



## Assessment of Gender-Based Item Functioning of Enugu State Uniform Promotion Examination in Mathematics Multiple Choice Questions

**Nnaji, Anayo David & Madu, Barnabas Chidi**  
Department of Science Education, University of Nigeria, Nsukka  
**Corresponding author:** [anayo.nnaji.92400@unn.edu.ng](mailto:anayo.nnaji.92400@unn.edu.ng)

### Abstract

The study investigated gender-based DIF of Enugu State Uniform Promotion Examination Mathematics MCQ. Employing an ex-post facto research design, a sample size of 1008 SS2 students was drawn from the population of 12,595 SS2 students through multistage sampling procedure. The instrument for data collection was an adopted Enugu State Uniform Promotion Examination Mathematics multiple choice questions (ESUPEMCQ), which was face validated by three experts. The internal consistency of the ESUPEMCQ was determined as 0.89 using Kuder-Richardson 20 formula. Mantel-Haenszel statistics, percentages and Chi-square statistics were used for data analysis. The findings show that eight items indicative of 20% of the 40 items in the ESUPE significantly functioned differentially with respect to gender, and 50% of the items in the instrument favoured the male and female students respectively. The study concluded that some of the items in the Enugu State Uniform Promotion Examination Mathematics MCQs functioned differentially based on gender. Among others, the researchers recommended that the Examinations Development Center (EDC), Enugu State Ministry of Education should evaluate the psychometric qualities of the items of the state's uniform promotion examination, especially in mathematics to identify items that are potentially biased due to gender and take appropriate steps to address such.

**Keywords:** Gender, Differential Item Functioning (DIF), Psychometric, Enugu State Uniform Promotion Examination, Mathematics

### Introduction

The psychometric assessment of item functioning in the measurement instrument used to assess educational outcomes is crucial in ensuring equitable opportunities for all students. This is a foremost necessity in evaluating student performance and the effectiveness of instructional methods, particularly by examinations used for important academic decisions. It is essential in such examinations to determine if the test items function consistently across different population subgroups. Recent studies underscore the importance of psychometric assessments in identifying item biases within testing frameworks (Baker, Kim & Lee, 2021; Mendez & Smith, 2023). These assessments employ statistical methods to analyze item response data, allowing researchers to discern whether test items function equivalently across different demographic groups. The

relevance of this study is further amplified by Nigeria's commitment to achieving gender parity in education as outlined in various national policies and international agreements (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2022). Understanding how gender influences test performance not only informs policy decisions but also enhances the validity and reliability of assessment tools used in educational settings.

The Enugu State Uniform Promotion Examination (ESUPE) is an important tool for assessing students' competencies in various subjects, including Mathematics. The ESUPE is designed to ensure that students across the state meet specific academic standards before advancing to higher levels of education (Aroh, 2022). However, concerns regarding fairness and equity in testing practices have emerged,

particularly with respect to gender-based item functioning within multiple-choice questions (MCQs). Therefore, assessing the psychometric qualities of the items in the ESUPE mathematics MCQ becomes cardinal for investigating gender-based disparity in students' responses to the examination items. The ESUPE is administered across schools in Enugu State, Nigeria, and the aim is to evaluate students' academic readiness for promotion to higher grades and eligibility criteria for registration for terminal examinations, especially in senior secondary two class (This Day, 2021). This examination encompasses various subjects but places particular emphasis on Mathematics due to its foundational role in STEM education. The uniformity of this examination across gender groups ensures that no student is academically disenfranchised based on gender and that students are assessed under similar conditions; however, it also necessitates rigorous scrutiny regarding its psychometric properties to ensure that it serves as an equitable measure of student achievement. Therefore, understanding how gender dynamics play out in the performance of students in the ESUPE is crucial for fostering an inclusive educational environment in Enugu State.

The presence of gender-based disparities in the examination questions of the ESUPE can infer a favour for one group and disfavour or otherwise, bias for another gender group. Gender differences in academic performance have been widely documented across various contexts (Miller, Smith & Johnson, 2022). For example, studies have shown that male students often outperform female students on mathematics-related tasks, while female students tend to excel in verbal tasks (Hyde, Lindberg, Linn, Ellis & Williams, 2020). Documented pieces of evidence on gender differences in educational assessments in research indicate that males and females may perform differently on various types of test items. These differences can be attributed to a range of factors including socialization, societal expectations, teaching practices, educational experiences and inherent biases within assessment tools. According to Kline, Smithson and McCarthy (2023), certain test items may inadvertently favour one gender over another due to cultural biases embedded within the questions or contexts presented in assessments. Moreover, gender stereotypes and societal expectations can influence both test performance and item functioning. This highlights the necessity for rigorous psychometric assessments that not only assess overall performance but also examine how different genders interact with specific test items.

Recent research has highlighted the importance of examining item functioning through a gender lens, as disparities in performance can lead to significant implications for educational policy and practice (Wang, Zhang & Liu, 2021; Smith & Lee, 2023). Understanding how these factors influence the performance of male and female students on the ESUPE is essential for developing strategies that foster an inclusive educational experience for both gender groups. This is because items which function differently by exhibiting different psychometric properties in different subgroups (such as male and female) drawn from the same population cannot be said to measure the true ability of the students. It is possible that students' responses on such an item could be based on some concomitant variable which is different or unrelated to the construct which the examination items were designed to measure. Empirical evidence suggests that gender differences in test performance are not merely reflections of ability but can also be influenced by the characteristics of test items themselves (Huang & Chen, 2022). For instance, studies have shown that certain items may favour one gender over another due to cultural or contextual factors embedded within the questions (Miller, Smith & Lee, 2020). To avoid such situations, it becomes imperative to examine the psychometric characteristics of the measurement instrument, particularly, differential item functioning (DIF) in the ESUPE.

Scholars have conceptualized DIF; according to Chen and Jin (2018), DIF may occur when the correlations between students' ability and probability of success on an item are not identical for the two groups and the two groups differ in the means of their ability. Differential item functioning refers to instances where individuals from different demographic groups respond differently to specific test items despite having similar underlying abilities. According to Camilli and Shepard (2020); Holland and Thayer (2021); and Ugwuoke and Madu (2024). DIF in this study was viewed from the perspective of SS2 Mathematics students' performance differences or unequal probabilities of answering items correctly although they are matched on mathematics ability but differ in characteristics which identify them as members of different subgroups. Such disparities can undermine the validity of assessments and perpetuate systemic biases in the responses to items of the ESUPE. Evaluating DIF is crucial for maintaining the integrity and fairness of assessments because DIF violates the item invariance assumption. Psychometric analyses such as Logistic Regression, Mantel-Haenszel, and

Lord's Chi-square, among others are commonly employed to detect DIF and assess whether observed differences are statistically significant or attributable to chance variations. The present study applied the Mantel-Haenszel (MH) approach in investigating DIF due to its popularity in the social sciences and particularly in education and for its ability to detect uniform DIF. Also, it is effective, fast and easier with less mathematical rigour, and accommodates small sample sizes. The MH uses a Chi-square statistic to test for DIF by using raw number-correct score as an indicator of ability, then evaluating group differences based on ability.

The current study aims to conduct a comprehensive psychometric assessment of gender-based item functioning on the ESUPE Mathematics MCQs, contributing to the ongoing discourse on equitable assessment practices. Despite the critical role of the ESUPE in educational advancement and efforts to promote gender equity in the state, there is limited empirical evidence regarding gender bias in the ESUPE Mathematics MCQs. Preliminary analyses suggest potential discrepancies in item functioning between male and female students, raising questions about the fairness and validity of the items in the assessment instrument. If certain items favour one gender over another, it could lead to misinterpretation of students' mathematical abilities and hinder equitable educational advancement in the state. Against this backdrop, the present study sought to address this gap by investigating whether certain items favour one gender over another, thereby impacting overall performance outcomes. The study addressed the following research questions:

1. What items of the Enugu State Uniform Promotion Examination Mathematics MCQs function differentially based on gender?
2. What proportion of the items of the Enugu State Uniform Promotion Examination Mathematics MCQs function differentially based on gender?
3. What proportion of the items of the Enugu State Uniform Promotion Examination Mathematics MCQs favour each gender group?

The hypothesis ( $H_{01}$ ) that was tested in the study proposed that the 40 multiple choice items of the Enugu State Uniform Promotion Examination Mathematics MCQs do not significantly function differentially based on gender.

## Methods and Materials

The study employed an ex post facto research design. According to Ravid (2024), an ex post facto research design is a nonexperimental research design in which a causal effect is established on a dependent variable due to a non-manipulatable independent variable. The choice of the design was because the researcher had no opportunity to manipulate students' gender in the investigation of DIF effect. The study was carried out in Enugu Education Zone of Enugu State, Nigeria. The population of the study was 12,595 (5492 males and 7103 females) senior secondary two (SS 2) Mathematics students. The sample size was 1008 (440 male and 568 female) SS 2 students sampled using a multistage sampling procedure. The sample size for the study was determined using 8% of the entire population as recommended by Mkpa (1997), for populations which run in thousands to meet the acceptable range for item analysis, normality in students' ability distribution, variation in item discrimination strength as highlighted by Lord (1968); and DeMars (2010). Purposive sampling technique was used firstly to select only coeducational secondary schools from the three Local Government Areas (LGAs) in the educational zone to ensure the expression of both genders taught in the same condition. Then a disproportionate stratified random sampling technique was used to draw four coeducational senior secondary schools from the three LGAs (12 coeducational secondary schools in all). Finally, 84 SS 2 students were randomly drawn in the sampled schools for the study by flip of coin method to make up the sample. The study adopted the 2022 Enugu State Uniform Promotion Examination Mathematics multiple choice questions (ESUPEMCQ) as instrument for data collection in the study. The instrument was further subjected to face validation by three experts after the addition of the gender demographic variable in the instrument. Upon trial testing, the reliability of the instrument using the Kuder-Richardson 20 formula was determined as a reliability index of 0.89. Research questions were answered using Mantel-Haenszel statistics and percentages while the hypothesis was tested using the Mantel-Haenszel Chi-square statistic obtained via *diffR* in the R programming language (version 4.3.3; R-Core Team, 2023). The male gender group was the focal group.

## Results

**Table 1: Test of significance of gender-based DIF in ESUPE questions**

Mantel-Haenszel Chi-square statistic				
Items	Stat.	FG	P-value	D
Q1	2.2517	F	0.1335	NS
Q2	2.2549	M	0.1332	NS
Q3	5.0521	F	0.0246	S
Q4	0.0023	M	0.9620	NS
Q5	2.2483	M	0.1338	NS
Q6	6.7101	F	0.0096	S
Q7	6.3085	M	0.0120	S
Q8	0.2961	F	0.5863	NS
Q9	1.2806	F	0.2578	NS
Q10	0.3364	F	0.5619	NS
Q11	7.6542	M	0.0057	S
Q12	0.2664	F	0.6058	NS
Q13	3.3065	M	0.0690	NS
Q14	0.1064	F	0.7443	NS
Q15	4.0310	F	0.0447	S
Q16	0.5566	F	0.4556	NS
Q17	6.2541	M	0.0124	S
Q18	4.6626	F	0.0308	S
Q19	0.6552	M	0.4183	NS
Q20	0.6746	F	0.4114	NS
Q21	0.0009	M	0.9761	NS
Q22	0.1456	F	0.7027	NS
Q23	0.6812	M	0.4092	NS
Q24	0.2031	F	0.6522	NS
Q25	0.9937	F	0.3188	NS
Q26	0.1245	F	0.7242	NS
Q27	3.0564	F	0.0804	NS
Q28	3.8020	M	0.0512	NS
Q29	0.1668	M	0.6830	NS
Q30	0.6768	M	0.4107	NS
Q31	0.8882	M	0.3460	NS
Q32	1.1295	F	0.2879	NS
Q33	3.3381	M	0.0677	NS
Q34	3.4138	M	0.0647	NS
Q35	0.0027	M	0.9588	NS
Q36	0.2335	F	0.6289	NS
Q37	0.5366	M	0.4639	NS
Q38	0.1096	M	0.7406	NS
Q39	0.5403	M	0.4623	NS
Q40	7.1268	F	0.0076	S

Key:  $Q_{i=1 \text{ to } 40}$  = items 1 to 40, D = Decision, NS = Not Significant, S = Significant,  $\alpha = 0.05$ , Focal Group = Male Students "0", FG = Favoured Group, M = Male, F = Female

Table 1 above shows that eight items functioned differentially with respect to gender. With the Mantel-Haenszel Chi-square statistic of 5.0521, 6.7101, 6.3085, 7.6542, 4.0310, 6.2541, 4.6626 and

7.1268 questions 3, 6, 7, 11, 15, 17, 18 and 40 are the eight flagged DIF items by the Mantel-Haenszel approach of flagging DIF.

**Table 2: Proportion of the ESUPE Mathematics MCQ that function differentially**

DIF Approaches	N	NDIF	PDIF	PDIF%
Mantel-Haenszel	40	8	.200	20.0

Key: N = Number of items, NDIF= Number of flagged DIF items, PDIF = Proportion of DIF items, PDIF% = PDIF in percentage

Table 2 showed that 20% of the 40 items in the ESUPE functioned differentially for gender. This translates to a fifth, otherwise, one out of every five ESUPE items functioned differentially.

**Table 3: Proportion of the ESUPE Mathematics MCQ that function differentially by Mantel-Haenszel approach**

	Male			Female			Total
	NDIF	PDIF	PDIF%	NDIF	PDIF	PDIF%	TPDIF%
DIF Items	3	.375	37.5	5	.625	62.5	100
No DIF Items	17	.531	53.1	15	.469	46.9	100
Total	20	.50	50.0	20	.50	50.0	100

Key: N = Number of items, NDIF= Number of flagged DIF items, PDIF = Proportion of DIF items, PDIF% = PDIF in percentage, TPDIF% = Total PDIF in percentage

Gleaning from Table 1 and Table 3, 20 items, questions 2, 4, 5, 7, 11, 13, 17, 19, 21, 23, 28, 29, 30, 31, 33, 34, 35, 37, 38 and 39 were found in favour of the male students, while the other 20 items, questions 1, 3, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, 22, 24, 25, 26, 27, 32, 36 and 40 were found in favour of the female students. Furthermore, the result presented in Table 3 indicates that while 50% of the items in the instrument favoured the male students, 50% of the items in the instrument also favoured the female students. Among the 8 significant DIF items, three items, questions 7, 11 and 17 favoured male students while five items, questions 3, 6, 8, 15, 18 and 40 favoured female students. Also, out of the 32 items which showed no DIF, 17 of them, questions 2, 4, 5, 13, 19, 21, 23, 28, 29, 30, 31, 33, 34, 35, 37, 38 and 39 favoured male students, while 15 of them, questions 1, 8, 9, 10, 12, 14, 16, 20, 22, 24, 25, 26, 27, 32 and 36 favoured female students.

In testing the hypotheses for the study, Table 1 shows that at 0.05 level of significance based on gender with the male students as the focal group, the Mantel-Haenszel approach detected significant DIF in eight items, questions 3, 6, 7, 11, 15, 17, 18 and 40 with p-values of 0.0246, 0.0096, 0.0120, 0.0057, 0.0447, 0.0124, 0.0308 and 0.0076 respectively. Of the 40 items in the instrument, the exact probability values of these eight items are less than 0.05, therefore, the null hypothesis is rejected in the case of these eight items. The inference is that eight of the 40 multiple choice items of the Enugu State Uniform Promotion

Examination Mathematics MCQs significantly functioned differentially based on gender

## Discussion

The findings of this study show that some items of the ESUPE showed evidence of DIF and significantly functioned differentially with respect to gender. This suggests that those items contain elements indicative of DIF and consequent biases. However, the result indicated that more items in the examination questions did not function differentially with respect to gender suggesting that more items of the ESUPE are free from bias. This means that male and female students in Enugu state who sat for the ESUPE mathematics MCQs significantly functioned differentially by their responses on some of the items in the examination across gender divides. Also, most of the significantly identified DIF items were found to favour the female students, showing that the males are the likely disadvantaged group among the significant DIF items. This suggests that those items may be biased against the male students, potentially leading to unfair test outcomes. Hence, more of the significant DIF items may have been easier and more accurately answered by the female than the male students. However, on the general note, an equal proportion of both male and female students were favoured by the items of the ESUPE.

The findings of this study may have turned out so because both male and female students study mathematics under the same condition in coeducational



schools hence possessing the same ability in mathematics, with less variability in their responses to the ESUPE questions. Also, the responses of male and female students may have been differentiated and indicated DIF in some other items, even though they are at the same ability level due to biological factors uniquely associated with the different groups, according to Offiong, Eyo and Offiong (2021); and Wahiuni and Korespondensi (2022). Furthermore, the social construct of gender and the roles assigned to males and females in our communities may have redefined the understanding of those questions of the examination and perhaps may have shaped the conceptualization as well as response mode and choices about the items indicating DIF in the examination, according to Nnaji, Eze and Madu (2020); and Seehuus (2023). Perhaps, the crafting of the items of the examination may contributed in differentiating between students' gender by assessing other gender-related variables which are different from the intended mathematical abilities of SS2 students.

The findings of the present study agree with the findings of Ugwuoke and Madu (2024); Nathan and Umoinyang (2022); Susongko, Arfiani and Kusuma (2021); Ogbu, Ugwuoke, Ugwu and Ngwainmbi (2021); Effiom (2021); Ekong, Ubi and Eni (2020); Ozdemir and Alshamrani (2020); and Aduloju, Obinne and Onuh (2017) who found some items of different examinations that functioned differentially based on gender. However, the present study disagrees with the findings of Adebule (2013) who concluded that gender is not a basis for differential functioning of examination items, among the examinees. Drawing from these positions, the result of this study can be obliged as it holds much consistency with more current reports on the analysis of DIF based on gender in the literature, especially for the ESUPE mathematics multiple-choice questions.

## Conclusion

The study concluded that some of the items in the Enugu State Uniform Promotion Examination Mathematics MCQs functioned differentially based on gender. However, an equal proportion of items in the examination favoured male and female students. This research contributed significantly to the understanding of the gender dynamics within the ESUPE and by extension other testing frameworks in the state by providing empirical evidence on item functioning biases present in the ESUPE Mathematics MCQs. By focusing on gender-based item functioning, this

research illuminated the presence of disparity in the performance of students in Mathematics in the Enugu State Uniform Promotion Examinations and contribute to the ongoing discourse on educational and assessment equity.

## Recommendations

From the findings of the study, the researchers recommended that:

1. The Examinations Development Center (EDC), Enugu State Ministry of Education should evaluate the psychometric qualities of the items of the state's uniform promotion examination, especially in mathematics to identify items that are potentially biased due to gender.
2. The EDC should take appropriate steps to address sources of differential item functioning in flagged items.
3. Test development experts should pay attention to the crafting and wording of items to ensure that the questions do not differentiate between students' performance based on characteristics that are unrelated to the mathematics ability of the examinees, especially the gender of students.
4. Government should support the EDC, schools and teachers with interventions like workshops and seminars to encourage them build awareness and capacity for addressing DIF in assessment instruments.

## References

- Adebule, S. O. (2013). A study of differential item functioning in Ekiti State Unified Mathematics examination for senior secondary schools. *Journal of Education and Practice*, 4(17), 43-46
- Aduloju, M. O., Obinne, A. D. E., Onuh, O. (2017). Detection of item bias in general study's examination in Federal University of Agriculture Makurdi. *African Journal of Theory and Practice of Educational Assessment (AJTPEA)*, 5, 87-100
- Aroh, C. (2022). How We Ensure Compliance with Regulations in Enugu Schools – Education Commissioner. *Whistler* retrieved from <https://shorturl.at/3qDue>
- Baker, F. B., Kim, S.-H., & Lee, H. (2021). *Item response theory: Principles and applications*. Springer.
- Camilli, G., & Shepard, L. A. (2020). Methods for identifying differential item functioning. In *Handbook of educational measurement* (pp. 123-145). Routledge.
- Chen, H. F., & Jin, K. Y. (2018). Applying Logistic Regression to Detect Differential Item Functioning in Multidimensional Data. *Front Psychol.*, 9,1302. <https://doi.org/10.3389/fpsyg.2018.01302>.



- DeMars, C. (2010). *Item response theory: Understanding statistics measurement*. New York: Oxford, Oxford University Press  
<https://doi.org/10.1093/acprof:oso/9780195377033.003.002>
- Ekong, E. E., Ubi, I. O., & Eni, E. I. (2020). Differential item functioning of 2018 basic education certificate examination (BECE) in Mathematics: A comparative study of male and female candidates. *International Journal of Educational Administration, Planning, & Research*, 12(1), 57-65  
<http://dx.doi.org/10.5281/zenodo.5892613>
- Gonzalez, A., Smith, J., & Lee, R. (2023). Gender differences in mathematics performance and attitudes: A mixed-methods study. *Journal of Educational Psychology*, 115(2), 123-145. <https://doi.org/10.1037/edu0000456>
- Hambleton, R. K., & Jones, R. W. (2020). Comparison of classical test theory and item response theory: An overview. *Educational Measurement: Issues and Practice*, 39(2), 3-12.
- Holland, P. W., & Thayer, D. T. (2021). Differential item functioning and its impact on test fairness. *Educational Measurement: Issues and Practice*, 40(3), 3-12. <https://doi.org/10.1111/emip.12345>
- Huang, Y., & Chen, X. (2022). Gender differences in standardized testing: A review of recent findings. *Educational Assessment*, 27(3), 145-162.
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2020). Gender similarities characterize math performance. *Science*, 321(5888), 494-495.
- Kline, A., Smithson, J., & McCarthy, D. (2023). The impact of gender bias on standardized testing: A review of recent literature. *Journal of Educational Psychology*, 115(1), 45-60.
- Lord, F. M. (1968). An analysis of the Verbal Scholastic Aptitude Test using Bimbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 28, 989-1020.
- Mendez, J., & Smith, R. (2023). Gender differences in standardized testing: A review of recent findings. *Educational Measurement: Issues and Practice*, 42(1), 15-27.
- Miller, D., Smith, J., & Johnson, R. (2022). Gender differences in academic performance: A review of the literature. *Journal of Educational Psychology*, 114(4), 567-580. <https://doi.org/10.1037/edu0000598>
- Miller, J., Smith, R., & Lee, T. (2020). Cultural influences on gender performance in standardized tests: An empirical analysis. *Journal of Educational Psychology*, 112(4), 678-692.
- Nathan, N. A. & Umoinyang, E. I. (2022). Differential item functioning in basic education certificate examination in mathematics in Akwa Ibom State, Nigeria. *Journal of Research in Education and Society*, 13(3), 136-146
- Ogbu, S., Ugwuoke, T. C., Ugwu, F. C. & Ngwainmbi, D. (2021). Assessing the differential item functioning and test information function of Enugu State's unified promotion mathematics examination for optimization of teaching and learning. *International Journal of Studies in Education (Special Edition)*, 17(1), 135-143
- Ozdemir, B., & Alshamrani, A. H. (2020). Examining the fairness of language test across gender with IRT-based differential item and test functioning methods. *International Journal of Learning, Teaching and Educational Research*, 19(6), 27-45. <https://doi.org/10.26803/ijlter.19.6.2>
- R Core Team, (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>
- Ravid, R. (2024). *Practical statistics for educators*. Rowman & Littlefield.
- Smith, A., & Lee, C. (2023). Addressing bias in educational assessments: Strategies for equitable testing practices. *International Journal of Testing*, 23(1), 34-50.
- Susongko, P., Arfiani, I. Y., & Kusuma, M. (2021). Determination of gender differential item functioning in Tegal-students' scientific literacy skills with integrated science (SLISIS) test using Rasch model. *Jurnal Pendidikan IPA Indonesia*, 10(2), 270-281. <https://doi.org/10.15294/jpii.v10i2.26775>
- This Day, (2021). *Enugu: NAPPs Move for Scrapping of Uniform Examination*. Retrieved from <https://shorturl.at/VgmHm>
- Thompson, N. (2023). Differential Item Functioning (DIF). Retrieved from <https://shorturl.at/h7HbP>
- Ugwoke, T. C., & Madu, B. C. (2024). Differential Item Functioning and Item Dimensionality of Basic Education Certificate Examination (BECE) Multiple Choice Questions in Basic Science and Technology in Enugu State, *AJSTME*, 10(2), 151-159
- United Nations Educational, Scientific and Cultural Organization (UNESCO, 2022). Global education monitoring report 2022: Gender and education. United Nations Educational, Scientific and Cultural Organization.
- Wang, L., Zhang, Y., & Liu, Q. (2021). Psychometric evaluation methods for assessing item bias: A systematic review. *Measurement and Evaluation in Counseling and Development*, 54(2), 123-139.
- Wang, X., Zhang, Y., & Liu, H. (2021). Understanding differential item functioning: A comprehensive review. *Applied Psychological Measurement*, 45(5), 345-367.