JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia)
DOI: http://dx.doi.org/10.15408/jp3i.v15i1.45690
http://journal.uinjkt.ac.id/index.php/jp3i

Exploring Gender Bias in the Dual Continua of Mental Health Measurement: Differential Item Functioning Analysis

Rahmat Aziz[1], Retno Mangestuti[1], Nada Alribdi[2], Maqdisi Firdausi Ali[3]
Department of Psychology, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia[1]
College of Education and Human Development, Princess Nourah bint Abdulrahman University, Saudi Arabia[2]
Master of Public Health, Monash University, Victoria, Melbourne, Australia[3]
mangestuti@uin-malang.ac.id

Abstract

Gender differences in mental health are widely reported, yet few studies have examined whether commonly used assessment instruments function equivalently across gender at the item level, particularly within the Dual Continua Model of Mental Health. The present study addresses this gap by evaluating gender-related measurement bias using regression-based Differential Item Functioning (DIF) analysis. A total of 1,674 university students from 32 Indonesian institutions completed the Azira Mental Health Scale (AMHS-24), which measures psychological well-being and psychological distress. DIF was assessed by controlling for latent trait levels to determine whether males and females respond differently to items at equivalent levels of psychological functioning. Results indicate that most well-being items exhibited no DIF, suggesting structural stability across gender for positive emotion, social relationship, and life satisfaction domains. In contrast, several distress items demonstrated uniform and nonuniform DIF, with one item showing strong magnitude. These findings suggest that the distress dimension may be more sensitive to gender-related response tendencies than the well-being dimension. By integrating the dual-continua framework with item-level psychometric analysis, this study clarifies whether observed gender differences reflect construct-relevant variance or differential item functioning, thereby contributing to the theoretical refinement and measurement fairness of mental health assessment.

Keywords: differential item functioning; dual continua model; gender bias; mental health assessment; psychological distress

Abstrak

Perbedaan gender dalam kesehatan mental telah banyak dilaporkan, namun sebagian besar penelitian masih berfokus pada perbandingan rerata skor tanpa menguji apakah instrumen pengukuran berfungsi setara pada tingkat item. Keterbatasan ini menjadi penting dalam kerangka Model Dual Continua Kesehatan Mental yang memandang kesejahteraan psikologis dan tekanan psikologis sebagai dua dimensi yang berbeda namun saling terkait. Penelitian ini bertujuan mengisi celah tersebut dengan mengevaluasi potensi bias gender melalui analisis Differential Item Functioning *(DIF) berbasis regresi. Sebanyak 1.674 mahasiswa dari 32 perguruan tinggi di Indonesia mengisi Azira Mental Health Scale (AMHS-24), yang mengukur kesejahteraan psikologis dan tekanan psikologis. Analisis DIF dilakukan dengan mengendalikan tingkat konstruk laten untuk menilai apakah laki-laki dan perempuan merespons butir secara berbeda pada tingkat kondisi psikologis yang setara. Hasil menunjukkan bahwa sebagian besar item kesejahteraan psikologis tidak menunjukkan DIF, mengindikasikan kestabilan struktural dimensi positif lintas gender. Sebaliknya, beberapa item tekanan psikologis menunjukkan DIF seragam maupun tidak seragam, dengan satu butir memiliki tingkat efek yang kuat. Temuan ini mengindikasikan bahwa dimensi tekanan psikologis relatif lebih sensitif terhadap perbedaan kecenderungan respons berbasis gender dibandingkan dimensi kesejahteraan. Dengan mengintegrasikan kerangka dual continua dan analisis psikometrik pada tingkat item, penelitian ini memberikan kontribusi terhadap penyempurnaan teoretis serta peningkatan keadilan pengukuran dalam asesmen kesehatan mental.*

Kata kunci: *fungsi butir diferensial, asesmen kesehatan mental, bias gender, model dual continua, tekanan psikologis*

Introduction

Gender bias in mental health assessments is a significant issue in contemporary psychometrics, as it may lead to systematic errors in diagnosis and treatment. Empirical evidence suggests that men and women exhibit different patterns of symptom reporting, which may affect the validity of commonly used assessment tools (Jane et al., 2007; Özel et al., 2024). For example, Women are more likely to report anxiety and emotional distress explicitly, increasing their likelihood of receiving specific psychiatric diagnoses, such as borderline personality disorder. In contrast, men tend to suppress or underreport psychological distress, often leading to undetected mental health conditions and inadequate interventions (Bacigalupe et al., 2024; Hennein et al., 2023). These discrepancies underscore the need for a more precise evaluation of gender bias in mental health measurement to enhance diagnostic accuracy and ensure equitable mental health care.

One approach to addressing this issue is the Dual Continua Model of Mental Health, which conceptualises psychological well-being and distress as two independent dimensions (Aziz & Mangestuti, 2023; Renshaw et al., 2024; Sholichatun et al., 2025). This model provides a more nuanced understanding of mental health beyond a single continuum of pathology to well-being. However, little research has examined whether this approach fairly represents gender differences, and the extent to which measurement tools based on this model are affected by gender bias remains unclear (Bruce & Weinraub, 2023; Vigod & Rochon, 2020). Furthermore, standardised mental health assessments may not adequately capture gender-specific symptom expressions, particularly in distress dimensions, leading to potential diagnostic inequities.

Research suggests traditional mental health scales may not equally capture symptom expressions across genders. Women's experiences of anxiety and depression tend to be more emphasised in clinical assessments, potentially increasing their likelihood of receiving specific psychiatric diagnoses. Meanwhile, distress indicators more common in men, such as emotional suppression or externalising behaviours, are often overlooked (Garb, 2021; Hennein et al., 2023). These disparities raise concerns about the fairness of mental health assessments and call for more refined psychometric approaches to detect potential biases in measurement.

The Differential Item Functioning (DIF) method identifies potential gender-based mental health measurement biases. By examining whether individuals with the same underlying psychological traits respond differently to specific test items, DIF helps distinguish between actual gender differences in mental health and biases in assessment tools (Garcia et al., 2021; Huang et al., 2022; Tsaousis et al., 2023). This is particularly important as mental health scales may not accurately capture gender-specific symptom expressions. While anxiety and depression are often emphasised in clinical assessments, distress manifestations more common in men, such as emotional suppression or externalising behaviours, may be overlooked. Addressing these disparities ensures that mental health assessments are fair and representative, reducing the risk of misdiagnosis and inadequate interventions across gender groups. Psychological constructs such as self-compassion may also influence how individuals respond to assessment items, as they are linked to healthier emotion regulation and mental health outcomes (Syafitri et al., 2024).

Although gender differences in mental health have been widely documented, particularly in relation to internalizing symptoms (Christiansen et al., 2022; Watkinson et al., 2024), most prior investigations have focused primarily on mean-level comparisons of psychological well- being or distress. Far fewer studies have examined whether the instruments used to measure these constructs operate equivalently across gender at the item level. This limitation is particularly critical within the dual-continua framework, which conceptualizes psychological well-being and psychological distress as distinct yet related dimensions (Elizabeth et al., 2022; Hermann et al., 2024). Without explicit evaluation of measurement fairness, observed gender differences may reflect Differential Item Functioning rather than substantive variation in latent constructs. Moreover, empirical work integrating the dual-continua model with regression-based Differential Item Functioning analysis remains limited, especially in non-Western contexts where

sociocultural norms surrounding emotional expression may influence response patterns (Anjara et al., 2021; Fitriana et al., 2023). Addressing this gap is essential to ensure that gender comparisons are psychometrically defensible and theoretically meaningful. The present study therefore examines gender-related DIF within the dual-continua model to evaluate item-level measurement fairness and to clarify whether observed differences reflect construct-relevant variance or potential measurement bias. In addition, considering evidence that work-life balance contributes to psychological well-being, incorporating such contextual variables may enhance the interpretability of gender differences in mental health assessments (Ogunola, 2022)

Accordingly, the present study investigates whether items within the psychological well-being and psychological distress domains of the AMHS-24 exhibit gender-related DIF. Using logistic regression to test uniform and nonuniform DIF, we evaluate whether individuals with equivalent levels of latent mental health respond differently based on gender. This approach allows us to determine whether observed gender patterns in distress arise from item-level bias rather than authentic differences in underlying mental health. The overarching goal is to strengthen the psychometric integrity of the instrument by ensuring that score differences reflect substantive psychological variation, not artifacts of measurement.

Methods

Research design

This study employed a quantitative cross-sectional survey design to examine gender-related measurement bias in the dual continua model of mental health. The design is appropriate for large-scale psychometric evaluation because it allows systematic assessment of response patterns across demographic groups (Jones, 2019). The research integrates descriptive analysis with trait-controlled Differential Item Functioning (DIF) to determine whether individual items operate differently for males and females after accounting for underlying levels of well-being or distress (Kwon & Sawatzky, 2017). Unlike group mean comparison designs, this approach focuses on item-level fairness and controls for latent trait variance, thereby enhancing the precision of bias detection (Verdam et al., 2016). The survey design also accommodates the diverse sample drawn from 32 universities, providing broad generalizability for evaluating the Azira Mental Health Scale (AMHS-24). Overall, this design combines population-level breadth with item-level analytical depth to ensure robust interpretation of gender effects in mental health measurement.

Procedures

Data for this study were collected in two phases. The first phase involved offline administration using paper-based surveys distributed in classroom settings at several universities in Malang under the supervision of trained research assistants to ensure standardized procedures. The second phase was conducted online through a secure digital platform and disseminated to students from various universities across Indonesia. All participants received information regarding the purpose of the study, their rights as voluntary respondents, and assurances of confidentiality before completing the instrument. To maintain balanced group comparisons for the DIF analysis, the number of male participants was used as the baseline, and female participants - whose numbers were greater in the initial pool - were selected proportionally based on male representation and institutional distribution. All procedures adhered to institutional ethical standards, and completed responses were checked for accuracy and securely stored for analysis.

Participants

A total of 1,674 university students from 32 higher education institutions across Indonesia participated in this study. These institutions represent a wide geographical distribution, spanning major urban centers such as Aceh, Medan, Padang, Palembang, Jakarta, Yogyakarta, Surabaya, Bandung, Cirebon, Malang, Pare-pare, Makassar, Madura, Mataram, Palangka Raya, Banjarmasin and Papua, ensuring broad

national representation. The sample consisted of 837 males and 837 females, providing balanced gender distribution suitable for Differential Item Functioning analysis. Gender distribution was proportionally balanced across participating institutions to ensure comparability between male and female groups in DIF analyses. Participants ranged in age from 19 to 25 years and were recruited from a wide range of academic programs, including those in the social sciences, natural sciences, and humanities. Data collection was conducted through voluntary participation during scheduled classroom sessions or institutional survey activities. All participants provided informed consent, and the final sample size exceeded recommended standards for item-level psychometric analyses.

Instrument

The Azira Mental Health Scale (AMHS-24) consists of 24 items assessing two major dimensions of the dual continua model: psychological well-being (12 items) and psychological distress (12 items). The well-being dimension comprises positive emotion, social relationship, and life satisfaction, whereas distress consists of anxiety, depression, and loss of control. All items are rated on a four-point Likert scale ranging from 1 (never) to 4 (always). Previous validation studies reported strong internal consistency, with Cronbach's alpha coefficients of .828 for psychological well-being, .875 for psychological distress, and .807 for the overall scale. Construct validity was supported through confirmatory factor analysis, with standardized loadings ranging from .302 to .718 (Aziz & Mangestuti, 2025). These indices indicate that the AMHS-24 is psychometrically suitable for item-level analysis such as Differential Item Functioning.

Data Analysis

Data analysis was conducted in two stages. First, descriptive statistics (mean, standard deviation, minimum, and maximum values) were calculated for all six subscales of the AMHS-24 to provide an initial overview of the distribution of psychological well-being and psychological distress. Total scores for each major dimension were computed by summing their respective 12 items, allowing the subsequent analyses to operate on aggregated indicators of trait levels.

The second stage examined Differential Item Functioning to identify potential gender-related item bias. DIF was evaluated using a trait-controlled regression approach, a method widely recommended for detecting uniform and nonuniform DIF in ordinal psychological scales. For each item, a linear regression model was estimated in which item responses were regressed on trait scores, gender, and the interaction between trait scores and gender. For well-being items, the psychological well-being total served as the matching variable, whereas the psychological distress total was used for distress items. Uniform DIF was identified when the gender coefficient was statistically significant after controlling for trait levels, while nonuniform DIF was identified when the interaction term was significant, indicating differences in slopes across gender groups.

A regression-based DIF framework was selected instead of an Item Response Theory (IRT) model because regression DIF procedures are well established for Likert-type psychological measures and do not require strict unidimensional assumptions (Jodoin & Gierl, 2001). This approach is particularly appropriate for multidimensional health and personality assessments in which item responses reflect attitudinal constructs rather than ability-based traits. To reduce potential contamination of the matching variable, total subscale scores were used as proxies for latent trait levels, consistent with conventional regression-based procedures. Given the strong internal consistency of both subscales, total scores were considered adequate approximations of the underlying constructs for trait-controlled DIF estimation. Nevertheless, total-score matching does not fully replicate latent trait estimation, and this limitation is acknowledged when interpreting the findings.

Effect size interpretation followed established guidelines for regression-based DIF analysis. In addition to statistical significance ($\alpha = .01$), the magnitude of the gender coefficient (β) was evaluated to determine practical relevance, with values below .10 interpreted as negligible, .10 – .30 as small, .30 – .50

as moderate, and above .50 as large DIF effects. This dual-criterion approach ensures that statistically detected DIF is interpreted in light of substantive impact rather than significance alone. Given the Likert-type nature of the data and the large sample size, OLS estimation was considered appropriate.

Results and Discussion

The findings of this study are presented in a structured sequence to provide a clear understanding of how gender influences item functioning within the dual-continua framework of mental health. The results begin with descriptive statistics summarizing the distribution of psychological well-being and psychological distress across the sample.

Descriptive Statistics

These descriptive patterns offer an initial perspective on the relative levels of positive functioning and distress experienced by participants. Following this, Differential Item Functioning analyses are presented to examine whether individual items operate differently for males and females after controlling for latent trait levels. Together, these analyses provide a comprehensive empirical basis for interpreting gender-related patterns in mental health measurement.

Table 1. Descriptive statistics of psychological well-being and psychological distress

Subscale	Item Count	Mean	SD	Min	Max
Psychological well-being (Total)	12	36.70	6.57	12	48.00
Positive emotion	4	12.27	2.47	4	16.00
Social relationship	4	12.31	2.77	4	16.00
Life Satisfaction	4	12.11	2.53	4	16.00
Psychological distress (Total)	12	27.52	7.67	12	48.00
Anxiety	4	10.60	2.56	4	16.00
Depression	4	8.29	3.23	4	16.00
Loss of control	4	8.64	3.17	4	16.00

Sources: Personal Data (2025).

The descriptive findings illustrate the distribution of mental health indicators across six subdomains aligned with the structure of the AMHS-24. All three well-being components—positive emotion, social relationships, and life satisfaction—show relatively similar means, indicating that students generally report balanced functioning across emotional, relational, and evaluative aspects of well-being. The total well-being score reflects moderate variability, suggesting consistent patterns of positive psychological functioning. For psychological distress, anxiety shows the highest subscale mean, whereas depression and loss of control exhibit wider variability, indicating uneven experiences of negative affect and self-regulatory difficulties among students. The greater dispersion in distress compared to well-being highlights a more heterogeneous expression of distress symptoms in the sample. These descriptive observations provide the contextual baseline for evaluating whether item responses remain comparable across gender when accounting for latent trait levels.

From a theoretical perspective, these findings can be interpreted as follows. The present study offers important insights into gender-related item functioning within the dual-continua model of mental health (Keyes, 2017; Keyes et al., 2012). The findings show that psychological well-being items display strong measurement invariance, whereas psychological distress items demonstrate more complex response patterns. This divergence reflects theoretical distinctions suggesting that well-being is a more universally experienced construct, while distress is influenced by individual and sociocultural factors (Diener et al., 2018). The DIF detected in distress items therefore likely reflects genuine differences in how males and females interpret or express negative emotional states rather than flaws in measurement. These results highlight the robustness of the well-being construct and the interpretive complexity inherent in distress assessment across gender groups.

DIF Analysis for Psychological Well-being

The DIF analysis was conducted separately for psychological well-being and psychological distress items in line with the dual-continua framework, which conceptualizes these domains as distinct constructs. This separation enables a focused evaluation of whether gender-related response differences reflect genuine variation in well-being or item-level bias. Table 2 presents uniform and nonuniform DIF statistics for the 12 well-being items. Rather than examining mean differences, the analysis assesses measurement equivalence by determining whether males and females with comparable levels of well-being endorse items similarly. Items flagged for DIF are further evaluated in terms of magnitude and implications for score interpretation.

Table 2. DIF analysis results for psychological well-being items (N = 1,674)

Item	β (Uniform DIF)	p (Uniform DIF)	β (Nonuniform DIF)	p (Nonuniform DIF)	Interpretation
PE 1	-	.354	-	.732	No DIF
PE 2	0.28	p < .005	-	.295	Small uniform DIF
PE 3	-0.26	p < .005	-	.476	Small uniform DIF
PE 4	-	.142	-0.01	p < .005	Small nonuniform DIF
SR 1	-	.676	-	.851	No DIF
SR 2	-	.758	0.04	p < .001	Small nonuniform DIF
SR 3	-	.710	-	.944	No DIF
SR 4	-	.611	-	.240	No DIF
LS 1	-	.807	-	.112	No DIF
LS 2	-	.729	-	.068	No DIF
LS 3	-	.660	-	.405	No DIF
LS 4	-	.145	-	.672	No DIF

Note. PE= Positive emotion, SR= Social relationship, LS= Life satisfaction. β values represent regression coefficients used to evaluate the magnitude of DIF effects.
Sources: Personal Data (2025).

The DIF analysis for the 12 psychological well-being items indicated that the majority of items function equivalently across male and female respondents. Overall, both uniform and nonuniform DIF effects were minimal, suggesting that individuals with comparable levels of well-being tend to endorse the items similarly regardless of gender. A small number of items showed statistically significant DIF; however, the magnitude of these effects ranged from negligible to small. Uniform DIF observed in several items reflects minor differences in item endorsement between gender groups, while nonuniform DIF indicates limited interaction effects with the underlying trait. Importantly, none of the items demonstrated moderate or large DIF effects, indicating that the well-being subscale remains largely invariant across gender. These findings support the use of well-being scores for gender-based comparisons, as the detected DIF is unlikely to meaningfully affect overall score interpretation. While minimal DIF is present in a few items, a different pattern emerges within the distress domain, which is examined in the following analyses.

The negligible to small DIF observed in the well-being subscale aligns with findings that positive psychological states, such as life satisfaction, social functioning, and emotional well-being, tend to be relatively stable across demographic groups (Nie et al., 2024; Velten et al., 2022). These constructs are widely regarded as core components of flourishing and are generally less constrained by cultural or gender norms. The small magnitude of DIF effects observed in a limited number of items suggests that these variations are unlikely to meaningfully influence total scores, consistent with previous research indicating that minor item-level differences do not compromise scale-level invariance (Ramli et al., 2024). Thus, the

AMHS-24 well-being items appear to capture fundamental aspects of positive functioning that are interpreted similarly by male and female respondents. Taken together, these findings support the conclusion that the well-being dimension of the AMHS-24 provides a gender-fair representation of positive psychological functioning, aligning with the study's objective to evaluate measurement equivalence within the dual-continua model.

DIF Analysis Results for Psychological Distress

While the well-being items demonstrated strong invariance, the distress items showed a more complex pattern. To further explore these differences, Table 3 presents the DIF results for the psychological distress items.

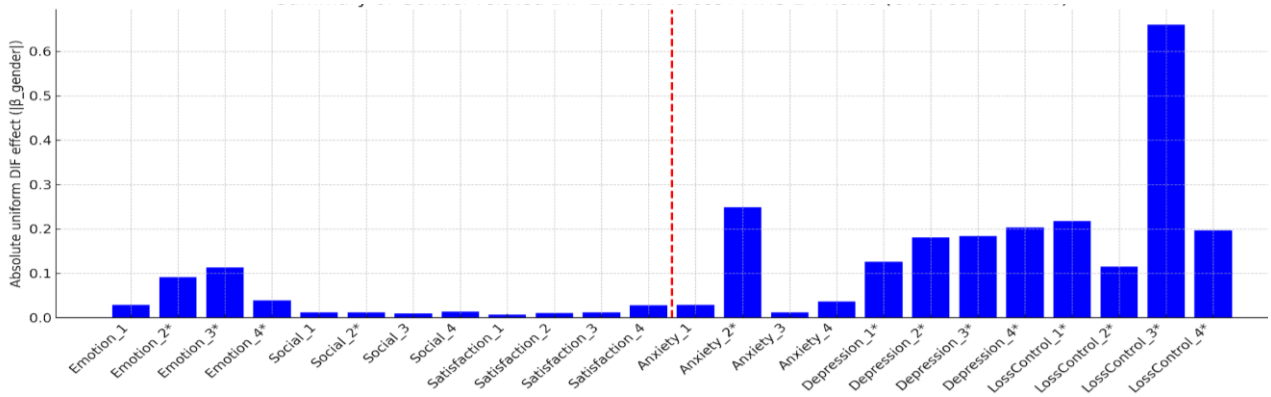
Table 3. DIF analysis results for psychological distress items (N = 1,674)

Item	β (Uniform DIF)	p (Uniform DIF)	β (Nonuniform DIF)	p (Nonuniform DIF)	Interpretation
AN 1	-	.436	-	.500	No DIF
AN 2	0.34	p < .001	-	.117	Moderate uniform DIF
AN 3	-	.711	-	.459	No DIF
AN 4	-	.220	-0.02	.022	Small nonuniform DIF
DE 1	0.18	p < .001	-	.337	Small uniform DIF
DE 2	0.36	p < .001	-	.401	Moderate uniform DIF
DE 3	0.41	p < .001	-	.773	Moderate uniform DIF
DE 4	0.38	p < .001	-	.171	Moderate uniform DIF
LC 1	0.35	p < .001	-	.291	Moderate uniform DIF
LC 2	0.15	p < .001	-	.574	Small uniform DIF
LC 3	0.62	p < .001	-	.055	Strong uniform DIF
LC 4	0.17	p < .001	-	.106	Small uniform DIF

Note. AN= Anxiety, DE= Depression, LC= Loss of control. β values represent regression coefficients used to evaluate the magnitude of DIF effects.
Sources: Personal Data (2025).

In contrast to the well-being domain, several psychological distress items demonstrated notable gender-related DIF. The results show that eight of the twelve distress items exhibited significant uniform DIF, indicating systematic differences in how males and females respond to symptoms even after accounting for overall distress levels. The magnitude of these effects ranged from small to strong, with several items reflecting moderate DIF. Items related to anxiety and depression, such as anxiety 2 and depression 2 - 4, showed moderate DIF, suggesting differences in response thresholds across gender. The most substantial finding emerged from loss of control 3, which demonstrated strong uniform DIF, indicating a pronounced discrepancy in how loss-of-control experiences are interpreted or reported.

The findings are interpreted as follows. The distress items exhibited greater DIF effects, with magnitudes ranging from small to strong, consistent with established evidence that gender shapes emotional expression and symptom reporting (Levkovich et al., 2025; Weiss et al., 2025). Females tend to be more expressive of internalizing symptoms, whereas males often adopt emotional suppression strategies due to social expectations of stoicism and restraint (Lee et al., 2025). These differing expressive norms may contribute to uniform DIF, even when underlying distress levels are comparable across gender groups. In addition, the presence of nonuniform DIF suggests that the relationship between distress severity and item endorsement varies between males and females. Rather than indicating measurement error, these patterns reflect broader psychological and sociocultural dynamics in how distress is experienced and communicated. To complement the item-level statistics presented above, the following figure provides a visual summary of the magnitude of uniform DIF across all AMHS-24 items.



Sources: Personal data (2025).

Figure 1. Magnitude of uniform DIF effects (β) across gender for AMHS-24 well-being and psychological distress Items

The figure displays the absolute value of the uniform DIF coefficient (|β_gender|) for each of the 24 AMHS-24 items. Items marked with an asterisk indicate statistically significant DIF at α = .01 based on the regression analysis. The dashed red line represents the threshold for moderate DIF (β = 0.30). The pattern shows that psychological well-being items tend to exhibit minimal DIF, whereas several psychological distress items demonstrate larger DIF effects, particularly those related to depression and loss of control. This visual pattern reinforces the earlier statistical findings and highlights several distress items that show moderate to strong DIF and require careful interpretation in relation to gendered psychological processes.

The strongest DIF was found in loss-of-control items, particularly those related to emotional regulation. Recent research on gendered coping styles shows that women are generally more likely to acknowledge emotional overwhelm, while men tend to minimize or underreport such experiences because of masculine norms that emphasize stoicism and restraint (García-Fernández et al., 2025; Levkovich et al., 2025). At the same time, higher male endorsement of several depressive symptoms at equivalent distress levels may reflect tension between socially imposed self-control and internal emotional conflict. These findings suggest that DIF in distress items may reflect psychological and sociocultural processes rather than solely measurement flaw highlighting how emotional norms shape self-report patterns and preparing the ground for interpreting these effects within Indonesia's collectivistic context.

The strongest DIF effect emerged for Item 23 ("Feel like crying when faced with problems"), which reflects an overt behavioral expression of emotional vulnerability. Unlike items describing internal affective states, this item refers to visible emotional display, which is socially and culturally gendered. Women are generally afforded greater social permission to express sadness openly, whereas men may experience implicit constraints due to norms emphasizing composure and restraint. Consequently, individuals with comparable levels of psychological distress may differ in endorsing this item because of gender-related display rules or response thresholds rather than differences in latent severity (Simon, 2020). Within Indonesia's collectivistic cultural context, where emotional expression is regulated by social expectations, such normative structures shape not only lived experiences of distress but also how respondents interpret self-report items (Anjara et al., 2021; Nila et al., 2025). Rather than indicating psychometric bias, the observed DIF may reflect culturally embedded emotion regulation norms influencing reporting patterns. In addition to sociocultural display rules, differential response thresholds or wording sensitivity may also contribute to this pattern, as overt behavioral expressions may activate different interpretive frames across gender.

Beyond the single strong DIF finding, several depression-related items (Depression 2-4) and one anxiety item (Anxiety 2), along with Loss of Control 1, exhibited moderate uniform DIF. Rather than indicating isolated item bias, this clustering suggests a patterned shift in response thresholds across gender within the distress domain. These items primarily describe internal affective states, which may be

differentially acknowledged or labeled across genders. Such patterns may reflect variations in symptom articulation or response style tendencies (Restar et al., 2020) including differences in emotional labeling, threshold sensitivity, or social desirability influences. Importantly, because the DIF effects are moderate and concentrated within the distress dimension, they point toward potential gender-related response tendencies embedded within specific symptom categories rather than pervasive measurement bias across the entire scale.

From a psychometric standpoint, the detection of DIF does not automatically necessitate item removal. The domain-specific pattern observed in this study further suggests that the dual-continua model not only distinguishes positive and negative functioning conceptually, but may also differentiate domains in terms of their susceptibility to sociocultural modulation at the item level. Contemporary measurement theory emphasizes that items showing small to moderate DIF can still be retained when they contribute substantively to construct representation and do not meaningfully distort overall score interpretation (Boone et al., 2014; Zumbo, 2007). Consistent with this view, the Standards for Educational and Psychological Testing (American Educational Research Association, 2014) recommend evaluating both statistical significance and practical impact before excluding items, as excessive item deletion may reduce construct coverage. Moreover, Zumbo (2007) argued that the presence of DIF should be viewed as an indicator of potential construct-relevant variance rather than as automatic evidence of measurement bias. The DIF-flagged distress items identified in this study therefore remain theoretically meaningful and psychometrically defensible, provided that their interpretation is guided by caution and contextual awareness.

These findings offer valuable implications for mental health practitioners, researchers, and university counseling services. Gender differences in item functioning suggest that variations in distress scores may partly reflect differences in emotional communication and socialization patterns rather than actual disparities in psychopathology (García-Fernández et al., 2025; Lee et al., 2025). Therefore, practitioners are encouraged to adopt gender-sensitive interpretation when screening for anxiety, depression, or emotional dysregulation to avoid over- or underestimating psychological impairment. From a research perspective, future studies should extend these findings by testing measurement invariance using confirmatory factor analysis (CFA) and item response theory (IRT) frameworks, ensuring the stability of the dual-continua model across demographic and cultural subgroups (Boone et al., 2014; Putnick & Bornstein, 2016). Ongoing psychometric refinement will not only improve the cultural adaptability of the AMHS-24 but also reinforce its capacity to serve as a comprehensive and equitable tool for assessing well-being and distress in Indonesia's diverse university populations.

Conclusion

This study examined gender-related measurement fairness within the dual-continua model of mental health using regression-based Differential Item Functioning (DIF) analysis. The findings indicate that psychological well-being items function equivalently across gender, supporting the structural stability and cross-group comparability of the positive functioning dimension. In contrast, several psychological distress items demonstrated uniform and, to a lesser extent, nonuniform DIF, particularly those reflecting emotional vulnerability and loss of control. These patterns suggest that observed gender differences in distress scores may partly reflect variations in response thresholds and socially conditioned norms of emotional expression rather than purely substantive differences in underlying psychological severity.

From a practical standpoint, these findings have important implications for both mental health assessment and instrument development. For practitioners, the presence of DIF in distress items indicates that score interpretation should be conducted with sensitivity to gender-related response tendencies, particularly when using self-report measures for screening or diagnostic purposes. Differences in distress scores between male and female respondents should not be interpreted solely as reflecting true disparities in psychological conditions without considering potential response bias. For instrument development, the results highlight the need to refine distress-related items, especially those involving overt emotional expression, to ensure that they capture a broader range of symptom manifestations across gender groups. Incorporating item-level evaluation procedures such as DIF analysis into routine validation processes can further enhance the fairness, accuracy, and interpretability of mental health instruments across diverse populations.

This study has several limitations that should be acknowledged. The cross-sectional design restricts the ability to assess the stability of DIF over time or across developmental contexts. In addition, the reliance on self-report data may introduce response biases, including social desirability and gendered norms of emotional disclosure, which can influence how participants interpret and endorse distress-related items. Methodologically, the regression-based DIF approach, while appropriate for Likert-type data, relies on total scores as proxies for latent traits and may not fully capture underlying construct variability compared to latent-variable approaches. Future research is therefore encouraged to employ longitudinal designs and alternative analytic frameworks, such as Item Response Theory or structural equation modeling, to validate and extend the present findings. Furthermore, cross-cultural and multi-group validation studies would help clarify whether the observed DIF reflects context-specific sociocultural dynamics or more generalizable patterns of gender-related response behaviour.

Acknowledgement

The authors sincerely acknowledge the financial support the Directorate General of Islamic Higher Education provided through Universitas Islam Negeri Malang for this research in 2024. This support has been instrumental in the successful completion of this study. The authors express their gratitude for the funding and resources made available.

Conflict of Interest

The authors declare that they have no conflicts of interest, financial or otherwise, that could have influenced this work. They affirm that no personal, professional, or institutional relationships create potential bias. This statement ensures the research's integrity, objectivity, and transparency, with findings based solely on academic and scientific principles.

Authors Contribution

RA and RM contributed to conceptualisation, methodology, data collection, and data analysis. NA provided critical insights and contributed to the interpretation of results. MFA participated in manuscript writing and review. All authors reviewed, edited, and approved the final version of the manuscript.

References

American Educational Research Association, A. P. A. & N. C. on M. in Education. (2014). *The Standards for Educational and Psychological Testing*. American Educational Research Association.

Anjara, S. G., Brayne, C., & Van Bortel, T. (2021). Perceived causes of mental illness and views on appropriate care pathways among Indonesians. *International Journal of Mental Health Systems*, *15*(1), 74. https://doi.org/10.1186/s13033-021-00497-5

Aziz, R., & Mangestuti, R. (2023). Profiling and action plan strategies for teachers' mental health. *Cyprus Turkish Journal of Psychiatry and Psychology*, *5*(2), 121–128. https://doi.org/10.35365/ctjpp.23.2.04

Aziz, R., & Mangestuti, R. (2025). Validating psychometric properties of dual-continua models for university student mental health assessment. *International Journal of Public Health Science*, *14*(1), 425–433. https://doi.org/10.11591/ijphs.v14i1.24601

Bacigalupe, A., Martín, U., Triolo, F., Sjöberg, L., Sterner, T. R., Dekhtyar, S., Fratiglioni, L., & Calderón-Larrañaga, A. (2024). Is the diagnosis and treatment of depression gender-biased? Evidence from a population-based aging cohort in Sweden. *International Journal for Equity in Health*, *23*(1), 252. https://doi.org/10.1186/s12939-024-02320-2

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. Springer Netherlands. https://doi.org/10.1007/978-94-007-6857-4

Bruce, M., & Weinraub, D. (2023). Implicit gender bias in the clinical judgment of psychopathy and personality disorders among licensed psychologists in the USA. *Journal of Personality Assessment*, *105*(6), 763–769. https://doi.org/10.1080/00223891.2023.2178928

Christiansen, D. M., McCarthy, M. M., & Seeman, M. V. (2022). Where sex meets gender: How sex and gender come together to cause sex differences in mental illness. *Frontiers in Psychiatry*, *13*, 856436. https://doi.org/10.3389/fpsyt.2022.856436

Diener, E., Oishi, S., & Tay, L. (2018). Advances in subjective well-being research. *Nature Human Behaviour*, *2*(4), 253-260. https://doi.org/10.1038/s41562-018-0307-6

Elizabeth, A., Hutton, J. L., Skues, J. A., Sullivan, L. Z., & Wise. (2022). Mental health research in the global construction industry: A scoping review using a dual-continuum model of mental health. *Mental Health and Prevention*, (28), 200249. https://doi.org/10.1016/j.mhp.2022.200249

Fitriana, T. S., Purba, F. D., Stolk, E., & Busschbach, J. J. V. (2023). Indonesia youth population norms for EQ-5D-Y-3 L, EQ-5D-Y-5 L and the PedsQL generic core scale: lower health related quality of life relates to high economic status and stress. *BMC Public Health*, *23*(1), 1124. https://doi.org/10.1186/s12889-023-16003-0

Garb, H. (2021). Race bias and gender bias in the diagnosis of psychological disorders. *Clinical Psychology Review*, *90*, 102087. https://doi.org/10.1016/j.cpr.2021.102087

Garcia, J., Gallagher, M., O'Bryant, S., & Medina, L. (2021). Differential item functioning of the Beck Anxiety Inventory in a rural, multi-ethnic cohort. *Journal of Affective Disorders*, *293*, 36–42. https://doi.org/10.1016/j.jad.2021.06.005

García-Fernández, M., Fuentes-Sánchez, N., Escrig, M. A., Eerola, T., & Pastor, M. C. (2025). Gender, emotion regulation, and cognitive flexibility as predictors of depression, anxiety, and affect in healthy adults. *Current Psychology*, *44*(7), 5685–5694. https://doi.org/10.1007/s12144-024-07240-6

Hennein, R., Poulin, R., Gorman, H., & Lowe, S. R. (2023). Gender Discrimination and Mental Health Among Health Care Workers: Findings from a Mixed Methods Study. *Journal of Women's Health (2002)*, *32*(7), 823–835. https://doi.org/10.1089/jwh.2022.0485

Hermann, V., Söderqvist, F., Karlsson, A. C., Sarkadi, A., & Durbeej, N. (2024). Mental health status according to the dual-factor model in Swedish adolescents: A cross sectional study highlighting associations with stress, resilience, social status and gender. *PLoS ONE*, *19*(3), 1–22. https://doi.org/10.1371/journal.pone.0299225

Huang, T.-W., Wu, P.-C., & Mok, M. (2022). Examination of gender-related differential item functioning through poly-BW indices. *Frontiers in Psychology*, *13*, 821459. https://doi.org/10.3389/fpsyg.2022.821459

Jane, J. S., Oltmanns, T. F., South, S. C., & Turkheimer, E. (2007). Gender bias in diagnostic criteria for personality disorders: an item response theory analysis. *Journal of Abnormal Psychology*, *116*(1), 166–175. https://doi.org/10.1037/0021-843X.116.1.166

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates Using an effect size measure with the logistic regression procedure for DIF eetection. *Applied Measurement in Education*, *14*(4), 329–349. https://doi.org/10.1207/S15324818AME1404_2

Jones, R. N. (2019). Differential item functioning and its relevance to epidemiology. *Current Epidemiology Reports*, *6*(2), 174–183. https://doi.org/10.1007/s40471-019-00194-5

Keyes, C. L. M. (2017). The dual continua model: The foundation of the sociology of mental health and mental illness. In T. L. Scheid & E. R. Wright (Ed.), *A handbook for the study of mental health: Social contexts, theories, and systems* (pp. 66–81). Cambridge University Press. https://doi.org/https://doi.org/10.1017/9781316471289.007

Keyes, C. L. M., Eisenberg D, Perry GS, Dube SR, Kroenke K, & Dhingra SS. (2012). The relationship of level of positive mental health with current mental disorders in predicting suicidal behavior and academic impairment in college students. *Journal of American College Health*, *60*(2), 126–133. https://doi.org/10.1080/07448481.2011.608393

Kwon, J.-Y., & Sawatzky, R. (2017). Examining gender-related differential item functioning of the Veterans Rand 12-item Health Survey. *Quality of Life Research*, *26*, 2877–2883. https://doi.org/10.1007/s11136-017-1638-x

Lee, J., Lim, Y., Seo, D. G., Lee, M. K., Schalet, B. D., Fischer, F., Rose, M., Kang, D., & Cho, J. (2025). A multinational comparison study of the patient-reported outcomes measurement information system anxiety, depression, and anger item bank in the general population. *International Journal of Methods in Psychiatric Research*, *34*(1), e70012. https://doi.org/10.1002/mpr.70012

Levkovich, I., Yatzkar, U., & Shenaar-Golan, V. (2025). Age and gender differences in emotional and behavioral functioning among youth referred to a psychiatric outpatient clinic at a public hospital. *Children*, *12*(6), 683. https://doi.org/10.3390/children12060683

Nie, Y.-Z., Zhang, X., Hong, N.-W., Zhou, C., Huang, Q.-Q., Cao, S.-Y., & Wang, C. (2024). Psychometric validation of the PERMA-profiler for well-being in Chinese adults. *Acta Psychologica*, *246*, 104248. https://doi.org/10.1016/j.actpsy.2024.104248

Nila, S., Webb, C., Suryobroto, B., & Carter, A. (2025). Cultural differences in self-reported empathy in Indonesia. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-16075-5

Ogunola, A. A. (2022). Quality of Work-Life and Work-Life Balance as Predictors of Employee Job Satisfaction. TAZKIYA Journal of Psychology, 10(1), 74-84 https://journal.uinjkt.ac.id/tazkiya/article/view/22499/pdf

Özel, B., Karakaya, E., Köksal, F., Altinoz, A. E., & Yilmaz-Karaman, I. G. (2024). Gender bias of antisocial and borderline personality disorders among psychiatrists. *Archives of Women's Mental Health*. https://doi.org/10.1007/s00737-024-01519-0

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Ramli, S. A., Zaremohzzabieh, Z., Idris, K., Bolong, J., & Abdullah, H. (2024). Psychometric validity, and measurement invariance of the PERMA model among youth in Malaysia. *Cogent Psychology*, *11*(1), 2316419. https://doi.org/10.1080/23311908.2024.2316419

Renshaw, T. L., Bolognino, S. J., & Clark, K. N. (2024). Exploring a dual-factor mental health screening model with children in grades 5–10. *School Psychology*. https://doi.org/10.1037/spq0000669

Restar, A., Jin, H., Breslow, A., Reisner, S. L., Mimiaga, M., & ... (2020). Legal gender marker and name change is associated with lower negative emotional response to gender-based mistreatment and improve mental health outcomes among trans populations. *SSM Popul Health*, *11*(100595). https://doi.org/10.1016/j.ssmph.2020.100595

Sholichatun, Y., Aziz, R., Baharuddin, B., Mulyadi, M., & Muhid, A. (2025). Building resilience to promote mental health in university students. *Health Education and Health Promotion*, *13*(2), 257–263. https://doi.org/http://dx.doi.org/10.58209/hehp.13.2.257

Simon, R. W. (2020). Gender, emotions, and mental health in the United States: Patterns, explanations, and new directions. *Society and Mental Health*, *2*(10), 97–111. https://doi.org/10.1177/2156869320926236

Syafitri, N., Lubis, R., Indrawan, Y. F., & Choong, T. C. (2024). Self-compassion: Unveiling mental health through emotion regulation in high-school students. TAZKIYA Journal of Psychology, 12(1), 1–16. https://doi.org/10.15408/tazkiya.v12i1.37794

Tsaousis, I., Alahmandi, M. T., & Asiri, H. (2023). Uncovering Differential Item Functioning effects using MIMIC and mediated MIMIC models. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1268074

Velten, J., Brailovskaia, J., & Margraf, J. (2022). Positive mental health scale: Validation and measurement invariance across eight countries, genders, and age groups. *Psychological Assessment*, *34*(4), 332–340. https://doi.org/10.1037/pas0001094

Verdam, M., Oort, F., & Sprangers, M. (2016). Item bias detection in the hospital anxiety and depression scale using structural equation modeling: Comparison with other item bias detection methods. *Quality of Life Research*, *26*(6), 1439–1450. https://doi.org/10.1007/s11136-016-1469-1

Vigod, S. N., & Rochon, P. A. (2020). The impact of gender discrimination on a woman's mental health. *EClinicalMedicine*, *20*(19), 100311. https://doi.org/https://doi.org/10.1016/j.eclinm.2020.100311

Watkinson, R. E., Linfield, A., Tielemans, J., Francetic, I., & Munford, L. (2024). Gender-related self-reported mental health inequalities in primary care in England: a cross-sectional analysis using the GP Patient Survey. *The Lancet Public Health*, *9*(2), e100–e108. https://doi.org/10.1016/S2468-2667(23)00301-8

Weiss, S. J., Nagle-Yang, S., Flynn, H., Cooper, B., Muzik, M., Simeonova, D. I., & Ozerdem, A. (2025). Gender differences in symptom profiles of individuals being treated for mood disorders. *Journal of Mood & Anxiety Disorders*, *12*, 100152. https://doi.org/10.1016/j.xjmad.2025.100152

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, *4*(2), 223–233. https://doi.org/10.1080/15434300701375832

Appendix

Instrument of AMHS-24 (English Version)

Aspect	Factors	No	Item in English
Psychological well-being	Positive emotion	01	Feel that everyday life is full of interesting things
		02	Feel that the future looks hopeful and promising
		03	Enjoying a relaxed and tension-free atmosphere
		04	Enjoy the activity or work done.
	Social relationship	05	Feeling that your friends appreciate your existence
		06	Feel comfortable in fostering social relationships with your friend
		07	Feel comfortable communicating with your friends.
		08	Feel comfortable because of the treatment from your friends.
	Life satisfaction	09	Feeling happy in living this life
		10	Feel satisfied in performing daily tasks.
		11	Enjoy what is going on in this life.
		12	Feel passionate about doing daily activities.
Psychological distress	Anxiety	13	Trying to calm down
		14	Being panicked in the face of unwanted situations
		15	You are finding yourself confused or frustrated.
		16	Feeling exhausted and helpless
	Depression	17	Feeling at rock bottom
		18	Spending time reflecting on something negative
		19	Using time to enjoy the feeling of despair
		20	Feeling useless in carrying out daily activities
	Loss of control	21	Assume that others would be better off if you were dead.
		22	Feeling that you have nothing to look forward to.
		23*	**Feel like crying when faced with problems.**
		24	You are feeling unable to restrain your anger.

*= DIF identified

Instrument of AMHS-24 (Indonesian Version)

Aspect	Factors	No	Item in Indonesian
Kesejahteraan psikologis	Emosi positif	01	*Merasa bahwa kehidupan sehari-hari penuh dengan hal yang menarik*
		02	*Merasa bahwa masa depan terlihat penuh harapan dan menjanjikan*
		03	*Menikmati suasana santai dan bebas dari ketegangan*
		04	*Menikmati kegiatan atau pekerjaan yang dilakukan*
	Hubungan social positif	05	*Merasa bahwa keberadaan anda dihargai oleh teman anda*
		06	*Merasa nyaman dalam membina hubungan sosial dengan teman*
		07	*Merasa nyaman berkomunikasi dengan teman anda*
		08	*Merasa nyaman karena perlakuan dari teman anda*
	Kepuasan hidup	09	*Merasa bahagia dalam menjalani kehidupan ini*
		10	*Merasa puas dalam melakukan tugas sehari-hari*
		11	*Menikmati apa yang terjadi dalam kehidupan ini*
		12	*Merasa bergairah dalam melakukan aktivitas sehari-hari*
Tekanan psikologis	Kecemasan	13	*Berusaha dan mencoba untuk tenang*
		14	*Menjadi panik ketika menghadapi situasi yang tidak diinginkan*
		15	*Menyadari diri sebagai orang yang bingung atau frustasi*
		16	*Merasa sebagai orang yang kelelahan dan tidak berdaya*
	Depresi	17	*Merasa berada pada titik yang terendah*
		18	*Menghabiskan waktu untuk merenungi sesuatu yang negatif*
		19	*Menggunakan waktu untuk menikmati rasanya putus asa*
		20	*Merasa sia-sia dalam melakukan aktivitas sehari-hari*
	Kehilangan kontrol	21	*Beranggapan bahwa orang lain akan lebih baik jika anda sudah mati*
		22	*Merasa tidak memiliki apa-apa dalam menatap masa depan*
		23*	***Merasa ingin menangis ketika menghadapi masalah***
		24	*Merasa tidak mampu menahan amarah*

***=** terindikasi DIF

http://journal.uinjkt.ac.id/index.php/jp3i