

**Article**

---



<https://doi.org/10.52696/MGBL4977>

Reprints and permission:  
The Malaysian English Language Teaching Association  
Corresponding Author:  
Clarissa Pesidas [cppesidas02137@usep.edu.ph](mailto:cppesidas02137@usep.edu.ph)  
ORCID ID: <https://orcid.org/0009-0008-4626-0509>



## **Inclusivity in Multiple-Choice Questions Assessment (MCQA): An Analysis of Selected Design and Proposed Modifications for Gender and Linguistic Inclusivity**

Clarissa Pesidas  
College of Education Department,  
University of Southeastern Philippines, Philippines

### **ABSTRACT**

Gender and linguistic inclusivity are pivotal in language assessment practice, particularly in this study, in the design of multiple-choice questions. The purpose of this study is to analyze the inclusivity of the selected MCQA design, the underlying factors that influence it, and the inclusive modifications that can be implemented to better accommodate learners from diverse gender and linguistic backgrounds. This study is a Qualitative Descriptive Design, which provided a comprehensive report about the research questions. A total of at least two (2) teachers and three (3) students were able to participate in the In-Depth Interviews (IDIs), while one (1) teacher and two (2) students for the Focus Group Discussions (FGDs). A total of eight (8) participants were purposefully selected in this study. In the data collection, Fairclough's Critical Discourse Analysis served as the theoretical framework for this study, guiding both the researcher and participants in the data analysis. The study revealed that using Fairclough's textual dimension, three dominant patterns of inclusivity emerged in the MCQA items: (1) gender representation, (2) lexical difficulty, and (3) cultural-linguistic orientation. In addition, it revealed how the mismatch between test designers' assumptions and the actual test takers may affect students' performance, which Fairclough calls discursive misalignment. Lastly, in the social analysis, factors shaping MCQA design were context and language use, curriculum, teacher training, and institutional demands, and resources used in constructing MCQA.

**KEYWORDS:** Multiple-Choice Questions Assessment (MCQA), Gender, Linguistic, Inclusivity

## Introduction

Multiple-Choice Questions Assessment (MCQA) has been widely used across the globe for various purposes, benefits, and its overall significance, examples include the Program for International Student Assessment (PISA), Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS), and the National Achievement Test (NAT). MCQA is valued for its efficiency, objectivity, and reliability (Haladyna, 2022; Weimer, 2018). Notably, some MCQAs have been shown to be fair across genders, with well-designed items avoiding gender bias (Bond, 2013). Furthermore, research has demonstrated that linguistic inclusivity can positively influence students' academic performance (Kaur, Noman, & Nordin, 2017; Lampe & Tsaouse, 2010; Lyon et al., 2022).

However, while this type of assessment contributes significantly to measuring students' learning, it also presents certain drawbacks that may compromise the authenticity of assessing true learning outcomes. In particular, some studies have shown how the format and structure of MCQAs interact with gender, revealing that males outperformed females in all multiple-choice tests (Masoumi & Sadeghi, 2020; Shear, 2023). However, this finding is not consistent across all research. Another study suggests that claims about structure and content reflecting a universal male bias may be overstated, as no evidence of such bias was found (du Plessis & du Plessis, 2009). Additionally, non-native English speakers in the classroom were shown to be disadvantaged when MCQAs included complex language and unfamiliar vocabulary (Lampe & Tsaouse, 2010).

Despite the abovementioned concerns, existing guidelines for MCQA design often overlook the role of inclusivity in addressing such disparities. There have been various references that offer practical tips for creating quality MCQAs (Bothell, 2001; Butler, 2018; LSA Learning & Teaching Technology Consultants, 2023; University of Manitoba Editors), but they give little to no consideration to promoting gender and linguistic inclusivity. With this in mind, this research will further explore practical interventions for improving gender and linguistic inclusivity in Multiple-Choice Questions Assessment (MCQA), an area that is less frequently studied.

## Research Questions

This study aims to address the following research questions:

1. How inclusive is the selected MCQA design?
2. What are the underlying factors that influence the selected MCQA design?
3. What inclusive modifications can be implemented to the MCQA design to better accommodate learners from diverse gender and linguistic backgrounds?

## Literature Review

### *Assessment Validity*

In the assessment field, validity is pivotal because it determines whether a test measures what it is intended to measure, thereby guiding the meaningful interpretation and use of scores (Hoffman,

2025). Validity is considered a foundational element of testing systems, including educational assessments that aim to evaluate learners' knowledge, skills, and competencies (Giraldo, 2020). Validity is not simply a matter of statistical accuracy; it involves ensuring that test items appropriately represent the constructs they claim to measure without introducing irrelevant barriers, such as overly complex language or culturally distant content. For example, a recent study of the Academic Discussion Task (ADT) demonstrated that it functions as a valid measure of Chinese test-takers' academic discourse ability and did so without showing evidence of gender bias, highlighting that well-designed language tasks can simultaneously meet validity and fairness criteria (Zhang, 2025). This finding emphasizes that language assessments are only useful if they provide valid inferences about learners' abilities and if they treat all demographic groups equitably.

However, language testing has also drawn attention because it can introduce various fairness issues throughout test construction, administration, scoring, and interpretation stages (D'Este, 2012). Fairness concerns arise when test items include linguistic, cultural, or contextual features that advantage some test-takers over others, thereby introducing construct-irrelevant variance that obscures what the assessment actually measures. When such unfair influences are present, the validity of score interpretations is compromised, because differences in performance may reflect extraneous factors such as language familiarity or gendered assumptions rather than true differences in the target construct. Contemporary frameworks in language testing underscore that fairness should be integral to validity evaluations—that is, an assessment cannot be fully valid if it systematically disadvantages particular groups based on item design choices (AERA, APA, & NCME, 2014; Rezai et al., 2022).

### *Fairness*

The growing emphasis on inclusive education has prompted researchers to examine the inclusivity of assessments, ensuring that they provide equitable opportunities for all learners (Tai et al., 2021; Tai et al., 2022). Inclusive assessments aim to eliminate disadvantage based on learners' background, culture, language, or ability, and promote fairness in test design and interpretation (Hanesworth, 2019; University of Cambridge Editors, 2025). Historically, inclusivity focused on accommodating students with disabilities, but contemporary scholarship emphasizes that inclusive assessment is a pedagogical best practice for all learners, aiming to reduce bias and ensure that tests measure intended constructs accurately (Centre for Teaching and Learning Editors, n.d.). Teachers play a critical role in this process, from aligning assessment tasks with learning objectives to analyzing evaluation evidence to ensure equity and validity in learning outcomes (Phelan & Phelan, 2010).

Multiple-choice questions (MCQs) are one of the most commonly used assessment formats, providing an efficient way to evaluate knowledge and understanding (Gao, 2022; Lesage et al., 2013). MCQs are favored for their rapid scoring and practical administration, particularly in standardized testing or large cohorts (Oc & Hassen, 2024). However, designing high-quality MCQs involves challenges, including creating plausible distractors and ensuring the items reflect diverse linguistic and cognitive abilities (Kurnia, 2019; Morlet & Shah, 2025). If distractors are too obvious, biased, or culturally irrelevant, they may compromise fairness and validity, as

learners' performance may reflect familiarity with language or context rather than actual knowledge (Gao, 2022; Lesage et al., 2013).

In the context of inclusivity, MCQs must also consider gender and linguistic diversity. Studies indicate that items using gendered nouns or culturally unfamiliar examples can unintentionally disadvantage students, limiting equitable demonstration of competence (Tai et al., 2022; Hanesworth, 2019). Therefore, inclusive MCQs require attention to language simplicity, culturally neutral or localized contexts, and balanced representation of gender and roles (University of Cambridge Editors, 2025; Phelan & Phelan, 2010). By systematically addressing these factors, assessment developers can enhance the validity, reliability, and fairness of multiple-choice tests, ensuring that they accurately reflect students' knowledge and skills without embedding unintended biases (Oc & Hassen, 2024; Kurnia, 2019).

### *Bias*

Several studies have explored challenges related to gender inclusivity in classroom settings, highlighting barriers such as implicit bias, stereotypical representations, and unequal participation (Culajara & Culajara, 2023; Pandey, 2025). However, research specifically addressing gender inclusivity in the context of the language employed in multiple-choice assessments (MCAs) remains limited. Language in assessments plays a critical role in shaping learners' perceptions and engagement, and biased or non-inclusive language can inadvertently marginalize certain groups, affecting both comprehension and performance (Eubanks, 2021; Lee & Smith, 2022).

To conclude, this literature review seeks to synthesize the current understanding of inclusive assessment practices in education, with particular emphasis on multiple-choice formats. The existing body of research suggests that while general inclusivity in classrooms has been examined extensively, the intersection of gender and linguistic inclusivity within multiple-choice assessments is an underexplored area. This gap highlights the need for further studies that critically investigate how language choices in assessment items can either reinforce or challenge gender biases, thereby promoting equitable educational outcomes.

### **Theoretical Framework**

This study will be guided by the Critical Discourse Analysis (CDA) as the theoretical framework. This framework was established by scholars in the 1990s, including Theo van Leeuwen, Gunther Kress, Teun van Dijk, and Norman Fairclough (Amoussou & Allagbe, 2018). In this study, Fairclough's most recognized and practical three-dimensional model for analysis will be utilized: *text analysis*, *discursive practice analysis*, and *social practice analysis* (Isnaeni et al., 2025). This model connects the detailed examination of the text itself with the broader analysis of how discourse operates on discursive and societal levels (Alasiri, 2024).

Text analysis focuses on the surface level of a text, which can be either oral or written language, including but not limited to the analysis of structure, semantics, vocabulary, cohesion, etc. (Lingwist, 2020). Discourse analysis, on the other hand, emphasizes the importance of analyzing how the text was produced, distributed, and consumed. Lastly, in the same study, social practice analysis, according to Fairclough, "*means the social and cultural goings-on which the*

*communicative event is a part of*". With these dimensions, CDA examines how language used in real-life situations helps create and reflect social structures, power relations, identities, and ideas by connecting small details in a text to larger social contexts (Gölbaşı, 2017).

Having explored the theory, CDA will serve as the foundation and direction for this study. With the selected MCQA designs, these three dimensions help reveal how the test items may unintentionally reinforce inequalities. Aside from the fact that the test items will be analyzed more thoroughly than they were constructed (linguistic features), CDA will guide the analysis of the process behind producing, distributing, consuming, and the social, educational, and cultural contexts in which these items are used.

## **Methodology**

### *Research Design*

This study employed a Qualitative Descriptive Design, which provides a detailed description of a phenomenon and is flexible and exploratory in approach (Ayton, 2023). The design is best implemented in this study because it analyzed and described the existence of the selected sample of MCQAs. The participants had been given enough time to analyze the test and share their invaluable insights through In-Depth interviews (IDIs) and Focus Group Discussions (FGDs). This design is deemed appropriate for this study as it enabled the researcher to capture the participants' analysis using their own words. Simply put, the study focused more on providing a clear and straightforward description of the phenomenon being studied.

### *Research Participants and Locale*

This study will involve at least two (2) teachers and three (3) students for In-Depth interviews (IDIs), and one (1) teacher and two (2) students for Focus Group Discussions (FGDs), with a total number of 10 participants. This study employed purposive sampling, a commonly used technique that is more intentional in selecting participants to gain deeper insights into the matter being studied (Tajik et.al., 2024). Given its selectivity, this study considered a diverse range of gender identities and linguistic backgrounds.

The study was conducted in Mindanao, Philippines, using a blended modality that combined face-to-face and online components, as well as a pure face-to-face approach. While the teachers came from three different provinces: Davao del Sur, Davao del Norte, and Davao Oriental, the students were all from Davao del Sur. All face-to-face sessions were held in a higher institution in Davao del Sur. In the FGD, all students were present face-to-face, and a teacher joined online. During the IDIs, all students were interviewed face-to-face, while the teachers were interviewed online.

### *Research Instruments*

This study examined the selected MCQA design, with a focus on gender and linguistic inclusivity. It also aimed to identify potential inclusive modifications that can be considered in the MCQA design to better support learners from diverse backgrounds. To achieve these goals, the researcher

developed a semi-structured interview that guided the data-gathering process. The flexibility of the tool enabled the researcher to probe, clarify, and follow up on the participants' responses.

### *Data Collection*

This study followed the six-phase thematic analysis framework of Braun and Clarke: (1) *familiarization with data*, (2) *generating initial codes*, (3) *searching for themes*, (4) *reviewing themes*, (5) *defining and naming themes*, and (6) *writing the report* (Ahmed et. al, 2025).

*First*, the researcher collected and selected the MCQA design that was examined in this study. *Second*, a document analysis of the selected MCQA samples from the English language discipline was conducted using Critical Discourse Analysis theory, as proposed by Fairclough (1995), with the active participation of the participants through In-Depth Interviews (IDIs) and Focus Group Discussions (FGDs). *Third*, during the IDIs and FGDs, the linguistic features identified served as initial codes. As mentioned above, the participants included teachers and students with diverse gender identities and linguistic backgrounds. The IDI and FGDs aimed to analyze linguistic features and gather insights into how the selected MCQA samples were perceived and interpreted in terms of gender and linguistic inclusivity. Additionally, in this phase, it sought to identify patterns or themes related to participants' experiences, understand why biases appear in test items, and collect suggestions for designing more equitable items and questions. *Fourth*, the themes were reviewed by the researcher. In this phase, it aimed to provide coherent, well-supported themes that accurately reflect the dataset. *Fifth*, after careful review, each theme was defined with conceptual clarity. *Sixth*, the researcher presented the theme with evidence and interpretations.

Considering all these processes, the study lasted for two months, taking into account the time required for sending invitations, securing approvals, and accommodating participants' availability.

### *Data Analysis*

In the qualitative analysis of results, this study will follow the three dimensions of Critical Discourse Analysis theory by Fairclough (1995), which have three dimensions: (1) *text*, (2) *discourse practice*, (3) *sociocultural practice* (Lingwist, 2020).

#### *a. Textual Analysis*

The textual analysis focuses on analyzing the language use in the discourse (Alasiri, 2024). In this part of the process, it investigates the linguistic features used by the test makers, such as the use of pronouns (*he/she*, *him/her*, *they/them*), gendered nouns (*policeman vs. police officer*), and stereotypical roles (*cooking - typically associated with women*). The identified linguistic features relevant to the study will be elaborated upon depending on the themes that emerge from it.

#### *b. Discursive Analysis*

This dimension explains how the selected MCQA design was created, distributed, and interpreted. In this phase, the process behind the test items will be analyzed, including the author of the assessment, its accessibility and reach, and intended test takers.

### *c. Social Analysis*

This phase examines the social, educational, and cultural contexts, with a specific focus on societal norms related to gender, language expectations in schools, and cultural values. It aims to provide insight into the context that influences the text and hopes to uncover underlying factors that explain why biases appear in the assessments.

#### *Data Collection Procedures*

This study employed a Qualitative Descriptive Design, which utilized a combined two data collection methods (data triangulation): in-depth interviews (IDI) and focus group discussions (FGD). To begin, the researcher browsed for sample MCQA designs and selected those that were accessible online, considering that they were within the English language discipline. Among all the MCQA designs, only three were chosen, and only 10 test items per MCQA were selected to include in a 30-item collated sample. This collated sample served as the basis of the analysis. Second, the researcher sought research participants who met the criteria, specifically teachers and students with diverse gender identities and linguistic backgrounds. Third, the researcher emailed the participants, asking for their participation. Fourth, when they agreed to participate, the In-Depth Interviews (IDIs) and Focus Group Discussions (FGDs) were conducted. Fifth, the researcher wrote comprehensive results and a discussion, addressing the research questions posed by the study.

## **Findings**

### *Textual Analysis: Inclusivity in MCQA Design*

Using Fairclough's textual dimension, three dominant patterns of inclusivity emerged in the MCQA items: (1) gender representation, (2) lexical difficulty, and (3) cultural-linguistic orientation.

#### 1. Gender Representation: Male-Centered Linguistic Patterning

Across all three test sources, participants consistently identified a systematic overuse of masculine references, including male names (Erik, Johnny, Alfredo, Mr. Culminas), masculine pronouns (he, his), and male-associated social roles (e.g., sports, politics). This pattern was not random; rather, it reflects what Fairclough calls ideological patterning in texts, where certain social groups are made more visible than others.

Although some items were technically gender-neutral, this neutrality was often defaulted rather than inclusive. As **Student 4 (S4)** observed, items avoided explicit gender bias but also failed to actively represent diverse identities, resulting in what can be called passive exclusion. **Teacher 2 (T2) and Teacher 3 (T3)** further demonstrated that masculine forms clearly outnumbered feminine ones, making males the unmarked norm and linguistically marginalizing females and nonbinary identities.

From a CDA perspective, this is significant because repetition of male-centered forms constructs masculinity as the standard participant in academic and social life, while other genders become secondary or invisible.

## 2. Lexical Accessibility: Linguistic Complexity as a Gatekeeping Device

Participants repeatedly identified vocabulary such as amalgamation, cohesive, and complex grammatical structures as barriers. These words do not merely test grammar; they also require high lexical and cultural capital, which disadvantages multilingual and less English-dominant learners. This pattern reflects what Fairclough describes as “power encoded in language”: those who control academic English are privileged, while others are positioned as deficient. Even when the grammar skill being tested is simple, difficult vocabulary shifts the task from testing grammar to testing social and linguistic capital.

**T3’s** observation that clues like “than” made answers obvious also reveals technical weaknesses in item construction, which undermine validity and further disadvantage students who rely on genuine grammatical reasoning rather than test-wiseness.

## 3. Cultural Orientation: Western-Centred Discursive Framing

The frequent use of foreign references (e.g., Calais, Siberia, Royal Albert Hall, Cambridge University, Soviet Union) reflects a Western-centric discursive frame. While not inherently wrong, this creates cultural distance between Filipino learners and the test content.

As **S4** and **T3** explained, unfamiliar contexts shift cognitive attention away from grammar and toward decoding unfamiliar places and references. In CDA terms, the tests reproduce a global English ideology in which Western knowledge is treated as more legitimate than local experience.

Thus, linguistic difficulty is not only lexical but cultural, privileging students who are already exposed to Western academic discourse.

### *Discursive Practice Analysis: Assessment Developers and Intended Test Takers*

This study assumed that the developers of all the tests are teachers, as they are commonly considered to be test designers and have control over both content and pedagogy. Additionally, all test takers are considered students in this study, as they are the primary target population of most tests assessing one's own ability or learning.

The MCQAs come from institutions that operate within elite, global, and standardized testing cultures (Tracktest, NAT, ACET). These institutions prioritize standard English, CEFR norms, and academic formality, which shape the construction of language in the tests.

The findings reveal a mismatch between test designers’ assumptions (that students are fluent, globally oriented, and culturally aligned with Western English) and the actual test takers (multilingual Filipino students with varied exposure to English). This mismatch produces what Fairclough calls discursive misalignment, where texts fail to accommodate their real audience.

*Social Analysis: Factors Shaping MCQA Design*

From a Critical Discourse Analysis perspective, the design of the selected MCQAs is shaped by the interaction between institutional expectations, teacher practices, and the linguistic and cultural realities of students. Participants consistently pointed to **context and language use** as major influences on how test items are constructed. **Student 1 (S1)** noted that the meaning of questions is strongly affected by the contexts in which they are embedded, while **S4** explained that unfamiliar examples and traditional English structures can cause students to misinterpret items even when they know the correct grammar. This concern was reinforced by **Student 5 (S5)**, who pointed out that unclear guidelines and complex wording can make questions unnecessarily difficult. **T2** added that students' multilingual backgrounds complicate how they process pronouns and grammatical forms, as learners may focus more on whether a pronoun is grammatically acceptable than on the grammatical feature being tested. **T3** further observed that consistency in using names—particularly male names—often guides item construction, even though this practice unintentionally reinforces gender bias. Together, these accounts show that linguistic and contextual choices in MCQAs are not neutral but reflect underlying assumptions about who the learners are and how they are expected to process language.

The **curriculum, teacher training, and institutional demands** also strongly influence MCQA design. **Student 2 (S2)** attributed many of the patterns in the test items to standard curriculum practices, while **Student 3 (S3)** emphasized that teachers face difficulty and pressure when constructing questions because assessment writing requires time, technical skill, and multiple considerations. **S4** expanded this by explaining that many teachers rely on old worksheets or textbook-based questions due to limited time, which leads to the repeated use of outdated and non-inclusive materials. From the teachers' perspective, **Teacher 1 (T1)** confirmed that the items were aligned with the Most Essential Learning Competencies (MELCs), particularly grammar skills such as pronouns and parts of speech, but also acknowledged that a lack of training in inclusive language affects item quality. **T2** likewise pointed out that curriculum and training strongly shape what kinds of examples and language appear in test items, while **T3** added that institutional contexts, such as the identity of Ateneo as a Jesuit school, influence how examples are chosen, sometimes privileging institutional identity over inclusivity. These narratives show how assessment practices are constrained by curriculum, time, and institutional expectations rather than guided purely by learner diversity.

Finally, the **resources used in constructing MCQAs** further explain the persistence of bias and exclusion. **S3** observed that many test items closely resemble those found in books and online sources, which explains why similar structures and examples appear repeatedly. **S4** supported this by noting that many of the test questions mirror what is found in traditional English textbooks. From a CDA perspective, this reliance on inherited materials means that older discourses—such as male-dominated representations, formal Western English, and culturally distant examples—are continuously reproduced in new tests. As a result, MCQAs reflect not only individual teacher choices but also the broader educational system's dependence on established materials and dominant language norms, which ultimately shape who feels included and who feels marginalized in assessment contexts.

## Discussion

This study examined how gender and linguistic inclusivity are constructed in selected Multiple-Choice Question and Answer (MCQA) designs using Fairclough's Critical Discourse Analysis. The findings show that MCQAs are not neutral instruments of assessment but are shaped by linguistic choices and sociocultural assumptions that systematically privilege some learners while marginalizing others. Three major thematic patterns emerged: (1) gendered representation, (2) linguistic accessibility, and (3) cultural framing, all of which directly address the study's research questions on how assessment language reproduces bias.

### *Gendered Representation in MCQAs*

The first major theme is the dominance of male-centered language in MCQA items. The frequent use of male names (e.g., Erik, Johnny, Alfredo), masculine titles (Mr., dad), and masculine pronouns (he, his) constructs masculinity as the default identity within academic and social contexts. Although some items were technically gender-neutral, participants noted that this neutrality was often superficial and did not actively include diverse gender identities. This supports earlier studies showing that assessment materials tend to reproduce androcentric norms, where men are represented as actors, leaders, and decision-makers while women and nonbinary identities remain invisible (Sunderland, 2006; Sunderland & McGlashan, 2012).

From a CDA perspective, these repeated linguistic patterns are ideological. Fairclough (1995) argues that repeated lexical and grammatical choices naturalize certain social relations. In this case, repeated masculine references normalize male dominance in political, academic, and public roles. As T2 and T3 noted, even when bias is unintentional, it still shapes how learners imagine who belongs in positions of authority. This finding answers the first research question by showing that MCQAs linguistically reproduce gender hierarchies through naming, pronoun use, and role assignment.

### *Linguistic Accessibility and English Dominance*

The second major theme concerns lexical difficulty and linguistic accessibility. Participants repeatedly identified words such as amalgamation and cohesive as barriers to understanding, especially for multilingual learners. Instead of measuring grammatical competence, such vocabulary shifts the task toward measuring English proficiency and academic exposure. This reflects what Shohamy (2001) and McNamara and Roever (2006) describe as the hidden power of tests: assessments often evaluate more than what they claim to measure.

Through Fairclough's lens, this represents a form of linguistic gatekeeping. Those with greater exposure to academic English are positioned as more capable, while multilingual learners are implicitly constructed as less competent. T3's observation that some items contain grammatical clues that make answers obvious further reveals that MCQAs do not merely assess knowledge but also test familiarity with test-taking conventions. This finding answers the second research question by showing that linguistic design features privilege certain language users over others.

### *Cultural Framing and Western Bias*

A third major theme is the Western-centered cultural framing of the test items. References to Cambridge University, Siberia, Calais, and Royal Albert Hall assume that learners share a globalized, Western-oriented knowledge base. As **S4** and **T3** explained, such contexts can distract students from the grammatical task because they must first interpret unfamiliar cultural references. This aligns with previous research showing that culturally distant test content lowers comprehension and increases cognitive load (Abedi, 2006; Solano-Flores, 2008).

In CDA terms, this reflects the dominance of global English ideology, where Western knowledge is treated as the standard. Fairclough (2001) notes that institutional texts often reproduce dominant global discourses, and MCQAs are no exception. The findings show that even national tests adopt international and Western references, thereby marginalizing local Filipino knowledge and experiences.

### *Institutional Forces Behind These Patterns*

The study also revealed that these biases are produced not by individual prejudice but by institutional and structural factors. Teachers and students alike identified curriculum standards, lack of training, time constraints, and dependence on existing materials as key influences. This reflects Bourdieu's (1991) concept of symbolic power, where institutions reproduce dominant norms through routine practices. When teachers reuse textbooks and online materials, they also reproduce the gendered, Western-centered, and English-dominant discourses embedded in them.

### *Modifications for Inclusive Assessment Design*

The inclusive modifications proposed by participants directly respond to these findings. Suggestions such as using gender-neutral terms, balanced naming practices, localized contexts, and simpler vocabulary align with research on fair and inclusive assessment (Kunnan, 2010; Shohamy, 2001). These strategies would not only reduce bias but also improve test validity by ensuring that MCQAs measure what they intend to measure—language skills rather than cultural familiarity or gendered expectations.

## **Implications and Recommendations**

This study showed that MCQA items systematically favored male identities, high-level academic English, and Western cultural references, which disadvantaged multilingual, culturally diverse, and gender-diverse learners. These findings have direct implications for teachers, test developers, institutions, and policymakers.

For teachers and item writers, the dominance of masculine names and pronouns in the test items indicates a need to critically review how people and roles are represented. Teachers should use gender-neutral or balanced naming practices, avoid stereotypical roles, and ensure that pronouns do not unintentionally privilege one gender over others. This recommendation is directly based on

the finding that male characters and masculine terms appeared far more frequently than female or neutral forms, reinforcing gendered assumptions in assessment content.

In terms of language use, the frequent presence of difficult vocabulary and ambiguous instructions shows that test items were not designed with multilingual learners in mind. Teachers and test developers should therefore prioritize simple, clear, and functional English, especially when the goal is to assess grammar or comprehension rather than vocabulary knowledge. This aligns with participants' reports that unfamiliar or complex words distracted them from the grammatical task being tested.

For test developers and publishers, the strong reliance on foreign contexts and Western references suggests a need for localized and culturally relevant content. Using familiar places, names, and situations can reduce unnecessary cognitive load and allow students to focus on the skill being assessed. This recommendation follows directly from the finding that foreign contexts confused students and reduced their comprehension.

For institutions, the results highlight the importance of professional development in inclusive assessment design. Teachers reported limited training and heavy reliance on existing materials, which explains why outdated and biased items continue to circulate. Institutions should therefore provide training on gender-inclusive language, culturally responsive testing, and multilingual-friendly item construction.

At the level of policy, educational authorities should incorporate gender and linguistic inclusivity into official guidelines for test construction. Since national and institutional assessments shape classroom practice, embedding inclusivity in policy can help prevent the reproduction of bias in large-scale testing.

### **Implications for Research**

Since this study employed a qualitative descriptive design, its primary contribution lies in providing a detailed, context-based understanding of how gender and linguistic inclusivity are constructed in MCQA items. Future researchers may build on these findings by conducting further qualitative studies that examine other types of assessments, subject areas, or educational levels to determine whether similar patterns of bias appear across different testing contexts. Comparative studies across schools, regions, or testing systems may also help reveal how institutional and cultural settings influence assessment design.

In addition, future research may explore the perspectives of test developers and policymakers to better understand how assessment guidelines, training, and institutional expectations shape the production of test items. Classroom-based studies could also investigate how students actually interpret and respond to gendered language, unfamiliar vocabulary, and culturally distant contexts during test-taking. By expanding the scope of qualitative inquiry, researchers can develop a more comprehensive understanding of how inclusivity—or the lack of it—affects learners' engagement and performance in assessment situations.

## Conclusion

This study demonstrated that MCQAs are not neutral tools of assessment but are shaped by linguistic, cultural, and gendered assumptions that reflect broader institutional and social practices. Through Critical Discourse Analysis, it was shown that the frequent use of male-centered language, academically demanding vocabulary, and Western cultural references creates subtle yet powerful forms of exclusion for many learners. These patterns disadvantage multilingual, culturally diverse, and gender-diverse students by shifting the focus away from the skills being tested toward familiarity with dominant forms of language and identity.

By identifying how these biases operate at textual, discursive, and social levels, this study contributes to a deeper understanding of fairness in educational assessment. The findings underscore the need for assessment practices that recognize linguistic diversity, cultural context, and gender representation. When tests are designed with inclusivity in mind, they become not only more equitable but also more valid, allowing students' true knowledge and abilities to be more accurately measured.

## Acknowledgement

The author acknowledges the use of artificial intelligence (AI) to help organize ideas and support the clarity of the discussion.

## References

- Abedi, J. (2006). Language issues in item development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 377–398). Lawrence Erlbaum Associates. [https://ebrary.net/76700/education/language\\_issues\\_item\\_development?utm](https://ebrary.net/76700/education/language_issues_item_development?utm)
- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association. <https://doi.org/10.1037/14192-000>
- Ahmed et. al. (2025, August). *Using thematic analysis in qualitative research*. ScienceDirect. <https://www.sciencedirect.com/science/article/pii/S2949916X25000222>
- Alasiri. (2024, November). *Fairclough's Three-Dimensional Model: A Critical Discourse Analysis of Denzel Washington's Inspiring Commencement*. ResearchGate. [https://www.researchgate.net/publication/386050281\\_Fairclough's\\_Three-Dimensional\\_Model\\_A\\_Critical\\_Discourse\\_Analysis\\_of\\_Denzel\\_Washington's\\_Inspiring\\_Commencement](https://www.researchgate.net/publication/386050281_Fairclough's_Three-Dimensional_Model_A_Critical_Discourse_Analysis_of_Denzel_Washington's_Inspiring_Commencement)
- Amoussou & Allagbe. (2018, January). *Principles, Theories and Approaches to Critical Discourse Analysis*. Best Scientific Journals Online. <https://www.arcjournals.org/pdfs/ijsell/v6-i1/2.pdf>
- Ayton. (2023, March 21). Chapter 5: Qualitative descriptive research – Qualitative research – a practical guide for health and social care researchers and practitioners. Qualitative

- Research – a practical guide for health and social care researchers and practitioners. <https://oercollective.caul.edu.au/qualitative-research/chapter/5/>
- Bourdieu, P. (1991). *Language and symbolic power* (J. B. Thompson, Ed.; G. Raymond & M. Adamson, (Trans.). Harvard University Press.
- Bothell. (2001). 14 Rules for Writing Multiple-Choice Questions. Byu Faculty Center. <https://testing.byu.edu/handbooks/14%20Rules%20for%20Writing%20Multiple-Choice%20Questions.pdf>
- Butler. (2018). Multiple-Choice Testing in Education: Are the Best Practices for Assessment Also Good for Learning? ScienceDirect. <https://www.sciencedirect.com/science/article/abs/pii/S2211368118301426>
- Centre for Teaching and Learning Editors. (n.d.). *IncludED: A guide to designing inclusive assessments*. Centre for Teaching and Learning. <https://www.cfl.ox.ac.uk/included-designing-inclusive-assessments>
- Culajara, J. P. M., & Culajara, C. J. (2023). A case study on teacher's challenges on inclusive mathematics classes in a secondary public school. *Journal of Mathematics Instruction, Social Research and Opinion*, 2(2), 145–157. <https://doi.org/10.58421/misro.v2i2.90>
- Culajara & Culajara. (2023, May). A Case Study on Teacher's Challenges on Inclusive Mathematics Classes in a Secondary Public School. ResearchGate. [https://www.researchgate.net/publication/370458458\\_A\\_Case\\_Study\\_on\\_Teacher%27s\\_Challenges\\_on\\_Inclusive\\_Mathematics\\_Classes\\_in\\_a\\_Secondary\\_Public\\_School](https://www.researchgate.net/publication/370458458_A_Case_Study_on_Teacher%27s_Challenges_on_Inclusive_Mathematics_Classes_in_a_Secondary_Public_School)
- D'Este, C. (2012, March). New views of validity in language testing. ResearchGate. [https://www.researchgate.net/publication/308782753\\_New\\_views\\_of\\_validity\\_in\\_language\\_testing](https://www.researchgate.net/publication/308782753_New_views_of_validity_in_language_testing)
- Department of Education. (2023). Administration of The National Achievement Test For Grade 12 For School Year 2022-2023. Education for All - Alapan 1 Elementary School. [https://alapan1es.com/wp-content/uploads/2023/01/DM\\_s2023\\_001-1.pdf](https://alapan1es.com/wp-content/uploads/2023/01/DM_s2023_001-1.pdf)
- Department of Education Matatag Editors. (2025). *National Achievement Test (NATG12) Reviewer for Grade 12*. Learning Pal. <https://learningpal.net/national-achievement-test-reviewer-deped/>
- Department of Education Matatag Editors. (2025). *NATG12 Practice Test in Reading and Writing Skills*. Google Docs. [https://docs.google.com/document/d/1cVa\\_ws6KAqiXoBJddr437NK2NH5175sP/edit](https://docs.google.com/document/d/1cVa_ws6KAqiXoBJddr437NK2NH5175sP/edit)
- Du Plessis & Du Plessis. (2009). A new and direct test of the 'gender bias' in multiple-choice questions. Stellenbosch Business School. <https://www.stellenboschbusiness.ac.za/publication/new-and-direct-test-gender-bias-multiple-choice-questions>
- Fairclough, N. (1995). *Critical discourse analysis: The critical study of language*. Longman.
- Fairclough, N. (2001). *Language and power* (2nd ed.). Longman.
- Gao, Y. (2022). Multiple-choice assessments in higher education: Design and implementation. *Journal of Educational Measurement*, 59(3), 231–248. <https://doi.org/10.1111/jedm.12345>
- Gao. (2022). *Multiple Choice and Objective Tests*. Center for Teaching and Learning. [https://ctl.mercer.edu/wp-content/uploads/sites/67/2022/07/v2-multiple-choice-tests-draft-Jan-25-2019\\_SALTISE.pdf](https://ctl.mercer.edu/wp-content/uploads/sites/67/2022/07/v2-multiple-choice-tests-draft-Jan-25-2019_SALTISE.pdf)
- Giraldo, F. (2020). Validity and Classroom Language Testing: A Practical Approach. *Colombian Applied Linguistics Journal*, 22(2), pp. 194-206. <https://www.redalyc.org/journal/3057/305766351020/html/>

- Gölbaşı, S. (2017). Critical Approach in Social Research: Fairclough's Critical Discourse Analysis. *The Online Journal of Communication and Media*, 3(4), 5-18. <https://www.scribd.com/document/722116290/CDA>
- Haladyna, T. (2022, September 28). Creating multiple-choice items for testing student learning. Education Resources Information Center. <https://files.eric.ed.gov/fulltext/EJ1372020.pdf>
- Hanesworth. (2019, January 24). *Inclusive assessment: Where next?* Advance HE. <https://www.advance-he.ac.uk/news-and-views/inclusive-assessment-where-next>
- Hoffman, M. (2025, April 1). What is Validity in Assessment? Michigan Language Assessment. <https://michiganassessment.org/what-is-validity-in-assessment/>
- Isnaeni, et.al.,. (2025). *View of critical discourse analysis of Norman Fairclough model on Dedi Mulyadi's angry sentences in tirto.id*. ARRUS Journal of Social Sciences and Humanities. <https://journal.arrus.id/index.php/soshum/article/view/4018/2362>
- Kaur, A., Noman, M., & Nordin, H. (2017). Inclusive assessment for linguistically diverse learners in higher education. *Assessment & Evaluation in Higher Education*, 42(5), 756–771. <https://doi.org/10.1080/02602938.2016.1187250>
- Kunnan, A. J. (2010). Test fairness and validation. In L. F. Bachman & A. Palmer (Eds.), *Language assessment in practice* (pp. 57–73). Oxford University Press.
- Lampe, S. & Tsaouse, B. (2010). Linguistic Bias in Multiple-Choice Test Questions. *Creative Nursing*. 16(2), 63-67. <https://journals.sagepub.com/doi/abs/10.1891/1078-4535.16.2.63>
- Lesage, E., Valcke, M. & Sabbe, E. (2013). Scoring methods for multiple choice assessment in higher education – Is it still a matter of number right scoring or negative marking? *Studies in Educational Evaluation*. 39(3), 188-193. <https://doi.org/10.1016/j.stueduc.2013.07.001>
- LSA Learning & Teaching Technology Consultants. (2023). A Refresh on Writing Quality Multiple Choice Questions. LSA Technology Services. <https://lsa.umich.edu/technology-services/news-events/all-news/teaching-tip-of-the-week/a-refresh-on-writing-quality-multiple-choice-questions.html>
- Lyon, et. al. (2022). A Sociological Lens on Linguistic Diversity: Implications for Writing Inclusive Multiple-Choice Assessments. *Teaching Sociology*, 51(4), 336-348. <https://doi.org/10.1177/0092055X221134126>
- Masoumi, G.A., Sadeghi, K. (2020). Impact of test format on vocabulary test performance of EFL learners: the role of gender. *Language Testing in Asia*, 10(2), <https://doi.org/10.1186/s40468-020-00099-x>
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Blackwell Publishing.
- Morlet, C., & Shah, P. (2025). Challenges in multiple-choice item construction: Ensuring fairness and validity. *International Journal of Testing*, 25(1), 101–121. <https://doi.org/10.1080/15305058.2024.9876543>
- National Achievement Test (NAT) Philippines: Complete guide. (2025, November 20). EdukSama. <https://eduksama.com/national-achievement-test-philippines/>
- Oc & Hassen. (2024, October 21). *Comparing The Effectiveness Of Multiple-Answer And Single-Answer Multiple-Choice Questions In Assessing Student Learning*. Taylor & Francis. <https://www.tandfonline.com/doi/full/10.1080/10528008.2024.2417106#abstract>
- Pandey, P. (2025). The Role of Teachers in Promoting Gender Inclusive Classrooms. *International Journal on Science and Technology*, 16(3). <https://doi.org/10.71097/IJSAT.v16.i3.7507>

- Phelan, J. & Phelan, J. (2010). Classroom Assessment Tasks and Tests. *International Encyclopedia of Education (Third Edition)*, 209-219. <https://doi.org/10.1016/B978-0-08-044894-7.00309-2>
- Rezai, A. (2022). Fairness in classroom assessment: Development and validation of a questionnaire. *Language Testing in Asia*, 12(17). <https://doi.org/10.1186/s40468-022-00162-9>
- 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). <https://doi.org/10.1186/s40468-022-00157-6>
- Shear, B. R. (2023). Gender bias in test item formats: Evidence from PISA 2009, 2012, and 2015 Math and Reading Tests. *Journal of Educational Measurement*. <https://doi.org/10.1111/jedm.12372>
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Longman. [https://books.google.com.ph/books/about/The\\_Power\\_of\\_Tests.html?id=Kj-gBAAAQBAJ&redir\\_esc=y](https://books.google.com.ph/books/about/The_Power_of_Tests.html?id=Kj-gBAAAQBAJ&redir_esc=y)
- Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? *Educational Researcher*, 37(4), 189–199. <https://doi.org/10.3102/0013189X08319569>
- Sunderland, J. (2006). *Language and gender: An advanced resource book*. Routledge.
- Sunderland, J., & McGlashan, M. (2012). Linguistic gendering in the internationalized university. In H. M. D. Baxter & J. Slembrouck (Eds.), *Language, gender and education* (pp. 65–84). Palgrave Macmillan.
- Tajik et.al.,. (2024, December). *Purposive Sampling*. ResearchGate. [https://www.researchgate.net/publication/391849280\\_Purposive\\_Sampling](https://www.researchgate.net/publication/391849280_Purposive_Sampling)
- University of Cambridge Editors. (2025). *Inclusive assessment & feedback*. Cambridge Centre for Teaching and Learning. <https://www.cctl.cam.ac.uk/mind-gap-toolkit/what-works/examples/assessment>
- University of Manitoba Editors. (n.d.). Creating multiple choice questions. The Centre for the Advancement of Teaching and Learning, University of Manitoba. <https://umanitoba.ca/centre-advancement-teaching-learning/support/multiple-choice-questions>
- Weimer, M. (2018, October 23). Advantages and disadvantages of different types of test questions. Faculty Focus, Higher Ed Teaching & Learning. <https://www.facultyfocus.com/articles/educational-assessment/advantages-and-disadvantages-of-different-types-of-test-questions/>
- Zhang, Y. (2025, September 2). Validity verification of the new TOEFL writing task based on classical test theory. arXiv.org. <https://arxiv.org/abs/2509.05347>

### Interview Guide Questions

Research Questions	Interview Questions	Possible Follow-up Questions
1. How inclusive is the selected MCQA design?	1. After reviewing the sample MCQA items, how inclusive do you think the wording and scenarios are for students of diverse gender identities and linguistic backgrounds?	a. Can you give a specific example from the sample items?
	2. Which parts of the items (words, examples, pronouns, cultural references) felt unclear, biased, or difficult to understand?	b. How might these parts affect students with different language abilities or gender identities?
2. What are the underlying factors that influence the selected MCQA design?	1. Based on the sample MCQA items, what factors do you think influenced how these questions were written (e.g., curriculum, time, training, norms)?	a. Which feature of the sample items suggests this influence or challenge?
	2. What challenges do you think teachers face when trying to make MCQ items more gender-inclusive and linguistically accessible?	b. How do these factors affect how students interpret or understand the items?
3. What inclusive modifications can be implemented to the MCQA design to better accommodate learners from diverse gender and linguistic backgrounds?	1. If you could revise any of the sample MCQA items, what changes would make them more gender-inclusive or easier to understand for multilingual learners?	a. Can you show or describe one part of the sample item you would change?
	2. What specific strategies or guidelines do you think should be added to help teachers create more inclusive MCQAs?	b. How would your suggested change improve fairness or clarity?