Dzakadzie, Y. (2025). Multiple-choice test development competencies of junior high school mathematics teachers in Ghana. *Frontiers in Education*, *10*, 1658971. https://doi.org/10.3389/feduc.2025.1658971

Rezigalla, A. A., Eleragi, A. M. E. S. A., Elhussein, A. B., et al. (2024). Item analysis: The impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education*, *24*, 445. https://doi.org/10.1186/s12909-024-05433-y

Rudolph, M. J., Daugherty, K. K., Ray, M. E., et al. (2019). Best practices related to examination item construction and post-hoc review. *American Journal of Pharmaceutical Education*, *83*(7), 7204. https://doi.org/10.5688/ajpe7204

Sanni, K.T, Aransi, W.O, Esan, O.M, & Ayodele, O.J (2024). Influence of Principal Leadership Styles, Teachers' Job Satisfaction, Work-Related Flow on Teachers' Productivity in Iwo Local Government Area of Osun State, Nigeria. Ilorin Journal of Education, 44(1), 251–262.

Shaibu L, Adeyemi O.V, I.E.JOY Obetta & Esan O.M (2024). An Assessment of learning environment as variables influencing students' academic performance in Ekiti State, Nigeria. IJEEI-International Journal of Educational Excellence and Innovation, 1(02), 37-52.

Shcheglova, I., Costley, J., Gorbunova, E., & Lange, C. (2025). Does activation of higher-order thinking skills lead to student dissatisfaction? *Innovations in Education and Teaching*

International, *62*(1), 31–44. https://doi.org/10.1080/14703297.2024.2332738

Skedsmo, G., & Huber, S. G. (2025). Exploring the complexities of assessment, evaluation, and accountability in education. *Educational Assessment, Evaluation and Accountability*, *37*, 1–4. https://doi.org/10.1007/s11092-025-09452-1

Stenberg, I., Sundgren, M., & Bostedt, G. (2025). The boy paradox: Motivational factors in gender differences and academic achievement. *Educational Psychology in Practice*, *41*(2), 211–230. https://doi.org/10.1080/02667363.2025.2469242

Suleiman, I. B., Okunade, O. A., Dada, E. G., et al. (2024). Key factors influencing students' academic performance. *Journal of Electrical Systems and Information Technology*, *11*, 41. https://doi.org/10.1186/s43067-024-00166-w

Tutz, G. (2025). Classical test theory. In *A short guide to item response theory models*. Springer. https://doi.org/10.1007/978-3-031-87271-6_7

Vera Gil, S. (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*. Advance online publication. https://doi.org/10.1007/s43076-024-00370-7

Vu, T., Magis-Weinberg, L., Jansen, B. R. J., et al. (2022). Motivation-achievement cycles in learning: A literature review and research agenda. *Educational Psychology Review*, *34*, 39–71. https://doi.org/10.1007/s10648-021-09616-7

Wong, R., Reichle, E. D., & Veldre, A. (2025). Prediction in reading: A review of predictability effects, their theoretical implications, and beyond. *Psychonomic Bulletin & Review*, *32*, 973–1006. https://doi.org/10.3758/s13423-024-02588-z

Young, G., Huijser, H., Mørcke, A. M., & Serbati, A. (2025). Measurement and transformation in academic development research. *International Journal for Academic Development*, *30*(4), 451–459. https://doi.org/10.1080/1360144X.2025.2577496

Zoucha, J., Himelfarb, I., & Tang, N.-E. (2025). Item classification by difficulty using functional principal component clustering and neural networks. *Educational and Psychological Measurement*, *85*(3), 429–457.