ORIGINAL ARTICLE

Rasch Analysis of the Indonesian PHQ-9 in ICU Patients with CHD

Aan Nuraeni[1], Suryani Suryani[2], Yanny Trisyani[1], Yulia Sofiatin[3], Helmy Hazmi[4], Firman Sugiharto[5]

[1] Department of Critical Care and Emergency Nursing, Faculty of Nursing, Universitas Padjadjaran, Bandung, West Java, Indonesia
[2] Department of Mental Health Nursing, Faculty of Nursing, Universitas Padjadjaran, Bandung 45363, Indonesia
[3] Department of Epidemiology, Faculty of Medicine, Universitas Padjadjaran, Bandung 45363, Indonesia
[4] Faculty of Medicine and Health Sciences Universiti Malaysia Sarawak (UNIMAS) 94300 Kota Samarahan Sarawak, Malaysia
[5] Doctoral Program, Faculty of Nursing, Universitas Padjadjaran, Bandung, West Java, Indonesia

ABSTRACT

Background: Screening and management of depression in patients with CHD is essential. PHQ-9 is a measurement that can be used, yet its reliability in investigating depression among CHD patients at the ICU has not been widely performed. **Purpose**: This study aimed to examine the construct validity of the Indonesian version of the PHQ-9 in CHD patients undergoing ICU treatment. **Methods**: Construct validity and differential item functioning (DIF) were analyzed using the Rasch model. A total of 102 CHD patients completed the Indonesian PHQ-9. Analyses included reliability and separation indices, infit-outfit MNSQ, ZSTD, point-measure correlations, unidimensionality, item-person maps, and DIF assessment. **Results**: Item and person reliabilities were 0.97 and 0.67, respectively, with a person separation strata of 2.25. Most items fit the Rasch model, with minor misfits (items 7 and 9 for MNSQ; items 1, 4, and 8 for ZSTD). All items had positive point-measure correlations and met at least one fit criterion. The measure explained 52.8% of the variance, with 5.6%–10.2% unexplained variance across the first five contrasts. Thresholds ranged from 1.35 to 2.38. Local dependence was low (r = 0.29), and no significant DIF was identified. Item 3 was the easiest; item 9 the most difficult. **Conclusion**: The Indonesian version of the PHQ-9 is a valid and reliable tool for detecting depression in ICU patients with CHD. However, its sensitivity to both mild and major depressive disorders remains limited. Instrument refinement, the use of complementary assessments, and a larger sample size are recommended for further analysis.
Malaysian Journal of Medicine and Health Sciences (2026) 22(2): 1-10.doi:10.47836/mjmhs.v22.i2.1535

Keywords: Coronary Heart Disease, Construct validity, Intensive-care, Rasch, PHQ-9

Corresponding Author:
Dr. Aan Nuraeni, BSN., MN
Email: aan.nuraeni@unpad.ac.id
Tel: +0227795596

INTRODUCTION

Coronary heart disease (CHD) is the most significant contributor to the burden of illness worldwide.[1] CHD continues to be the primary cause of mortality in countries across every socioeconomic group.[2] Although there is a global decline in incidence, the prevalence of risk factors associated with the disease has increased due to the increasing rates of hypertension in developing countries and obesity globally.[2], [3] This condition demands careful attention, as it could potentially increase the prevalence of the disease in the future.

Improvements and advances in the treatment of coronary heart disease (CHD), such as revascularization, the placing of drug-related stents, antiplatelet drugs, and complex imaging, have increased treatment effectiveness but have not prevented or reduced the occurrence of CHD.[4] Fioranelli et al. argued that the view of CHD has shifted from atherosclerosis as a passive process of cholesterol accumulation in the coronary arteries resulting in occlusion or thrombosis to the role of inflammation and the immune system in the formation and destabilization of the plaque that causes CHD.

[5] Worsening outcomes in current CHD patients are primarily associated with chronic inflammation rather than blood cholesterol levels. Chronic inflammation linked to heart disease is also believed to be connected to a high occurrence of depression.[5] Chronic depression is recognized as a significant cardiovascular risk factor that contributes to a poor outcome in patients with myocardial infarction.[5]–[7] Moreover, depression can result in significant functional impairments, poor adherence to treatment, reduced engagement in cardiac rehabilitation, and a diminished quality of life.[8]–[11] The condition emphasizes the importance of recognition, improved management of depression, as well as providing referrals for patients with Coronary Heart Disease (CHD).

Depression screening can be conducted at the onset of the patient's therapy in the acute phase, particularly when the patient is receiving treatment in the intensive care unit. The goal is to facilitate early depression management and referrals in a more cost-effective manner.[12] Furthermore, severe depressive illness frequently manifests in individuals following myocardial infarction or coronary artery bypass grafts.[12] Additional findings from one of the cardiac intensive care units in Indonesia revealed that a third of CHD patients experienced depression, with the majority of cases being classified as moderate to severe.[13]

Patients with CHD receiving treatment in the ICU have a precarious physical condition, requiring safe and effective examinations, including the appropriate selection of questionnaires. Recommended assessment tools for evaluating depression in CHD patients who have been tested were the Patient Health Questionnaire-9 (PHQ-9), Hospital Anxiety and Depression Scale-Depression (HADS-D), Beck Depression Inventory-II (BDI-II), and Cardiac Depression Scale (CDS).[14]–[16] All these measures have been confirmed as valid and reliable for detecting depression, particularly in patients with heart problems. However, the PHQ-9 is more effective in reducing missed diagnoses and is also more cost-effective than the HADS-D.[15], [17]

The BDI-II is more effective than HADS-D for detecting major depressive disorder (MDD).[18] Nevertheless, the BDI-II is not ideal for screening in clinical settings such as ICUs due to its time-consuming completion and assessment, especially in environments with fast patient turnover and limited resources. Similarly, although CDS is the only validated measure designed specifically for a cardiac population, it has great sensitivity and specificity. However, it is not yet ideal for clinical use, so a shorter instrument needs to be made.[14] Recently, a shorter version of CDS has been developed but has not been generally adopted, especially in Indonesia.

The PHQ-9 remains advantageous, especially in clinical settings, compared with other depression measurement tools. The PHQ-9 construct has been studied in non-coronary heart disease (CHD) patients and, to a lesser degree, in CHD patients. But to our knowledge, assessment of the instrument for CHD patients undergoing treatment in the ICU is still limited. Moreover, differences related to physical and psychological conditions, culture, and literacy levels can affect the accuracy of screening methods.[19]

Additionally, there has been little testing of the PHQ-9 construct in individuals with CHD. A study conducted in China analyzed the psychometric characteristics of the PHQ-9 using classical test theory (CTT).[15] Several problems with CTT compared to Rasch analysis are that it treats ordinal data as interval levels, scale ratings depend on the sample, and it's not possible to clearly confirm CTT assumptions.[20] A study has examined the psychometric characteristics of the PHQ-9 in CHD patients through Rasch analysis. However, these findings are not generalizable as they were specific to patients awaiting CABG procedures and included individuals with different sociodemographic backgrounds. This study was conducted to assess the construct validity and Differential Item Functioning (DIF) of the PHQ-9 instrument among CHD patients receiving care in the cardiac ICU in Indonesia, taking into account these specified factors.

MATERIALS AND METHODS

Study design and participants

This is a cross sectional study conducted from October 2023 to December 2023. The participants consist of patients with coronary heart disease (CHD) who were receiving treatment in a cardiac intensive care unit at a referral hospital in West Java, Indonesia. The sample selection employed a convenience sampling method, with the following inclusion criteria: (1) respondents aged 18 years or older, (2) diagnosed with either chronic coronary disease or acute coronary disease, and (3) having a relatively stable condition, as indicated by being free of or experiencing minimal chest pain, assessed using a numeric rating scale (NRS).

Instrument

The PHQ-9 is a self-administered questionnaire used to identify symptoms of depression, aligning with the criteria for major depressive disorder as outlined in DSM-IV[21]. The instrument consists of nine questions, which include: lack of interest, depressed mood, sleeping difficulties, tiredness, appetite problems, concentration problems, psychomotor agitation/retardation, negative feelings about self and suicidal ideation. The questionnaire uses a four point for respondents to indicated the frequency of their symptoms over the past two weeks.[21], [22], with the options including: 0 for "not at all," 1 for "several days," 2 for "more than half the days," and 3 for "nearly every day."

Data analysis
The data for the Rasch analysis was processed using Winsteps software, version 5.4.3. Rasch analysis is a mathematical method that accurately calculates assessments of item difficulty and individual aptitude using linear logit measures, specifically for ordinal data.[22] Essentially, the Rasch model requires that an instrument meets the specific calibration measurements, which includes the calibration of respondent, questions, and measurement scales.[23]

Construct validity in Rasch Model analysis is assessed by reliability measures such as person and item reliability, and separation. The validity of the Rasch model is determined by arguments related to data fit with the Rasch model, the uni-dimensionality test, rating scale analysis, good model fit, and local independence.[23], [24] In addition, for a valid measurement, it is necessary to identify any bias that may arise from the characteristics of the respondent. In Rasch modeling, detecting this bias is done through Differential Item Function (DIF) analysis.[23], [25] Then, we utilize the item person map (Wrightmap) to plot item difficulty and individual abilities on the same axis as the logits, enabling the assessment of item fit.

To test whether the data could fit the Rasch model, fit statistics, eg, information-weighted fit statistics (infit) mean square (MnSq) and outlier-sensitive fit statistics (outfit) MnSq were used. An item with infit or outfit MnSq out of the 0.5–1.5 range was considered a misfit. Additionally, the standardized fit statistic (ZSTD) value for infit and outfit, which shows the item's deviation from expected performance under the model should range between -2 to 2.[26], [27] The dimensionality of the scale was examined using principal component analysis (PCA) of the standardized residuals. To indicate unidimensionality, there should be an absence of any meaningful pattern in the residuals. Unidimensionality is assessed by examining the total variance explained by the main construct, the variance unexplained in comparisons from the first to the fifth after the main factor, and the eigenvalues from the Principal Component Analysis (PCA).

We analyzed the rating scale by evaluating the "Andrich Threshold" value. The investigation examined whether respondents could differentiate between the Likert scales 0 ("not at all"), 1 ("several days"), 2 ("more than one half of the days"), and 3 ("nearly every day"). Acceptable response scores should monotonically increase average difficulties (average measure) and step difficulties (step measure).[22] The value of Andrich Treshold that moves from NONE to negative, and leads to positive in sequence indicates the given option is valid (monotonic). Furthermore, the Andrich Treshold index criterion, ranging from 1.4 to 5 logits, indicates that the scale range of options is scalable and does not need to be merged or split.[28] Local dependency, referring to the items containing a latent trait other than depression, was tested using the correlation (r) of the Rasch residuals between each pair of items; r ≤0.3 was considered acceptable.[29]

RESULTS

Characteristic of Study
Table 1 presents the demographic and clinical characteristics of the respondents. The majority were male (64.7%) and most were within the 46–55 age group (46.1%). A large proportion of respondents (96.1%) were diagnosed with ST-elevation myocardial infarction (STEMI). Regarding the severity of depression as assessed by the PHQ-9, mild depression was most prevalent (44.1%), followed by moderate depression (36.3%). Major depressive disorder was reported in 7.8% of respondents, while moderately severe depression was the least common, affecting only 1.0% of the sample.

Table I: Characteristics of the Subjects

Characteristic	N (%) or Mean (SD)
Gender, n (%)	
Male	66 (64.7 %)
Female	36 (35.3 %)
Age, n (%)	
18 – 35 years old	5 (4.9 %)
36 – 45 years old	6 (5.9 %)
46 – 55 years old	47 (46.1 %)
56 – 65 years old	44 (43.1 %)
Diagnosis, n (%)	
STEMI	98 (96.1 %)
NSTEMI	4 (3.9 %)
Severity of depression, n (%)	
Minimal depression	11 (10.8 %)
Mild depression	45 (44.1 %)
Moderate depression	37 (36.3 %)
Moderately severe depression	1 (1.0 %)
Major depressive disorder	8 (7.8 %)

Abbreviation: STEMI, ST Elevation Myocardial Infarction; NSTEMI, Non ST Elevation Myocardial Infarction

Reliability Results
Reliability and Separation
Item reliability was high at 0.97, indicating excellent internal consistency across items. Person reliability was 0.67, reflecting acceptable consistency in respondents' responses. The Cronbach's alpha value ranged from 0.6 to 0.7, supporting the overall reliability of the instrument (Table 2). These indicators collectively demonstrate strong item quality and reasonable response consistency.[28]
The person separation index was 1.44, corresponding to a person reliability of 0.67. This suggests a limited ability to differentiate respondents into distinct levels of the measured trait, potentially indicating that the

respondents were not clearly stratified into separate levels.[30], [31] However, based on the person strata calculation (Table 2), the instrument is capable of distinguishing between two levels of trait severity.[23], [30] In contrast, the item separation index was 6.00, well above the acceptable threshold of 3, indicating that the instrument can differentiate up to six levels of item difficulty, demonstrating strong discriminative capacity.

Table II: Summary of Reliability Rasch measurement model on Patients Health Questionaire-9 (PHQ-9) (N=102)

Parameter (with quality criteria)	Scale (9 items)
RELIABILITY	
Model fit: summary of items	
Item mean in logits (criteria: 0.0 logits)	0.00 (SD=0.19)
Item reliability	0.97
Item separation reliability (criteria: good, 0.81-0.90); very good, 0.91-0.94; excel-lent, >0.94)	0.97
Item separation index (criteria: > 3)	6.00
Separate item strata = [(4 x separation in-dex) + I]/3 (criteria: fair, 2-3; good, 3-4; very good, 4-5; excellent, >5)	8.33
Model fit: summary of persons	
Person mean logits	-1.44 (SD=0.56)
Person reliability	0.67
Person separation reliability (criteria: air, 2-3; good, 0.81-0.90; very good, 0.91-0.94; excellent, >0.94)	0.67
Person separation index (criteria: >2)	1.44
Separate person strata = [(4 x separation index) + I]/3 (criteria: fair, 2-3; good, 3-4; very good, 4-5; excellent, >5)	2.25
Alpha cronbach (criteria ≥0.67)	0.67

Validity Results

Evaluating data fit to the model

An additional method to assess the model data fit is to examine the mean square (MNSQ), standardized fir statistics (ZSTD), and point measure correlation values as in Table 2. Regarding item fit (Table 2), both the infit and outfit MNSQ values, as well as the ZSTD values in the items and persons are close to the ideal model values of 1 and 0, respectively.[28] Table 3 displays the item fit for PHQ-9, which assesses depression in Coronary Heart Disease (CHD) patients undergoing hospital treatment.

Table III: Summary of Validity Rasch measurement model on Patients Health Questionaire-9 (PHQ-9) (N=102)

Parameter (with quality criteria)	Scale (9 items)
VALIDITY	
Model fit: summary of items	
Infit-Outfit MNSQ (criteria: 0.5 – 1.50)	1.12 – 0.96
Infit-Outfit ZSTD (Criteria+2 – (-) 2)	(-) 0.03 – (-) 0.29
Item model fit MNSQ range extremes (criteria: good, 0.5-1.5; very good, 0.71-1.4; excellent, 0.77-1.3)	
Point measure correlation (criteria: 0.4-0.85)	0.23 - 0.72
Model fit: summary of persons	
Infit-Outfit MNSQ (criteria: 0.5 – 1.50)	0.98 – 0.96
Infit-Outfit ZSTD (Criteria+2 – (-) 2)	(-) 0.14 – 0.02
Unidimensionality	
Raw variance in data explained by measures (criteria: > 50%)	52.8%
PCA eigenvalue for first contrast (crite-ria: > 2.0 indicates presence of anoth-er dimension; ≤2 supports unidimen-sional scale)	1.938
Unexplained variance in contrasts 1st-5th of PCA of residuals (criteria: fair, 10 – 15%; good, 5–10%; very good, 3–5%; excellent, < 3%)	5.6 – 10.2%
Adjacent threshold distance (criteria: Andrich Threshold index 1.4-5 logits)	Scale 0 – 1 was 0 to -2.04 = 2.04 logits Scale 1 – 2 was -2.04 to 0.34 = 2.38 logits Scale 2 – 3 was 0.34 to 1.69 = 1.35 logits
Average measure (criteria: increases monotonically across rating scale)	Yes
Probability curve graph (criteria: dis-tinct curve on each response category and each peak is ≥0.5 logits)	Yes
Local independency (criteria: <0.3)	0.29
Differential Item Functioning (DIF) (Criteria >5%)	Age: >5% Gender: >5% Diagnosis: >5%

Abbreviation: MNSQ, Mean Square; ZSTD, Z Standard

The analysis reveals that, according to the mean square (MNSQ) values for the infit and outfit, all items meet the established criterion except for items 7 and 9 (see Table 4). Whereas the standardized fit statistics (ZSTD) for infit and outfit, which should fall within the range of -2 to 2 to indicate good fit, do not meet the criterion for items 1, 4, and 8. Point measure correlation showed positive values for all items except for items 5, 7, 8, and 9, which has values below 0.4. No items across the three measures failed to meet the established criteria, indicating that all items are consistent with the Rasch model.[23]

Table IV: Item Fit for PHQ-9

Item	Measure or Logit (SE)	Infit Mean Square	Infit ZSTD	Outfit Mean Square	Outfit ZSTD	Point Measure Correlation
1. Little interest or pleasure.	0.14	0.65	-3.02	0.68	-2.72	0.64
2. Feeling down, depressed, or hopeless.	0.14	1.29	2.04	1.23	1.66	0.69
3. Trouble falling or staying asleep or sleeping too much.	0.14	0.74	-2.17	0.76	-1.97	0.72
4. Feeling tired or having little energy	0.15	0.53	-4.17	0.54	-3.94	0.65
5. Poor appetite or overeating	0.15	1.29	1.98	1.26	1.71	0.35
6. Feeling bad about self	0.18	1.09	0.66	1.00	0.07	0.49
7. Trouble concentrating	0.16	0.02	0.21	1.03	0.25	0.38
8. Moving or speaking so slowly	0.17	1.31	2.00	1.48	2.63	0.34
9. Thoughts better of dead	0.46	2.11	2.19	0.68	-0.33	0.23

Abbreviation: ZSTD, Z Standard; SE, Standard Error

Unidimensionality and rating scale

Table 2 shows that the measure explaind 52.8% of the raw variance in the data, satisfying one of the criteria for unidimensionality. Similarly, the unexplained variance in the first to fifth contrast should not be higher than 15%[28]. In this study, the unexplained variance ranged between 5.6% and 10.2%, staying well within the acceptable limit. The PCA eigenvalue for the first contrast is less than or equal to 2. Every facet of this dimensional analysis suggests a unidimensional scale. The politomial or scale values in the instrument were effectively distinguished by respondents as evidenced by their adherence to the Andrich Threshold index and their monotonical progression across the rating scale. This is reinforced by the presence of distinct strata items and persons with values exceeding two, demonstrating clear differentiation within the data.

Item and person mapping

Figure 1 illustrates the Wrightmap of items and persons in the PHQ-9. The majority of respondents have logit scores below 0.0, suggesting a tendency to agree with the statements in the questionnaire, indicative of signs of depression. The statements that were most difficult for respondents to concur with involved "thoughts better of dead" and "feeling about self." However, the issue of "trouble falling, staying asleep, or sleeping too much" was commonly experienced by nearly all respondents.

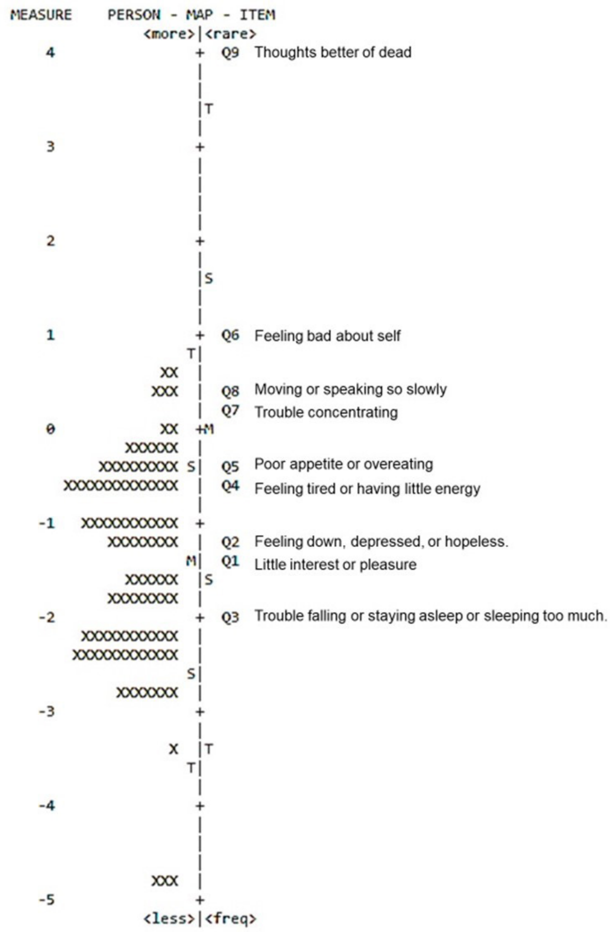

Figure 1: Wright map

Differential item functioning (DIF)
Individual characteristic can introduce bias towards a specific item or instrument. Rasch analysis addresses this by identifying such biases through Differential Item Functioning (DIF). An item exhibits bias related to a particular respondent's characteristic if the probability of the DIF for that item is less than 5% (<0.05).[25] This study examined three characteristics of the respondents: gender, age, and type of CHD. The likelihood of DIF occurring across the entire item on the PHQ-9 was found to be less than 0.05. No bias was detected in the three characteristics of the respondents.

Ethics approval and consent to participate
The study has received ethical approval from the Research Ethics Committee of Universitas Padjadjaran under reference number 253/UN6.KEP/EC/2023. All participants were given informed consent to secure their willingness to take part in the study. Participants also have the privilege to opt out of the procedure without facing any consequences. The data from each respondent was evaluated collectively to ensure the anonymity of the participants.

DISCUSSION
This study evaluated the psychometric properties of the PHQ-9 instrument. The PHQ-9 instrument was administered to CHD patients receiving care in an ICU at a hospital in Indonesia. The Rasch analysis demonstrated that the instrument generally fit the model well, with acceptable unidimensionality, model fit statistics, and reliable item performance. However, some challenges were noted in respondent-level reliability and the identification of misfitting items. Below is a detailed explanation of the validity of each construct:

Reliability and separation
The psychometric properties of the PHQ-9 in this study demonstrate strong internal consistency and item quality, with an excellent item reliability of 0.97, indicating that the items function consistently across the measurement scale. Although the person reliability was lower at 0.67, it still reflects acceptable consistency in respondents' responses and is supported by a Cronbach's alpha ranging from 0.6 to 0.7, further indicating good reliability (Table 2).[25] However, the person separation index of 1.44—corresponding to the person reliability—suggests limited capacity of the instrument to differentiate respondents into distinct levels of the latent trait, as it falls below the commonly accepted threshold of 2 for adequate separation [30], [31]. Nevertheless, based on the person strata presented in Table 2, the instrument remains capable of distinguishing between two levels of trait severity. This implies that the PHQ-9 can be used as a general screening tool to detect depression in ICU patients with coronary heart disease. However, it may not be sensitive enough to distinguish between different levels of depression severity. Therefore, it is recommended to use the PHQ-9 alongside other assessment tools or clinical interviews, especially when more detailed evaluation is needed.

Further analysis of item and person distribution highlights a potential misalignment in measurement targeting. The mean item difficulty was centered at 0.00 logits with a narrow standard deviation of 0.19, indicating homogeneity in item difficulty levels. In contrast, the person mean was notably lower at -1.44 logits with a wider standard deviation of 0.56, reflecting that most participants reported relatively mild depressive symptoms with some individual variability (Table 2). This discrepancy suggests poor targeting, a situation in Rasch measurement theory where item difficulty is not well-aligned with the respondents' ability levels, thereby reducing the precision of measurement. [32] As Linacre (2011) notes, when an instrument is not well-matched to the population's ability level, it may fail to accurately capture subtle variations, particularly at the lower end of the trait continuum.[33] Thus, while the PHQ-9 shows strong item reliability and acceptable overall consistency, its effectiveness in distinguishing levels of depression in ICU patients with coronary heart disease (CHD)—who tend to exhibit milder symptoms—appears limited. This misalignment underscores the need to consider revising or supplementing the PHQ-9 with items better targeted to this population in order to enhance sensitivity, validity, and clinical utility.

Data fit to the model
The current analysis evaluated item and person fit using multiple Rasch-based indices, including infit and outfit Mean Square (MNSQ), Z-standardized fit statistics (ZSTD), and point-measure correlations. While there is no universally agreed-upon standard for fit criteria in Rasch analysis[26], commonly accepted thresholds for MNSQ range from 0.5 to 1.5, which indicate productive measurement.[26] Our results show that the majority of items and persons fall within these limits, indicating that the PHQ-9 overall functions well in this critically ill sample.

The Point-measure Correlation, that requires all scores or measurements across variables be in the same unit of measurement showed positive results, indicating that item scoring is consistent with the underlying latent variable. This consistency is further evidenced by the entirely positive measurement results presented in Table 3 and 4. However, point-measure correlation should be interpreted as complementary rather than superior to MNSQ fit.[25]In our study, while the point-measure correlations were all positive, suggesting a coherent structure, the MNSQ values provided more specific insights into item-level behavior. Furthermore, Closer examination of individual item fit revealed that item 7 (trouble concentrating) demonstrated infit values below 0.5, while item 9 (thoughts of being better off dead) exhibited an outfit MNSQ of 2.11. While a low infit (item 7) may not inherently compromise validity, it indicates possible overfit or redundancy. An MNSQ infit less than 0.5 does not necessarily diminish the quality of the item but can affect reliability.[25] More notably, the high outfit for item 9 suggests unexpected response patterns that could reflect random error or external influences not captured by the latent trait. Item number 9 value exceeds 1.5 due to the majority of respondents disagreeing with the statement, indicating a high level of random distortion in the measurement system.[25], [26] This is probably a result of the strong religious culture in Indonesia that forbids suicidal behaviour or thoughts. Cultural and religious norms around discussing suicidal ideation—especially within populations that are critically ill and potentially more spiritually inclined—may result in underreporting due to stigma or perceived inappropriateness. These cultural influences likely contributed to the misfit and raise concerns about the clinical sensitivity and interpretability of this item in similar populations. Nurses or clinicians using the PHQ-9 in such contexts should be cautious in interpreting responses to item 9 and consider supplementing it with culturally adapted tools or follow-up questions during assessment.

Unidimensionality, local independence, and rating scale

Unidimensionality refers to the extent to which a test accurately evaluates the specific underlying characteristic that it claims to measure, without capturing undesired constructs.[26] The primary approach to assess unidimensionality in Rasch analysis is through calculating the principal component analysis of residuals (PCAR).[25], [26] This examination assesses the instrument's effectiveness in measuring the intended dimensions. Evidence of unidimensinality is demonstrated by a raw variance value of 50%, or more,[34] as observed in this study at 52.8%, along with a PCA eigenvalue of 1.938 for the first contrast (Table 2). These findings indicate that the instrument measures a single dimension, thereby confirming its unidimensionality.

The examination of local independence yielded a value of less than 3, suggesting that the items in the questionnaire are not interdependent.[35] The item distribution on the wrightmap (Figure 1) shows that the questions are not clustered together, indicating that they are independent of each other. Meanwhile, scale analysis is a method used to assess the respondent's ability to distinguish between responses on a given scale.[25], [26] The PHQ-9 consists of four scales: 0 ("not at all"), 1 ("several days"), 2 ("more than half the days"), and 3 ("nearly every day"). An evaluation using the Andrich Threshold Index was conducted to ascertain if participants could differentiate between these scores. The distance between adjacent threshold that satisfied the criteria ranged from 1.4 to 5 logits.[34] The results indicate that the Andrich threshold value in Table 2 transitioned from none to negative, and then to positive in a sequential pattern, with consecutive threshold distances ranging between 1.35 and 2.38. The finding suggest that the measurements condition were generally consistent, and the scale has been suitable for the respondents.

Wright map: Item and person mapping

The Wright map (Figure 1) illustrates the distributions of item difficulty and person ability levels.[23], [36] Figure 1 shows the difficulty levels of knowledge-related items on the right side of the vertical line on the Wright map. The items are arranged in ascending order of difficulty on a logit scale, with item 3 being the easiest and item 9 being the most challenging, reflecting the person's ability to respond to the items. On the left side of the map, the respondents' ability levels were displayed. Item 3, which assesses the frequency of sleep problems experienced by responders over two weeks, was the issue most commonly reported. Conversely, the item that the majority of respondents were least likely to endorse was related to having "thoughts of being better off dead."

The study indicated a correlation between sleep-related issues and depression in patients with CHD, highlighting sleep disturbances as the most prevalent issue among respondents. This study support the review by Mezick, Hall, and Matthews, which suggested that prolonged periods of inadequate sleep, irregular sleep patterns, and depression are closely connected issues that have been cited as potential risk factors for the development of cardiovascular disease and diabetes.[37]

Conversely, participants identified the item suggesting that death is preferable to life as the most difficult to agree on. This aligns to the study by Jiraniramai, Wongpakaran, Angkurawaranon, et al. conducted in Thailand.[22] In the current research, the reluctance of respondents to agree with this item is likely influenced by their religious beliefs, with the majority being Muslims. Islamic teachings forbid giving up on life and, more severely, having suicidal intentions.

Based on these findings, cultural and religious factors may obscure the detection of more severe depressive symptoms, as individuals may be reluctant to acknowledge or report them. This presents a clinical challenge for nurses and healthcare providers in accurately identifying patients with major depression, as cultural norms may suppress disclosure of critical symptoms such as suicidal ideation. Thus, Healthcare professionals should be culturally sensitive when assessing depression, particularly in populations where religious or cultural beliefs may inhibit the expression of certain symptoms. Reliance solely on standardized tools like the PHQ-9 may result in underestimation of severe depression. Therefore, incorporating culturally adapted approaches and conducting in-depth clinical interviews are recommended to ensure accurate diagnosis and appropriate mental health interventions.

Differential item functioning (DIF)
DIF is a crucial validity construct that was examined in this study, focusing on the probability scores of DIF for gender, age, and diagnosis, with each item showing a probability over 5% (>0.05). These results indicates no significant bias in the items based on the respondents' characteristics, suggesting that the instrument is suitable for use across these demograhics groups without requiring modification.[22] However, it is worth to note that some respondents were elderly, yet no bias was found based on age or related characteristics. This contrast with previous research, which have found DIF in older population and those with multimorbidities.[38] The disparities could be due to the relatively small sample size of this study.[39] Forkmann et al. have suggested that the sensitivity to detect DIF increases with samples size[38]; therefore, further biasas could be identified with the expansion of the study samples.

Strengths and Limitations
The strengths of this study lie in its homogeneous participants group, allowing the generalization of results to comparable participants. To the best of our knowledge, this study is the first to examine PHQ-9 among CHD patients receiving treatment in the ICU in Indonesia. Furthermore, the use of Rasch Analysis for construct testing offers several advantages over the classical test theory (CTT), enhancing the robustness of our findings. However, this study has several limitations. First, the small sample size restricts the ability to draw more sensitive conclusions about unidimensionality and differential item functioning (DIF). Second, the absence of identified comorbidities in the sample may introduce bias in the DIF analysis. Lastly, the study was conducted in a single setting, which limits the strength of generalization to other ICUs or healthcare contexts.

CONCLUSION

The Indonesian version of the PHQ-9 instrument, used in this study to detect depression among ICU patients with coronary heart disease (CHD), demonstrated good construct validity. This conclusion is supported by assessments of reliability and item/person separation, with item reliability rated as very good and person reliability as fairly good. The instrument also exhibited unidimensionality, and no issues related to local independence were identified, alongside the use of an appropriate rating scale.

However, certain considerations should be taken into account when applying this instrument to this specific population in Indonesia. Its effectiveness in distinguishing between levels of depression among ICU patients with CHD—who tend to exhibit milder symptoms—appears limited. Additionally, cultural factors pose challenges in validating severe depression, particularly due to respondents' reluctance or discomfort in selecting the item related to "thoughts of being better off dead." This suggests the need for either refinement of the instrument or the incorporation of complementary assessment tools to achieve more sensitive and precise results for this population. Furthermore, a larger sample size is necessary to allow for a more robust analysis of Differential Item Functioning (DIF).

ACKNOWLEDGEMENT

All authors thank Universitas Padjadjaran, Bandung, West Java, Indonesia, who has facilitated the study. We would also like to thank the nurses and respondents who agreed to participate in the research.

REFERENCES

1. H. Dai et al., "Global, regional, and national burden of ischaemic heart disease and its attributable risk factors, 1990-2017: Results from the Global Burden of Disease Study 2017," Eur. Hear. J. - Qual. Care Clin. Outcomes, vol. 8, no. 1, pp. 50–60, 2022, doi: 10.1093/ehjqcco/qcaa076.
2. A. N. Nowbar, M. Gitto, J. P. Howard, D. P. Francis, and R. Al-Lamee, "Mortality from ischemic heart disease: Analysis of data from the world health organization and coronary artery disease risk factors from NCD risk factor collaboration," Circ. Cardiovasc. Qual. Outcomes, vol. 12, no. 6, pp. 1–11, 2019, doi: 10.1161/CIRCOUTCOMES.118.005375.
3. W. Adisasmito, V. Amir, A. Atin, A. Megraini, and D. Kusuma, "Geographic and socioeconomic disparity in cardiovascular risk factors in Indonesia: Analysis of the basic health research 2018," BMC Public Health, vol. 20, no. 1004, pp. 1–13, 2020, doi: 10.1186/s12889-020-09099-1.
4. M. M. Burg et al., "The 'Perfect Storm' and Acute Coronary Syndrome Onset: Do Psychosocial Factors Play a Role?," Prog. Cardiovasc. Dis., vol. 55, no. 6, pp. 601–610, 2013, doi: https://doi.org/10.1016/j.pcad.2013.03.003.
5. M. Fioranelli, A. G. Bottaccioli, F. Bottaccioli, M. Bianchi, M. Rovesti, and M. G. Roccia, "Stress and inflammation in coronary artery disease: A review psychoneuroendocrineimmunology-based," Front. Immunol., vol. 9, no. SEP, 2018, doi: 10.3389/fimmu.2018.02031.
6. J. H. Lichtman et al., "Depression as a risk factor for poor prognosis among patients with acute coronary syndrome: systematic review and recommendations: a scientific statement from the American Heart Association," Circulation, vol. 129, no. 12, pp. 1350–1369, 2014, doi: 10.1161/CIR.0000000000000019.
7. S. L. Kuhlmann et al., "Prevalence, 12-Month Prognosis, and Clinical Management Need of Depression in Coronary Heart Disease Patients: A Prospective Cohort Study," Psychother.

Psychosom., vol. 88, no. 5, pp. 300–311, 2019, doi: 10.1159/000501502.
8. A. Nuraeni, R. Mirwanti, A. Anna, and I. Nurhidayah, "Determinant Factors of Depression in Patients with Coronary Heart Disease," Padjadjaran Nurs. J., vol. 7, no. 3, pp. 246–254, 2019, doi: DOI : 10.24198/jkp.
9. D. Nurhamsyah, Y. Trisyani, and A. Nur'aeni, "Factors Affecting Quality of Life of Post-Acute Coronary Syndrome Patients in Indonesia," Open Access Maced. J. Med. Sci., vol. 9, no. T6, pp. 203–206, 2021.
10. V. Vaccarino et al., "Depression and coronary heart disease : 2018 position paper of the ESC working group on coronary pathophysiology and microcirculation," Eur. J. Cardiovasc. Nurs., vol. 41, pp. 1687–1696, 2020, doi: 10.1093/eurheartj/ehy913.
11. T. Mьller-Tasch et al., "Depression Is a Major Determinant of Quality of Life in Patients With Chronic Systolic Heart Failure in General Practice," J. Card. Fail., vol. 13, no. 10, pp. 818–824, Dec. 2007, doi: 10.1016/j.cardfail.2007.07.008.
12. D. M. Colquhoun et al., "Screening, referral and treatment for depression in patients with coronary heart disease: A consensus statement from the national heart foundation of Australia," Med. J. Aust., vol. 198, no. 9, pp. 1–7, 2013, doi: 10.5694/mja13.10153.
13. R. Amni, Y. Akbar, A. Nuraeni, and M. R. Akbar, "Tingkat depresi pada pasien sindrom koroner akut di Unit Perawatan Intensif Jantung : analisis menggunakan skala depresi jantung [Level of depression in acute coronary syndrome patients in the Cardiac Intensive Care Unit: analysis using the cardiac depres," J. Keperawatan, vol. 20, no. 3, pp. 31–41, 2022, doi: 10.35874/jkp.v20i3.1063.
14. D. L. Hare et al., "Psychometric assessment of the Cardiac Depression Scale Short Form in cardiac outpatients," Eur. J. Cardiovasc. Nurs., vol. 16, no. 3, pp. 249–255, 2017, doi: 10.1177/1474515116652759.
15. J. Yuan, R. Ding, L. Wang, L. Sheng, J. Li, and D. Hu, "Screening for depression in acute coronary syndrome patients : A comparison of Patient Health Questionnaire-9 versus Hospital Anxiety and Depression," J. Psychosom. Res., vol. 121, no. March, pp. 24–28, 2019, doi: 10.1016/j.jpsychores.2019.03.018.
16. D. L. Hare and C. R. Davis, "Cardiac depression scale: Validation of a new depression scale for cardiac patients," J. Psychosom. Res., vol. 40, no. 4, pp. 379–386, 1996, doi: 10.1016/0022-3999(95)00612-5.
17. F. Kendel, M. Wirtz, A. Dunkel, E. Lehmkuhl, R. Hetzer, and V. Regitz-Zagrosek, "Screening for depression: Rasch analysis of the dimensional structure of the PHQ-9 and the HADS-D," J. Affect. Disord., vol. 122, no. 3, pp. 241–246, 2010, doi: 10.1016/j.jad.2009.07.004.
18. A. Bunevicius, M. Staniute, J. Brozaitiene, and R. Bunevicius, "Diagnostic accuracy of self-rating scales for screening of depression in coronary artery disease patients," J. Psychosom. Res., vol. 72, no. 1, pp. 22–25, 2012, doi: 10.1016/j.jpsychores.2011.10.006.
19. Y. K. Woldetensay et al., "Validation of the Patient Health Questionnaire (PHQ-9) as a screening tool for depression in pregnant women: Afaan Oromo version," PLoS One, vol. 13, no. 2, pp. 1–15, 2018, doi: 10.1371/journal.pone.0191782.
20. J. Hobart and S. Cano, "Improving the evaluation of therapeutic interventions in multiple sclerosis: The role of new psychometric methods," Health Technol. Assess. (Rockv)., vol. 13, no. 12, 2009, doi: 10.3310/hta13120.
21. K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The PHQ-9: Validity of a brief depression severity measure," J. Gen. Intern. Med., vol. 16, no. 9, pp. 606–613, 2001, doi: 10.1046/j.1525-1497.2001.016009606.x.
22. S. Jiraniramai, T. Wongpakaran, C. Angkurawaranon, W. Jiraporncharoen, and N. Wongpakaran, "Construct validity and differential item functioning of the phq-9 among health care workers: Rasch analysis approach," Neuropsychiatr. Dis. Treat., vol. 17, pp. 1035–1045, 2021, doi: 10.2147/NDT.S271987.
23. B. Sumintono and W. Widhiarso, Aplikasi Pemodelan Rasch pada Assessment Pendidikan [Application of Rasch Modeling to Education Assessment]. Bandung: Trim Komunikata .ublishing House, 2015.
24. B. Sumintono, "Rasch Model Measurements as Tools in Assesment for Learning," vol. 173, no. Icei 2017, pp. 38–42, 2018, doi: 10.2991/icei-17.2018.11.
25. B. Sumintono and W. Widhiarso, Application of the Rasch Model to Research in the Social Sciences, Revised Ed., no. November. Cimahi: Trim Komunikata Publishing House, 2014.
26. V. Aryadoust, L. Y. Ng, and H. Sayama, "A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research," Lang. Test., vol. 38, no. 1, pp. 6–40, 2021, doi: 10.1177/0265532220927487.
27. W. Ling Lee, K. Chinna, and B. Sumintono, "Psychometrics assessment of HeartQoL questionnaire: A Rasch analysis," Eur. J. Prev. Cardiol., vol. 28, no. 12, pp. E1–E5, 2021, doi: 10.1177/2047487320902322.
28. B. Sumintono and W. Widhiarso, Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisi revisi). Trim Komunikata Publishing House, 2014.
29. K. B. Christensen, G. Makransky, and M. Horton, "Critical values for Yen's Q 3: Identification of local dependence in the Rasch model using residual correlations," Appl. Psychol. Meas., vol. 41, no. 3,

pp. 178–194, 2017.
30. B. D. Wright and G. N. Masters, Rating scale analysis. MESA press, 1982.
31. M. A. P. Souza, W. J. Coster, M. C. Mancini, F. C. M. S. Dutra, J. Kramer, and R. F. Sampaio, "Rasch analysis of the participation scale (P-scale): usefulness of the P-scale to a rehabilitation services network," BMC Public Health, vol. 17, no. 1, p. 934, 2017, doi: 10.1186/s12889-017-4945-9.
32. T. G. Bond and C. M. Fox, Applying the Rasch model: Fundamental measurement in the human sciences. Psychology Press, 2013.
33. M. Linacre, "A user's guide to WINSTEPS: Rasch-model computer programs," winstep.com, 2011.
34. W. P. Fisher, "Rating scale instrument quality criteria," Rasch Meas. Trans., vol. 21, no. 1, p. 1095, 2007.
35. E. Brentani and G. Silvia, "Unidimensionality in the Rasch model how to detect and interpret," Stat. anno LXVII, n. 3, vol. 67, no. 3, pp. 253–261, 2007, doi: 10.6092/issn.1973-2201/3508.
36. N. S. Jimam et al., "Use of Rasch Wright map to understand the quality of Healthcare Workers' Knowledge, Attitudes, and Practices for Uncomplicated Malaria (HKAPIUM)," J. Pharm. Bioresour., vol. 18, no. 3, pp. 237–244, 2021, doi: 10.4314/jpb.v18i3.8.
37. E. J. Mezick, M. Hall, and K. A. Matthews, "Are sleep and depression independent or overlapping risk factors for cardiometabolic disease?," Sleep Med. Rev., vol. 15, no. 1, pp. 51–63, 2011, doi: 10.1016/j.smrv.2010.03.001.
38. T. Forkmann, S. Gauggel, L. Spangenberg, E. Br∆hler, and H. Glaesmer, "Dimensional assessment of depressive severity in the elderly general population: Psychometric evaluation of the PHQ-9 using Rasch Analysis," J. Affect. Disord., vol. 148, no. 2–3, pp. 323–330, 2013, doi: 10.1016/j.jad.2012.12.019.
39. J.-S. Lai, J. Teresi, and R. Gershon, "Procedures for the analysis of differential item functioning (DIF) for small sample sizes.," Eval. Health Prof., vol. 28, no. 3, pp. 283–294, Sep. 2005, doi: 10.1177/0163278705278276.