












Research Article

The Arabic Version of the Patient Health Questionnaire-2 (PHQ-2): Psychometric Evaluation Among Mothers of Children With Intellectual Disabilities

Amira Mohammed Ali ¹, Saeed A. Al-Dossary ², Musheer A. Aljaberi ^{3,4},
 Heba Emad El-Gazar ⁵, Carlos Laranjeira ^{6,7,8}, Haitham Khatatbeh ⁹,
 Mohamed Ali Zoromba ^{10,11}, Rasmieh Alamer ⁹, Faten Amer ¹², Annamaria Pakai ¹³,
 and Feten Fekih-Romdhane ^{14,15}

¹Department of Psychiatric Nursing and Mental Health, Faculty of Nursing, Alexandria University, Smouha, Alexandria 21527, Egypt

²Department of Psychology, College of Education, University of Ha'il, Ha'il 1818, Saudi Arabia

³Research Centre Innovations in Care, Rotterdam University of Applied Sciences, Rotterdam, the Netherlands

⁴Department of Internal Medicine, Section Nursing Science, Erasmus University Medical Center (Erasmus MC), Rotterdam 3015 GD, the Netherlands

⁵Department of Nursing Administration, Faculty of Nursing, Port Said University, Port Said, Egypt

⁶School of Health Sciences, Polytechnic University of Leiria, Campus 2, Morro do Lena, Alto do Vieiro, Apartado 4137, Leiria 2411-901, Portugal

⁷Centre for Innovative Care and Health Technology (ciTechCare), Polytechnic University of Leiria, Campus 5, Rua das Olhalvas, Leiria 2414-016, Portugal

⁸Comprehensive Health Research Centre (CHRC), University of Évora, Évora 7000-801, Portugal

⁹Faculty of Nursing, Yarmouk University, Irbid, Jordan

¹⁰College of Nursing, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia

¹¹Psychiatric and Mental Health Nursing Department, Mansoura University, Mansoura, Egypt

¹²Faculty of Pharmacy, An-Najah National University, Nablus, State of Palestine

¹³Institute of Nursing Sciences, Basic Health Sciences and Health Visiting, Faculty of Health Sciences, University of Pécs, Pécs 7622, Hungary

¹⁴Faculty of Medicine of Tunis, Tunis Al Manar University, Tunis, Tunisia

¹⁵Department of Psychiatry Ibn Omrane, The Tunisian Center of Early Intervention in Psychosis, Razi Hospital, Tunis, Tunisia

Correspondence should be addressed to Annamaria Pakai; annamaria.pakai@etk.pt

Received 14 November 2024; Revised 19 May 2025; Accepted 4 June 2025

Academic Editor: Eirini Lidoriki

Copyright © 2025 Amira Mohammed Ali et al. Nursing Research and Practice published by John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Aim: Mothers of children with intellectual disabilities are particularly vulnerable to mental distress due to demanding and exhausting caregiving. However, in the Arab world, they are seldom screened for depression because of limited diagnostic resources. Addressing the urgent need for brief and reliable screening tools, this study evaluated the psychometric properties of the Arabic version of the Patient Health Questionnaire-2 (PHQ-2) among 85 Saudi mothers.

Design: A cross-sectional study.

Methods: The construct, convergent, and divergent validity of the PHQ-2 was examined through a latent variable model (LVM), while its cutoff score was examined through receiver-operating characteristic (ROC) curve analysis.

Results: The unidimensional PHQ-2 (item loadings > 0.7) was positively predicted by stress and negatively predicted by high mood and happiness, supporting its convergent and divergent validity. The PHQ-2 effectively predicted low mood, poor sleep

quality, nightmares, high stress, low general physical health, and willingness to join a psychological support program (area under the curve [AUC] range = 0.72–0.84, p values < 0.001). The best balance between sensitivity and specificity was achieved at the PHQ-2 threshold ≥ 2.5 , while the cutoff ≥ 3.5 demonstrated a higher positive predictive value (PPV) for all outcomes (range = 30.0–78.8 vs. 23.0–70.8).

Conclusions: The PHQ-2 is a brief, valid tool, which at cutoffs ≥ 2.5 and ≥ 3.5 can reliably detect clinically significant depression and related psychological and physical adverse effects. Mothers scoring ≥ 3.5 may require a clinician-based examination for depression, and they may benefit from specific mental health literacy interventions. However, the results should be interpreted with caution given convenience sampling, a small sample size, and elevated distress levels in the current population. These limitations highlight the need to replicate the study with larger, randomly selected samples from more diverse populations.

Implications for Practice: Nurses can efficiently screen for depression and its mental/physical sequelae, as well as monitor response to treatment using only two items. The study provided two well-interpreted cutoffs of the PHQ-2, with real-world implications for mental health screening in under-resourced settings.

Reporting Method: The study adhered to STROBE guidelines.

Patient or Public Contribution: No patient or public contribution.

Keywords: Arabic/Saudi; cutoff score/cutoff point; depression/low mood; mothers/children with mental or intellectual disabilities; Patient Health Questionnaire 2-item scale/PHQ-2; psychometrics/construct validity; receiver-operating characteristic curve/ROC analysis; stress/sleep/nightmares/physical health/psychological support

Summary

- What This Paper Adds:
 - Validating a brief scale for use in busy and resource-limited clinical settings.
 - The PHQ-2 may lower nonresponse and delayed diagnosis of depression in 22 Arab countries.

1. Introduction

Depression is a multifactorial disorder that arises from a complex interplay among genetic, sociodemographic, lifestyle, personality, and environmental factors [1, 2]. Major depressive disorder (MDD), one of the top 10 causes of years lost to disability and a second cause of global disease burden, prevails in 5% of the general population—up to 300 million people worldwide [2–5]. Its prevalence has exponentially increased over the last few decades (by up to 150%) [2, 5]. Depression entails a condition of generalized inflammation and oxidative stress, which promotes the development and exacerbation of numerous physical and mental illnesses such as coronary heart disease, diabetes, and Alzheimer's disease [4, 6]. Depression is frequently undetected/untreated, especially in low- and middle-income countries [3, 7]. This is probably due to lack of mental health literacy (poor knowledge of a mental disorder and stigma), which discourages early detection of symptoms and engagement in culturally effective education and treatment [8]. Validated brief measures that reliably identify depression are essential to mitigate its adverse effects on health, quality of life, and productivity [5, 8].

Caring for children, especially those with intellectual disabilities, can be highly stressful [9]. Those children frequently exhibit poor sleep symptoms as well as disruptive behaviors (irritability and running out of the house at night) [10]. Therefore, the mothers are continually alert, and most of them sleep with their children in the same room or the same bed. Accordingly, they may spend a substantial

proportion of nighttime in caregiving [11]. Mothers may also be burdened with working outdoors to support their living, in addition to the demands of caring for their healthy children and meeting their social and family responsibilities, which put their psychological and physical wellbeing at jeopardy [12].

Sleep problems such as insomnia, difficulty initiating/maintaining sleep, are highly prevalent worldwide, affecting up to 22% of adults. Daytime impairment is reported in approximately 15% of those affected [13]. Nightmares are distressing mental imagery, which evoke nocturnal awakening, nonrestorative sleep, and sleep avoidance. They occur weekly in 2%–5% of adults, and they are highly associated with other sleep disturbances such as insomnia [12, 13] and poor sleep quality—low levels of self-perceived satisfaction with sleep. Caregiving women are at a great risk of poor sleep and nightmares secondary to their children's poor sleep, disruptive behaviors, and caregiving stress, particularly at nighttime [11, 12]. Longitudinal and experimental studies show that chronically poor sleep increases the risk of developing future depression [2, 14, 15]. It operates through a mechanism that impedes the removal of brain wastes, giving rise to neuronal damage [15]. Additionally, poor sleep is associated with adverse psychological effects such as cognitive decline and poor emotional regulation, which promote the development of depressive symptoms over time [16].

Poor sleep also increases vulnerability to poor physical health, which manifests by uncomfortable symptoms (e.g., fatigue, headache, gastrointestinal discomfort) and physical disorders (e.g., diabetes, hypertension) [11]. In line, poor physical health predisposes to defective coping and mental frailty [17]. On the other hand, depression promotes the development and deterioration of up to 29 physical morbidities (e.g., of the endocrine, musculoskeletal, and circulatory systems) by discouraging self-care, help-seeking, and adherence to prescribed treatments [4–18]. Indeed, mothers of children with intellectual disabilities and their children experience frequent hospitalization. During their children's admission (e.g., to critical care units), the mothers remain

actively involved in caregiving, often with limited perceived support from healthcare professionals, which negatively impacts the continuity and quality of their sleep [19].

Among various common measures, the Patient Health Questionnaire 9-Item scale (PHQ-9) tightly matches the DSM-IV diagnostic criteria for depression. Despite its relative brevity, it may not yield the required response when used frequently in some clinical settings because of time constraints [20]. Research indicates that two items from the Primary Care Evaluation of Mental Disorders (PRIME-MD) Screening Questionnaire, which address depressed mood and loss of interest, are used in general practice, and they may effectively capture depression. However, the binary nature (yes/no) of these items may underestimate the severity of depression or the extent of symptom change over time. Thanks to its four-point response format, the PHQ-2 may overcome these shortcomings [21]. Many studies testing the PHQ-2 diagnosed MDD through standard diagnostic methods such as the Composite International Diagnostic Interview (CIDI) [3], the Mini International Neuropsychiatric Interview (MINI) [20], and the Structured Clinical Interview for DSM-IV (SCID) [21, 22]. The performance of the PHQ-2 was also favorable relative to the PHQ-9 and other long measures of depression such as Beck Depression Inventory [20–23]. It successfully reflected response to treatment [20] as well as longitudinal changes in the course of depression, e.g., spontaneous remission, resistance/lack of change, or condition deterioration [21]. However, reports on its cutoff scores, which range from ≥ 2 to ≥ 3 , are inconsistent in terms of the accuracy of their diagnostic performance [23, 24].

The PHQ-9 and its short form PHQ-2 have been validated in up to 80 languages [3]. Until the current moment, psychometric testing is entirely lacking for the PHQ-2 in Arabic—a language spoken by around 500 million people in 22 countries [25, 26]. Moreover, economic and political instabilities in the Middle East and North Africa triggered the displacement of large numbers of Arabs into the USA, Australia, France, Germany, the UK, among others [27, 28]. Mental and physical health problems are documented among immigrants and refugees, stressing the need for well-tested Arabic measures [29]. Therefore, this study addresses a previously unexplored context by evaluating the psychometric properties of the Arabic version of the PHQ-2 as a brief measure for diagnosing depression. As per the literature, it hypothesized that the Arabic PHQ-2 would have a unidimensional structure, which would strongly correlate with measures of mood, stress, poor sleep, nightmares, and poor general physical health. The study included a measure of happiness to test the divergent validity of the PHQ-2. It also aimed to test its criterion and predictive validity by determining its cutoff in relevance to the aforementioned constructs.

2. Methods

2.1. Design and Participants. Using an anonymous online survey, this cross-sectional study recruited a convenience sample of 85 Saudi mothers of children with intellectual disabilities. Invitations to take part in the study were disseminated through parent groups on Facebook and

WhatsApp as well as advertisements posted in special childcare centers in the Kingdom of Saudi Arabia. Potential participants were informed about the commencement of a free psychological support intervention. Only mothers were encouraged to take part. Participants eligible for inclusion in the study were those aged 18 years or older who had children with intellectual disabilities, could speak Arabic, and owned a smartphone. They were asked to complete the survey even if they did not wish to take part in the supportive intervention. Before completing the survey, they signed a digital informed consent form, which reassured them about voluntary participation and confidentiality of their data. The study protocol was approved by the Research Ethics Committee of Ha'il University (11/9/2023: H-2023-367).

2.2. Measures. A self-administered questionnaire was used for data collection. It elicited information on the socio-demographic characteristics of the mothers (age, marital status, desire to take part in a support program). The questionnaire had many sections, which explored the clinical characteristics of the participants including the following measures:

The Patient Health Questionnaire-2 (PHQ-2) is a two-item version of the PHQ-9. Its items measure depressive symptoms over the last two weeks: “losing interest” and “feeling down, depressed, or hopeless.” Items are rated on a four-point response format ranging from 0 (*not at all*) to 3 (*almost every day*). The minimum and maximum total scale scores range from 0 to 6; higher scores reflect greater symptom severity [21].

Mood was assessed by a single item “Rate your mood on a scale from 0 (*very bad/the worst mood ever*) to 10 (*excellent/the best mood ever*).” Higher scores of this item reflect good mood and vice versa [30].

Happiness was assessed by a single item “Rate the extent to which you generally feel happy over the past week on a scale from 0 (*very unhappy*) to 10 (*extremely*).” Higher scores reflect greater happiness [31].

Stress was assessed by a single item “In the past year, how would you rate the amount of stress in your life (at home and at work/study)?” Mothers were requested to rate their experience on a scale ranging from 1 (*no stress*) to 6 (*extreme stress*) [32].

Sleep quality was assessed by a single-item measure: “Rate the quality of your sleep on a scale from 0 (*very poor*) to 10 (*excellent*)” [33].

Perceived physical health was assessed by a single item “How do describe your general physical health?” They rated their response on a five-point scale ranging from 1 (*poor*) to 5 (*excellent*) [34].

Mothers were asked to describe their experience of dreams during the past month by a single item (no dreams/neutral content and bad dreams/nightmares) [12]. Mothers' desire to participate in a supportive program was assessed by a single binary (yes, no) item.

2.3. Statistical Analysis. Frequencies and percentages were used to describe categorical variables. The distribution of quantitative variables was checked. Accordingly, normally

and nonnormally distributed variables were reported as mean \pm standard deviation and median (interquartile range: Q1–Q3), respectively.

The internal consistency of the PHQ-2 was measured by coefficient α , while item–total correlations were used to address its convergent validity at the item level. We calculated intraclass correlation coefficients (ICCs) with 95% confident intervals using mean-rating ($k = 85$), absolute-agreement, and two-way random-effects model [35, 36].

Confirmatory factor analysis (CFA) of scales that comprise fewer than four items results in noninformative saturated models, i.e., the values of chi-square (χ^2) and degree of freedom (DF) equal zero [7, 37]. As a remedy, the CFA model capturing the construct validity of the PHQ-2 was nested within a latent variable model (LVM) by anchoring it to the measures of mood, stress, and happiness, which were included as exogenous variables in the model. This approach also allows testing the convergent and divergent validity of the PHQ-2 [27]. Considering the model as well as fitting was based on the thresholds of < 3 , < 0.05 , and > 0.95 for the following fit indices: χ^2/DF (CMIN/DF), (root-mean-square error of approximation [RMSEA] and standardized root-mean-square residual [SRMR]), and (Comparative Fit Index [CFI] and Tucker–Lewis Index [TLI]), respectively [1, 25].

For predictive validity testing, receiver-operating characteristic (ROC) curve analysis was conducted using the PHQ-2 as a continuous variable to predict low mood, poor sleep quality, nightmares, poor physical health, stress, and willingness to join a supportive intervention. For this purpose, the categorical form of the continuous outcome variables was obtained by categorizing them according to their quartile scores (Table 1): low mood and poor sleep quality < 4 , poor physical health < 3 , and high stress ≥ 5 —this is because quartile scores may operate to a great extent like statistically determined cutoff scores [1, 37].

The diagnostic potential of the PHQ-2 was considered based on various fit indices of different ROC models: area under the curve (AUC), sensitivity and specificity for all possible cutoff points, and the Youden index (sensitivity + specificity) – 1 [1, 18]. In the next step, two categorical forms of the PHQ-2 were defined based on cutoff points resulting from the analysis (≥ 2.5 and ≥ 3.5). Then, χ^2 index was used to determine the positive predictive values (PPVs) and the negative predictive values (NPVs) of each threshold. The analyses were performed in SPSS Version 22 and Amos Version 26. Significance of the results was set at probability levels below 0.05 in two-tailed tests.

3. Results

Our mothers had a mean age of 40.8 ± 8.3 years. Most of the respondents were married (87.1%) and willing to receive psychological support (63.5%). More than one-fifth of the mothers (22.4%) frequently experienced nightmares. Table 1 displays the clinical characteristics of the mothers.

As shown in Table 2, the frequency of losing interest (Item 1) was greater than that of feeling blue, depressed, or hopeless (Item 2). Both items had strong significant positive

TABLE 1: Clinical characteristics of the participants ($N = 85$).

Characteristics	Mean \pm SD	MD (Q1–Q3)
PHQ-2	3.2 (1.1)	4 (3–5)
Mood	5.7 (2.9)	6 (4–8)
Sleep quality	5.6 (2.5)	5 (4–7)
Perceived physical health	4.0 (1.8)	3 (3–4)
Stress	3.5 (1.7)	3 (2–5)
Happiness	6.0 (2.9)	6 (5–8)

Note: MD: median and PHQ-2: Patient Health Questionnaire 2-item scale. Abbreviation: SD, standard deviation.

correlations with each other and with the total score of the PHQ-2, indicating scale homogeneity and convergent validity at the item level.

The LVM displayed excellent fit ($\chi^2 = 3.01$, $DF = 3$, $p = 0.390$, $CMIN/DF = 1.0$, $CFI = 0.999$, $TLI = 0.999$, $RMSEA = 0.007$, $RMSEA$ 95% CI: 0.000 to 0.184, $SRMR = 0.025$). It predicted 70% of the variance in depression. Figure 1 shows that the loadings of both items of the PHQ-2 were high, indicating strong correlation with their domain-specific factor, i.e., adequate construct validity. Convergent validity was expressed by a significant direct positive effect of stress on the PHQ-2. Because higher scores of the single item measuring mood reflect good mood, this item had a significant negative direct effect on the PHQ-2 (i.e., higher PHQ-2 scores reflect low mood), lending further support to its concurrent/convergent validity. Meanwhile, divergent validity was expressed by a significant direct negative effect of happiness on the PHQ-2.

The reliability of the PHQ-2 was adequate (coefficient $\alpha = 0.76$). The values of item–total correlations were also high ($r = 0.61$), indicating adequate convergent validity at the item level. The coefficient α of the PHQ-2 was comparable to the average measures ICC, indicating good scale reliability. In contrast, the single measures ICC was at the lower end of the acceptable range (0.60, 95% CI: 0.45–0.73), suggesting potential variability due to rater error.

The results of predictive/criterion validity as tested by the ROC analysis are shown in Table 3 and Figure 2. For all outcomes, the best balance between the highest values of both sensitivity and specificity was achieved at a PHQ-2 threshold of ≥ 2.5 . AUC values indicated good diagnostic accuracy of the PHQ-2 for mood and sleep quality. The diagnostic accuracy of the PHQ-2 was moderate for nightmares, general physical health, stress, and joining a supportive program. Sensitivity values were high for all outcomes, albeit specificity values were lower, particularly for joining a supportive program.

4. Discussion

Brief measures enable the efficient monitoring of psychiatric symptoms, which may be routinely performed to identify unmet treatment needs. This practice would enhance the quality and effectiveness of psychological interventions while offering essential evidence of service efficacy to insurance providers and funding agencies [20]. Research emphasizes that the performance of the PHQ-2 is favorable

TABLE 2: Item analysis and descriptive statistics of the Patient Health Questionnaire-2 (PHQ-2) among mothers of children with intellectual disabilities (N = 85).

Items	Mean (SD)	Median (Q1–Q3)	Skewness (SE)	Kurtosis (SE)	Shapiro–Wilk	Correlations	
						1	2
1. Little interest or pleasure in doing things	2.08 (1.04)	2 (1–3)	0.62 (0.26)	-0.76 (0.52)	0.83**	—	—
2. Feeling down, depressed, or hopeless	1.96 (0.97)	2 (1–2)	0.80 (0.26)	-0.29 (0.52)	0.82**	0.607**	—
3. Patient Health Questionnaire-2	4.05 (1.80)	4 (3–5)	0.81 (0.26)	-0.15 (0.52)	0.88**	0.904**	0.889**

**Correlations are significant at the level of 0.01.

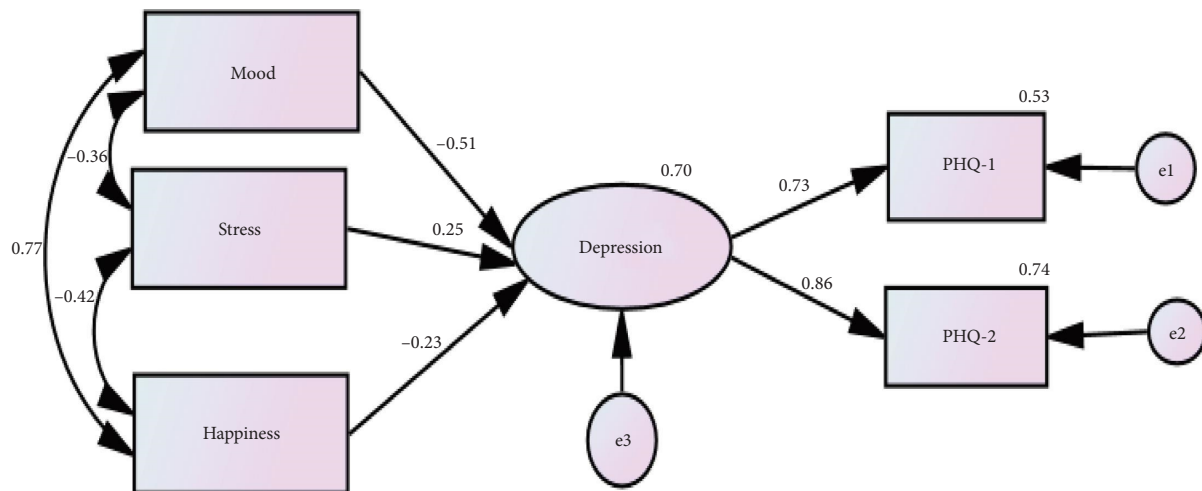


FIGURE 1: Latent variable model exploring the construct (convergent and divergent) validity of the Patient Health Questionnaire-2 (PHQ-2) among mothers of children with intellectual disabilities.

relative to standard diagnostic interviews and well-established measures of depression; it is also sensitive to change (e.g., in response to treatment or spontaneous remission) [3, 20–24, 38]. Filling a gap in the current literature, this study is the first to examine the psychometric properties of the PHQ-2 in an Arab culture and among mothers of children with disabilities. Our LVM confirms the unidimensional structure of the scale as well as its convergent and divergent validities. Examining its cutoff relative to six outcome criteria shows that the PHQ-2 at the cutoff point ≥ 2.5 achieved the best balance between the highest values of both sensitivity and specificity, albeit higher PPV was noticed at the cutoff ≥ 3.5 (Table 3).

Different values of the optimal cutoff score of the PHQ-2 have been reported in the literature: ≥ 1 among Spanish Speaking Immigrants in Chile [3], ≥ 2 among Spanish primary care patients and Iranian students [5, 22]; and ≥ 3 among Italian cardiovascular inpatients, German medical outpatients, and Australian recipients of internet-delivered cognitive behavior therapy for common mental disorders [20, 21, 39]. Such variations are congruent with notions that cutoff scores are not universal, and they should be determined for each region and for each disease condition [40]. In our sample, the PHQ-2 mean score (Table 1) was greater than previously recorded cutoffs for detecting clinically significant depression in people with physical and mental conditions, i.e., ≥ 3 cardiovascular disorders and common mental disorders [20, 21, 39]. Meanwhile, classifying the respondents based on the cutoffs detected in the present study indicates that 76.5% and 58.8% of our mothers had PHQ-2 scores ≥ 2.5 and ≥ 3.5 , respectively. Those scoring above the cutoffs are typically referred to the advanced diagnostics of depression. However, a cutoff of 2.5 may be less favorable in under-resourced settings due to the high cost of advanced diagnostic procedures. Nevertheless, it is crucial not to overlook any case of depression during screening, and a lower PHQ-2 cutoff (≥ 2) demonstrated effectiveness in achieving this goal [23, 24].

For all outcomes, the PHQ-2 at the threshold ≥ 3.5 still had acceptable sensitivity (Table 3). Particularly, sensitivity remained the same as that noticed at the threshold ≥ 2.5 for two cordial outcomes: low mood and poor sleep quality. However, at this threshold, specificity considerably decreased for the rest of the outcomes, ranging from 0.44 to 0.53. This was subsequently followed by reductions in the Youden index. According to the criteria of cutoff selection, which count on maximizing both sensitivity and specificity, this denotes reduction in the optimal diagnostic performance of the scale at a cutoff of 3.5. This result is consistent with that of a meta-analysis, which shows that 19 out of 21 studies report a PHQ-2 cutoff ≥ 3 with pooled sensitivity and specificity of 0.76 (95% CI = 0.68–0.82) and 0.87 (95% CI = 0.82–0.90), respectively [24]. However, it is not possible to concomitantly increase both sensitivity and specificity for increasing cutoff scores of a given test that has continuous (or multiple) results. A test that is more specific, i.e., with increased cutoff value and diagnostic potential for true positive cases (PPV), will be a less sensitive test, i.e., with a reduced capacity for identifying true negative cases [40]. In this respect, sensitivity and specificity of the PHQ-2 cutoff ≥ 2 pooled from 17 studies (0.91 (95% CI = 0.85–0.94) and 0.70 (95% CI = 0.64–0.76), respectively) were associated with considerably lower heterogeneity than the cutoff ≥ 3 ($I^2 = 43.2\%$ vs. 81.8%) [24]. However, in a previous study, PHQ-2 cutoffs that achieved the best balance between sensitivity (73%) and specificity (89%) at an AUC signifying excellent fit (0.85) exhibited exceptionally poor PPV (4%) despite the optimal NPV (100%) [3]. PPV is generally the metric of most utility in applied contexts, and low PPV indicates that the scale detects too many false positives [41]. Low mood is a main depressive symptom, and poor sleep is also a symptom that commonly manifests in depression and closely related sleep disorders [16, 21]. Thus, based on the greater benefit of the test “ruling-out the disease,” other indicators were considered in our trade-off for the higher cutoff: PPV and NPV were higher at the cutoff ≥ 3.5 for outcomes most relevant to MDD (low mood and poor sleep

TABLE 3: Cutoff scores of the Patient Health Questionnaire-2 (PHQ-2), along with goodness-of-fit indices associated with receiver-operating characteristic (ROC) curve analysis among mothers of children with intellectual disabilities (N = 85).

	AUC	SE	p	AUC 95% CI	Cutoff	Sensitivity	Specificity	Youden index	PPV %	NPV %	Cutoff	Sensitivity	Specificity	Youden index	PPV %	NPV %
Mood	0.84	0.05	0.001	0.75–0.94	2.5	0.96	0.69	0.65	35.4	95	3.5	0.96	0.44	0.40	46.0	97.1
Sleep quality	0.83	0.05	0.001	0.73–0.92	2.5	1.00	0.71	0.71	23.1	100	3.5	1.00	0.50	0.50	30.0	100
Nightmares	0.75	0.07	0.001	0.62–0.88	2.5	0.95	0.71	0.66	27.7	95.0	3.5	0.79	0.53	0.32	30.0	88.6
General physical health	0.72	0.07	0.004	0.59–0.85	2.5	0.95	0.71	0.66	27.7	95.0	3.5	0.79	0.53	0.32	30.0	88.6
Stress	0.73	0.07	0.001	0.60–0.85	2.5	0.89	0.71	0.60	35.4	85.0	3.5	0.73	0.53	0.26	38.0	80
Joining support program	0.74	0.05	0.001	0.63–0.84	2.5	0.85	0.61	0.46	70.8	60.0	3.5	0.72	0.36	0.08	78.8	57.1

Abbreviations: AUC, area under the curve; CI, confidence interval; NPVs, negative predictive values; PPVs, positive predictive values; SE, standard error.

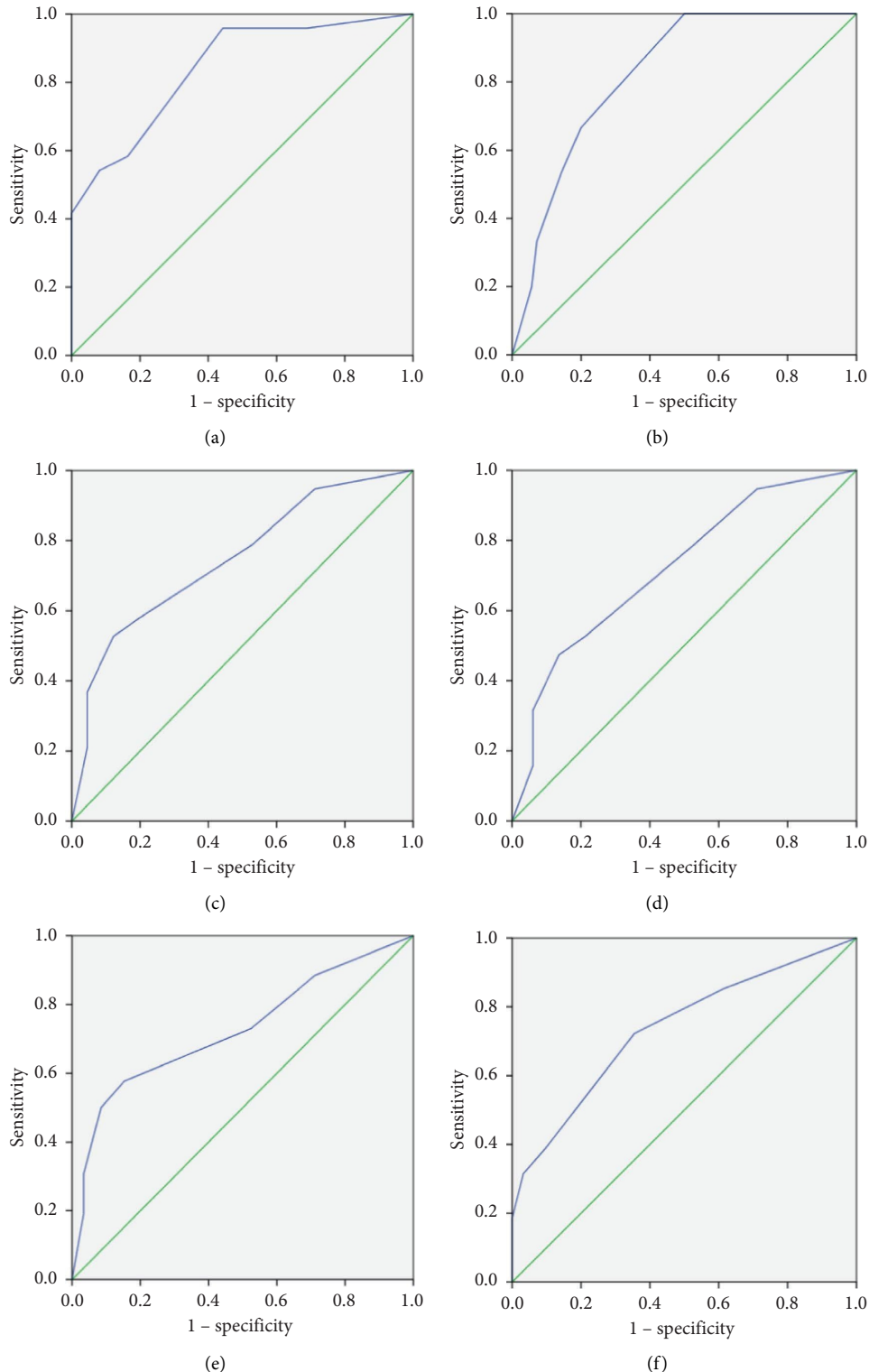


FIGURE 2: Receiver-operating characteristic (ROC) curve exploring the predictive/criterion validity of the Patient Health Questionnaire-2 (PHQ-2) among mothers of children with intellectual disabilities. (a) Mood. (b) Sleep quality. (c) Nightmares. (d) General physical health. (e) Stress. (f) Joining a supportive program.

quality)—sensitivity remained high (Table 3), ensuring that most cases of depression were accurately identified. High PPV and NPV signify that the scale does not detect too many false positive or false negative cases [42]. The PHQ-2 cutoff

score of ≥ 3.5 demonstrated greater specificity for these two outcome variables, increasing the likelihood of correctly identifying true cases of depression while minimizing false positives. This may reduce reliance on more costly

confirmatory diagnostic procedures. Thus, this cutoff would be more convenient in our Arab context, which is low in investments focused on diagnosing and treating mental diseases. In other words, it may not be possible/acceptable in low resource settings to consider lower cutoffs that address cases with slight severity [7, 43]. Additionally, research associates intense depressive symptoms with low sleep quality, which frequently follows nightmares and ruminations over caregiving-related stressors [12]. Congruent with the literature, PPV of the PHQ-2 at the cutoff ≥ 3.5 for all the outcomes was higher than that of the cutoff ≥ 2.5 , with slight reductions in NPV (Table 3), suggesting that this score is still capable of identifying those who frequently experience nightmares, poor physical health, or high stress.

The PHQ-2 mean score was also higher among those who intended to participate in the psychological support program (Mann–Whitney $U = 442.5$, $z = -3.69$, $p < 0.001$). However, mothers' willingness to take part in the support program was slightly higher among those who scored ≥ 2.5 than those who scored ≥ 3.5 on the PHQ-2 (54.1% vs. 45.9%). In line, those with higher scores of depression/anxiety were less likely to seek help (e.g., for pain treatment) because they may consider treatment as ineffective [18]. Severe depression is associated with greater alterations in mental processes involving poor self-efficacy (i.e., that they can seek help), focusing on negative events in past treatment experiences (e.g., drug resistance), lack of expression of negative emotions, lack of awareness of suicidal behaviors, and possibly greater stigma [44–46]. Therefore, those scoring ≥ 3.5 on the PHQ-2 may benefit from educational interventions that focus on depression literacy and help seeking [45].

4.1. Strength, Implications, and Limitations. This study, which adheres to STROBE reporting guidelines, has the merit of addressing a previously unexplored research area: the psychometric performance of PHQ-2 in Arabic and in a highly distressed sample—caregiving mothers of children with intellectual disabilities. It employed methodologically robust statistical approaches (LVM and ROC analysis) for evaluating the construct, convergent, and divergent validities of the scale and determining its cutoff. The results confirm the usability of the PHQ-2 for predicting low mood as well as distressful psychological problems such as poor sleep quality, nightmares, stress, and poor physical health. Accordingly, mothers scoring ≥ 2.5 should undergo advanced assessment for mental morbidities, while those scoring ≥ 3.5 should be given a higher priority in resource-limited settings. Because the PHQ-2 is ultra-brief, it may be combined with other brief measures to evaluate a wide range of outcomes that may be crucial for the assessment of depression and its management in routine care and research.

The study comes with some limitations, which should be considered while interpreting the results as they may limit their generalizability. While former studies [3, 20–23] compared the properties of the PHQ-2 relative to the full-length scale, other lengthy measures of depression, or diagnostic interviews (e.g., CIDI, MINI, SCID), a major

limitation in our study is that no gold-standard diagnostic tool was used to confirm depression diagnosis—because of access and logistic constraints, it was not possible to have a clinician-based diagnosis for concurrent validity testing, especially as the mothers did not attend mental or public health settings. Alternatively, we counted on a single-item measure of mood. Brief measures of all other outcome measures were used to overcome poor response rates, which are evident in surveys in the Arab region [28]—particularly, mothers are busy spending a long time in childcare and may be reluctant to contribute because they do not confide in surveys or perceive them as personally useful [47]. It is true that single-item measures cannot undergo reliability or construct validity testing, but these measures may capture global or unidimensional constructs at predictive validity equal to specific multiple-item measures [48, 49]. In agreement with this logic, cutoffs revealed by our analysis are comparable to those previously reported in studies validating the PHQ-2 against standardized clinical interviews [23, 24], which casts credibility on cutoffs estimated in the current analysis. Nevertheless, it is necessary that future studies overcome this major limitation by clinically validating the diagnosis of depression.

The study is also amenable to the risk of selection bias for many reasons. The sample size was relatively small ($N = 85$). However, according to the role of thumb, which requires 10 responses per item for testing the structure of a tool [50], this sample may credibly support the testing of our model, which comprised eight items in total. On the other hand, the characteristics of the participants may diverge from the general mother population as we included only those who can use smart phones to access the internet (educational level, seeking information, and support for self-care/childcare). Despite disseminating the link of the survey through Facebook groups of the parents of special need children in other countries (Jordan, Egypt, Kuwait, Qatar, Morocco, and Yemen), we failed to obtain responses from any of those countries. Owing to the help of childcare centers that were personally contacted by the second author, the sample came only from Saudi Arabia. Saudi Arabia is a large country with a great heterogeneity in socioeconomic standards, urbanization, educational (e.g., special childcare) and healthcare services, customs, support systems, among others [51]. Therefore, our nonrandom sample may not even represent Saudi and Arab mothers of children with intellectual disabilities. Future studies that overcome these limitations will provide credible information on the psychometrics of the PHQ-2.

5. Conclusions

The Arabic PHQ-2 is a reliable unidimensional measure, which demonstrates adequate convergent and divergent validities. The best balance between sensitivity and specificity in the ROC analysis revealed two thresholds of the PHQ-2 (≥ 2.5 and ≥ 3.5). The latter demonstrated considerable reductions in specificity for different outcomes except for two outcomes that are tightly related to MDD (low mood and poor sleep quality). Nonetheless, it had slightly higher

PPV for all outcomes than the cutoff ≥ 2.5 . Accordingly, the cutoff ≥ 2.5 may be used in depression screening to avoid missed diagnosis, while PPV denotes that the cutoff ≥ 3.5 may be more accurate for diagnosing severe form/clinically significant depression as it manifests primarily with low mood as well as poor sleep quality. Further investigations of the PHQ-2 in relevance to lengthy measures of depression or standardized diagnostic interviews in larger samples from other Arab countries may support these conclusions.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Ethics Statement

The data collection protocol was approved by the Research Ethics Committee of Ha'il University (11/9/2023: H-2023-367), and all the participants signed a digital informed consent before taking part in the study. The present study was conducted according to the Declaration of Helsinki.

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions

Conceptualization, Amira Mohammed Ali, Annamaria Pakai, and Feten Fekih-Romdhane; data curation, Saeed A. Al-Dossary and Feten Fekih-Romdhane; formal analysis, Feten Fekih-Romdhane and Musheer A. Aljaberi; funding acquisition, Saeed A. Al-Dossary; investigation, Saeed A. Al-Dossary; methodology, Amira Mohammed Ali, Heba Emad El-Gazar, and Mohamed Ali Zoromba; project administration, Amira Mohammed Ali and Saeed A. Al-Dossary; resources, Haitham Khatatbeh and Rasmieh Alamer; software, Haitham Khatatbeh; supervision, Amira Mohammed Ali; validation, Rasmieh Alamer, Carlos Laranjeira, and Faten Amer; visualization, Carlos Laranjeira; writing – original draft, Carlos Laranjeira, Haitham Khatatbeh, Musheer A. Aljaberi, and Mohamed Ali Zoromba; writing – review and editing, Faten Amer, Heba Emad El-Gazar, and Annamaria Pakai.

Funding

The research has been funded by the Scientific Research Deanship at University of Ha'il, Saudi Arabia through project number RG-23 195.

Acknowledgments

The research has been funded by the Scientific Research Deanship at the University of Ha'il, Saudi Arabia, through project number <<RG-23 195>>.

References

- [1] A. M. Ali, S. A. Al-Dossary, C. Laranjeira, et al., "Effects of Hormonal Replacement Therapy and Mindfulness-Based Stress Reduction on Climacteric Symptoms Following Risk-Reducing Salpingo-Oophorectomy," *Healthcare* 12, no. 16 (2024): 1612, <https://doi.org/10.3390/healthcare12161612>.
- [2] Y. J. Um, Y. Kim, Y. Chang, et al., "Association of Changes in Sleep Duration and Quality With Incidence of Depression: A Cohort Study," *Journal of Affective Disorders* 328 (2023): 64–71, <https://doi.org/10.1016/j.jad.2023.02.031>.
- [3] A. Errazuriz, R. Beltrán, R. Torres, and A. Passi-Solar, "The Validity and Reliability of the PHQ-9 and PHQ-2 on Screening for Major Depression in Spanish Speaking Immigrants in Chile: A Cross-Sectional Study," *International Journal of Environmental Research and Public Health* 19, no. 21 (2022): 13975, <https://doi.org/10.3390/ijerph192113975>.
- [4] P. Frank, G. D. Batty, J. Pentti, et al., "Association Between Depression and Physical Conditions Requiring Hospitalization," *JAMA Psychiatry* 80, no. 7 (2023): 690–699, <https://doi.org/10.1001/jamapsychiatry.2023.0777>.
- [5] I. Gómez-Gómez, I. Benítez, J. Bellón, et al., "Utility of PHQ-2, PHQ-8 and PHQ-9 for Detecting Major Depression in Primary Health Care: A Validation Study in Spain," *Psychological Medicine* 53, no. 12 (2022): 5625–5635, <https://doi.org/10.1017/S0033291722002835>.
- [6] A. M. Ali and H. Kunugi, "Bee Honey Protects Astrocytes against Oxidative Stress: A Preliminary In Vitro Investigation," *Neuropsychopharmacology Reports* 39, no. 4 (2019): 312–314, <https://doi.org/10.1002/npr2.12079>.
- [7] A. M. Ali, S. A. Al-Dossary, F. Fekih-Romdhane, et al., "Psychometric Evaluation of the Arabic Version of the Eight-Item Center for Epidemiological Studies Depression Scale (CESD-8): Specific Cultural Considerations for the Assessment of Depression," *International Journal of Nursing Studies Advances* 8 (2025): 100310, <https://doi.org/10.1016/j.ijnsa.2025.100310>.
- [8] V. Lopez, K. Sanchez, M. O. Killian, and B. H. Eghaneyan, "Depression Screening and Education: An Examination of Mental Health Literacy and Stigma in a Sample of Hispanic Women," *BMC Public Health* 18, no. 1 (2018): 646, <https://doi.org/10.1186/s12889-018-5516-4>.
- [9] H. Khatatbeh, T. Al-Dwaikat, H. Alfatafta, A. M. Ali, and A. Pakai, "Burnout, Quality of Life and Perceived Patient Adverse Events Among Paediatric Nurses During the COVID-19 Pandemic," *Journal of Clinical Nursing* 32, no. 13-14 (2022): 3874–3886, <https://doi.org/10.1111/jocn.16540>.
- [10] S. Alarifi, L. Denne, N. Alatifi, and R. P. Hastings, "Interventions Addressing Challenging Behaviours in Arab Children and Adults With Intellectual Disabilities and/or Autism: A Systematic Review," *Journal of Applied Research in Intellectual Disabilities: JARID* 37, no. 2 (2024): e13205, <https://doi.org/10.1111/jar.13205>.
- [11] M. F. Almeida, S. Mello, M. B. Zonta, and A. C. Crippa, "Cerebral Palsy and Sleep: Nonpharmacological Treatment and Impact on the Life of Caregivers: An Integrative Review," *Arquivos de Neuro-Psiquiatria* 82, no. 03 (2024): 001–009, <https://doi.org/10.1055/s-0044-1781464>.
- [12] A. N. Leggett, M. Morley, and S. F. Smagula, "It's Been a Hard Day's Night: Sleep Problems in Caregivers for Older Adults: Sleep Problems in Caregivers for Older Adults," *Current Sleep Medicine Reports* 6, no. 1 (2020): 1–10, <https://doi.org/10.1007/s40675-020-00164-0>.
- [13] J.-P. Delage, J. Côté, W.-G. Journault, A. Lemyre, and C. H. Bastien, "The Relationships Between Insomnia, Nightmares,

- and Dreams: A Systematic Review,” *Sleep Medicine Reviews* 75 (2024): 101931, <https://doi.org/10.1016/j.smrv.2024.101931>.
- [14] T. Fu, C. Wang, J. Yan, Q. Zeng, and C. Ma, “Relationship Between Antenatal Sleep Quality and Depression in Perinatal Women: A Comprehensive Meta-Analysis of Observational Studies,” *Journal of Affective Disorders* 327 (2023): 38–45, <https://doi.org/10.1016/j.jad.2023.01.125>.
- [15] O. Semyachkina-Glushkovskaya, I. Fedosov, T. Penzel, et al., “Brain Waste Removal System and Sleep: Photobiomodulation as an Innovative Strategy for Night Therapy of Brain Diseases,” *International Journal of Molecular Sciences* 24, no. 4 (2023): 3221, <https://doi.org/10.3390/ijms24043221>.
- [16] K. O’Leary, L. M. Bylisma, and J. Rottenberg, “Why Might Poor Sleep Quality Lead to Depression? A Role for Emotion Regulation,” *Cognition and Emotion* 31, no. 8 (2017): 1698–1706, <https://doi.org/10.1080/02699931.2016.1247035>.
- [17] A. M. Ali, A. A. Alkhamees, E. S. Abd Elhay, S. M. Taha, and A. O. Hendawy, “COVID-19-Related Psychological Trauma and Psychological Distress Among Community-Dwelling Psychiatric Patients: People Struck by Depression and Sleep Disorders Endure the Greatest Burden,” *Frontiers in Public Health* 9 (2022): 799812, <https://doi.org/10.3389/fpubh.2021.799812>.
- [18] A. M. Ali, S. A. Al-Dossary, A. M. Almarwani, M. Atout, R. Al-Amer, and A. A. Alkhamees, “The Impact of Event Scale-Revised: Examining its Cutoff Scores Among Arab Psychiatric Patients and Healthy Adults Within the Context of COVID-19 as a Collective Traumatic Event,” *Healthcare* 11, no. 6 (2023): 892, <https://doi.org/10.3390/healthcare11060892>.
- [19] R. J. Graham, D. M. Pemstein, and M. A. Curley, “Experiencing the Pediatric Intensive Care Unit: Perspective From Parents of Children With Severe Antecedent Disabilities,” *Critical Care Medicine* 37, no. 6 (2009): 2064–2070, <https://doi.org/10.1097/CCM.0b013e3181a00578>.
- [20] L. G. Staples, B. F. Dear, M. Gandy, et al., “Psychometric Properties and Clinical Utility of Brief Measures of Depression, Anxiety, and General Distress: The PHQ-2, GAD-2, and K-6,” *General Hospital Psychiatry* 56 (2019): 13–18, <https://doi.org/10.1016/j.genhosppsy.2018.11.003>.
- [21] B. Löwe, K. Kroenke, and K. Gräfe, “Detecting and Monitoring Depression With a Two-Item Questionnaire (PHQ-2),” *Journal of Psychosomatic Research* 58, no. 2 (2005): 163–171, <https://doi.org/10.1016/j.jpsychores.2004.09.006>.
- [22] M. Ghazisaeedi, H. Mahmoodi, I. Arpaci, S. Mehrdar, and S. Barzegari, “Validity, Reliability, and Optimal Cut-Off Scores of the WHO-5, PHQ-9, and PHQ-2 to Screen Depression Among University Students in Iran,” *International Journal of Mental Health and Addiction* 20, no. 3 (2022): 1824–1833, <https://doi.org/10.1007/s11469-021-00483-5>.
- [23] B. Levis, Y. Sun, C. He, et al., “Accuracy of the PHQ-2 Alone and in Combination With the PHQ-9 for Screening to Detect Major Depression: Systematic Review and Meta-Analysis,” *JAMA* 323, no. 22 (2020): 2290–2300, <https://doi.org/10.1001/jama.2020.6504>.
- [24] L. Manea, S. Gilbody, C. Hewitt, et al., “Identifying Depression With the PHQ-2: A Diagnostic Meta-Analysis,” *Journal of Affective Disorders* 203 (2016): 382–395, <https://doi.org/10.1016/j.jad.2016.06.003>.
- [25] A. M. Ali, A. A. Alkhamees, H. Hori, Y. Kim, and H. Kunugi, “The Depression Anxiety Stress Scale 21: Development and Validation of the Depression Anxiety Stress Scale 8-Item in Psychiatric Patients and the General Public for Easier Mental Health Measurement in a Post-COVID-19 World,” *International Journal of Environmental Research and Public Health* 18, no. 19 (2021): 10142, <https://doi.org/10.3390/ijerph181910142>.
- [26] A. Sabah, M. A. Aljaberi, J. Hajji, et al., “Psychometric Properties of the Benign and Malicious Envy Scale: A Cross-Cultural Study in the Arab Maghreb Countries,” *Acta Psychologica* 248 (2024): 104397, <https://doi.org/10.1016/j.actpsy.2024.104397>.
- [27] A. M. Ali, S. Al-Dossary, C. Laranjeira, et al., “Loneliness Among Dementia Caregivers: Evaluation of the Psychometric Properties and Cutoff Score of the Three-Item UCLA Loneliness Scale,” *Frontiers in Psychiatry* 16 (2025): 1526569, <https://doi.org/10.3389/fpsy.2025.1526569>.
- [28] S. Elshahat and K. B. Newbold, “Physical Activity Participation Among Arab Immigrants and Refugees in Western Societies: A Scoping Review,” *Preventive Medicine Reports* 22 (2021): 101365, <https://doi.org/10.1016/j.pmedr.2021.101365>.
- [29] P. Zeinoun, D. Iliescu, and R. El Hakim, “Psychological Tests in Arabic: A Review of Methodological Practices and Recommendations for Future Use,” *Neuropsychology Review* 32, no. 1 (2022): 1–19, <https://doi.org/10.1007/s11065-021-09476-6>.
- [30] P. Gertler and R. L. Tate, “Are Single Item Mood Scales (SIMS) Valid for People With Traumatic Brain Injury?” *Brain Injury* 34, no. 5 (2020): 653–664, <https://doi.org/10.1080/02699052.2020.1733087>.
- [31] M. W. Fordyce, “A Review of Research on the Happiness Measures: A Sixty Second Index of Happiness and Mental Health,” *Social Indicators Research* 20, no. 4 (1988): 355–381, <https://doi.org/10.1007/BF00302333>.
- [32] A. J. Littman, E. White, J. A. Satia, D. J. Bowen, and A. R. Kristal, “Reliability and Validity of 2 Single-Item Measures of Psychosocial Stress,” *Epidemiology* 17, no. 4 (2006): 398–403, <https://doi.org/10.1097/01.ede.0000219721.89552.51>.
- [33] E. Snyder, B. Cai, C. DeMuro, M. F. Morrison, and W. Ball, “A New Single-Item Sleep Quality Scale: Results of Psychometric Evaluation in Patients With Chronic Primary Insomnia and Depression,” *Journal of Clinical Sleep Medicine* 14, no. 11 (2018): 1849–1857, <https://doi.org/10.5664/jcsm.7478>.
- [34] C. Agyemang, S. Denktas, M. Bruijnzeels, and M. Foets, “Validity of the Single-Item Question on Self-Rated Health Status in First Generation Turkish and Moroccans Versus Native Dutch in the Netherlands,” *Public Health* 120, no. 6 (2006): 543–550, <https://doi.org/10.1016/j.puhe.2006.03.002>.
- [35] A. M. Ali, R. A. Alameri, T. Brooks, et al., “Cut-Off Scores of the Depression Anxiety Stress Scale-8: Implications for Improving the Management of Chronic Pain,” *Journal of Clinical Nursing* 32, no. 23–24 (2023): 8054–8062, <https://doi.org/10.1111/jocn.16878>.
- [36] T. K. Koo and M. Y. Li, “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research,” *Journal of Chiropractic Medicine* 15, no. 2 (2016): 155–163, <https://doi.org/10.1016/j.jcm.2016.02.012>.
- [37] A. M. Ali, S. A. Al-Dossary, M. A. Aljaberi, et al., “The Arabic Satisfaction With Life Scale (SWLS) and Its Three-Item Version: Factor Structure and Measurement Invariance Among University Students,” *Acta Psychologica* 255, no. 255 (2025): 104867, <https://doi.org/10.1016/j.actpsy.2025.104867>.
- [38] J. I. Hlynsson and P. Carlbring, “Diagnostic Accuracy and Clinical Utility of the PHQ-2 and GAD-2: A Comparison With Long-Format Measures for Depression and Anxiety,” *Frontiers in Psychology* 15 (2024): 1259997, <https://doi.org/10.3389/fpsyg.2024.1259997>.
- [39] M. Giuliani, A. Gorini, S. Barbieri, F. Veglia, and E. Tremoli, “Examination of the Best Cut-Off Points of PHQ-2 and GAD-2 for Detecting Depression and Anxiety in Italian Cardiovascular Inpatients,” *Psychology and Health* 36, no. 9 (2021): 1088–1101, <https://doi.org/10.1080/08870446.2020.1830093>.

- [40] F. Habibzadeh, P. Habibzadeh, and M. Yadollahie, "On Determining the Most Appropriate Test Cut-Off Value: the Case of Tests With Continuous Results," *Biochemia Medica* 26, no. 3 (2016): 297–307, <https://doi.org/10.11613/bm.2016.034>.
- [41] E. G. Chin, E. M. Buchanan, C. Ebesutani, and J. Young, "Depression, Anxiety, and Stress: How Should Clinicians Interpret the Total and Subscale Scores of the 21-Item Depression, Anxiety, and Stress Scales?" *Psychological Reports* 122, no. 4 (2019): 1550–1575, <https://doi.org/10.1177/0033294118783508>.
- [42] J.-K. Ahn, Y. Kim, and K.-H. Choi, "The Psychometric Properties and Clinical Utility of the Korean Version of GAD-7 and GAD-2," *Frontiers in Psychiatry* 10 (2019): 127, <https://doi.org/10.3389/fpsy.2019.00127>.
- [43] A. M. Ali, S. A. Al-Dossary, C. Laranjeira, et al., "Cardiometabolic Morbidity (Obesity and Hypertension) in PTSD: A Preliminary Investigation of the Validity of Two Structures of the Impact of Event Scale-Revised," *Journal of Clinical Medicine* 13, no. 20 (2024): 6045, <https://doi.org/10.3390/jcm13206045>.
- [44] A. M. Ali, A. A. Alkhamees, S. Hallit, T. N. Al-Dwaikat, H. Khatatbeh, and S. A. Al-Dossary, "The Depression Anxiety Stress Scale 8: Investigating its Cutoff Scores in Relevance to Loneliness and Burnout Among Dementia Family Caregivers," *Scientific Reports* 14, no. 1 (2024): 13075, <https://doi.org/10.1038/s41598-024-60127-1>.
- [45] T. McLaren, L.-J. Peter, S. Tomczyk, H. Muehlan, G. Schomerus, and S. Schmidt, "The Seeking Mental Health Care Model: Prediction of Help-Seeking for Depressive Symptoms by Stigma and Mental Illness Representations," *BMC Public Health* 23, no. 1 (2023): 69, <https://doi.org/10.1186/s12889-022-14937-5>.
- [46] P. Shi, A. Yang, Q. Zhao, Z. Chen, X. Ren, and Q. Dai, "A Hypothesis of Gender Differences in Self-Reporting Symptom of Depression: Implications to Solve Under-Diagnosis and Under-Treatment of Depression in Males," *Frontiers in Psychiatry* 12 (2021): 589687, <https://doi.org/10.3389/fpsy.2021.589687>.
- [47] J. J. Gengler, M. Tessler, R. Lucas, and J. Forney, "Why Do You Ask? the Nature and Impacts of Attitudes Towards Public Opinion Surveys in the Arab World," *British Journal of Political Science* 51, no. 1 (2021): 115–136, <https://doi.org/10.1017/S0007123419000206>.
- [48] B. B. Hoepfner, J. F. Kelly, K. A. Urbanoski, and V. Slaymaker, "Comparative Utility of a Single-Item Versus Multiple-Item Measure of Self-Efficacy in Predicting Relapse Among Young Adults," *Journal of Substance Abuse Treatment* 41, no. 3 (2011): 305–312, <https://doi.org/10.1016/j.jsat.2011.04.005>.
- [49] J. Song, E. Howe, J. R. Oltmanns, and A. J. Fisher, "Examining the Concurrent and Predictive Validity of Single Items in Ecological Momentary Assessments," *Assessment* 30, no. 5 (2023): 1662–1671, <https://doi.org/10.1177/10731911221113563>.
- [50] G. O. Boateng, T. B. Neilands, E. A. Frongillo, H. R. Melgar-Quinonez, and S. L. Young, "Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer," *Frontiers in Public Health* 6 (2018): 149, <https://doi.org/10.3389/fpubh.2018.00149>.
- [51] A. Abdul Salam, I. Elsegaey, R. Khraif, and A. Al-Mutairi, "Population Distribution and Household Conditions in Saudi Arabia: Reflections From the 2010 Census," *SpringerPlus* 3, no. 1 (2014): 530, <https://doi.org/10.1186/2193-1801-3-530>.