The Patient Health Questionnaire-4 (PHQ-4) as a Brief Screening Tool for Anxiety and Depression: A Study of Iranian Samples

Mahboubeh Dadfar[1] Yahya Turan[2] Arash Askari[3] Mohsen Roshanpajouh[4]

Abstract

The aims of the present study were (a) to develop a Farsi version of the Patient Health Questionnaire-4 (PHQ-4), (b) to explore correlations for anxiety and depression with psychological well-being, mental health, and happiness, and (c) to explore gender-related differences. A sample of 488 subjects representing the general population of Iran responded to the PHQ-4, the World Health Organization-five Well-Being Index, the Self-Rating Scale of Mental Health, and the Self-Rating Scale of Happiness. Cronbach's α was .86 for the PHQ-4. Using a principal components analysis (PCA) with varimax rotation, one factor was extracted, labelled "Anxious/Depressed." Using exploratory factor analysis, the scale has both a one-factor and a two-factor structure. Confirmatory factor analysis showed the PHQ-4 has good fit indices for both structures. Cronbach's α was high for both structures. The PHQ-4 scores correlated negatively with psychological well-being, mental health, and happiness scores, indicating good validity. The sex-related difference in the PHQ-4 was not statistically significant. The PHQ-4 had good psychometric properties in the present general population sample from Iran. This study provides evidence for the usefulness of the Farsi version of the PHQ-4 for assessing the prevalence of anxiety and depression symptoms in Iranian community residents.

Keywords: Anxiety, Depression, Psychological well-being, Mental health, Happiness, Iran

1. Introduction

Mental health, as defined in 2022, is a state of mental well-being that allows individuals to cope with the stressors of life, recognize their abilities and competencies, learn effectively, work well, and contribute to society (Mental health, 2022). Mental health and mental disorders are generally determined by a combination of social, psychological, and biological factors that interact with each other (Fryers et al., 2005; Mental health, 2022; Sıncar et al., 2020). Mental health can be influenced by various individual and societal factors such as economic disadvantage, employment, working conditions, housing conditions, and (lack of) social support (Dohrenwend, 1990; Fryers et al., 2005).

[1] *Iran University of Medical Sciences, Tehran, Iran*

[2] Bandırma Onyedi Eylül University, Bandırma, Turkey

[3] Corresponding author: Arash Askari, Department of Psychiatry, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran; email: askaria@mums.ac.ir or arash7askari@gmail.com

[4] *Iran University of Medical Sciences, Tehran, Iran*

Globally, a significant portion of the burden of global health is attributed to mental disorders. Among mental disorders, depression and anxiety are reported as the most frequent disabling mental health conditions (Dohrenwend, 1990). In 2019, depression and anxiety ranked among the top 25 causes of disability for all age groups (Institute for Health Metrics and Evaluation [IHME], 2024; Murray, 2020; Santomauro et al., 2021; WHO, 2023).

In the DSM-5, depressive disorders are defined as the presence of a sad, empty, or restless mood accompanied by somatic and cognitive changes that significantly affect an individual's functioning (Black & Grant, 2014, p. 155). Depression is often characterized by the presence of negative thoughts about oneself, the world, and the future, often accompanied by feelings of hopelessness, guilt, or worthlessness (Beck & Alford, 2009, p. 65). Depressive mood is influenced by biological, social, and emotional changes (Haehner et al., 2024). Depression is a common mental disorder. An estimated 3.8% of the population experience it, including 5% of adults (4% in men and 6% in women) and 5.7% of adults aged 60 and older (Depressive Disorder (Depression), 2023).

Depression is a challenging condition to diagnose because it can manifest with a wide range of symptoms, including symptoms such as headaches, constipation, loss of appetite, back pain, or chronic fatigue. Rapid diagnosis and treatment of depression are crucial. Research has shown that many individuals who die by suicide had recently visited a primary care physician, but they were not diagnosed with depression and antidepressant treatment was not initiated for them (Beck & Beck, 1972).

The General Health Services Study of Psychological Problems (PPGHC), conducted by the World Health Organization (WHO) in various cultures to detect common psychological disorders in primary care settings, reports that almost half of depression and anxiety cases occur in the same individuals at the same time (Sartorius et al., 1996). Although it has been suggested that these two disorders can be separated (Wetzler & Katz, 1989), research has shown that a significant portion of individuals with Generalized Anxiety Disorder (GAD) also meet the criteria for depression, and vice versa (Sartorius et al., 1996). It is reported that up to 85% of depression patients also experience significant anxiety symptoms, and up to 90% of individuals with anxiety disorders develop depression (Gorman, 1996). The symptoms of depression and anxiety disorders seen in patients can meet the criteria for both mental disorders, making it challenging to differentiate between the two. However, it is essential to identify and treat both conditions because they are associated with significant morbidity and mortality. In this context, primary care health facilities play an important role in maintaining mental health, detecting patients with these disorders, and providing treatment (Tiller, 2013).

Anxiety is defined as a pervasive negative mood that affects emotional, cognitive, physical, behavioral, and relational states in various ways. In clinical literature, this term is applied to the presence of fear or anxiety that is disproportionate to the situation (Black & Grant, 2014, p. 123). GAD, the most commonly encountered mental disorder in the community, has a prevalence ranging from 2.8% to 8.5% in general medical practice and from 1.6% to 5.0% in the general population- although the diagnosed cases are merely the extreme end of a bell-shaped distribution, and any categorical distinction between “normal” and GAD is arbitrary. Until recently, due to its inability to be fully diagnosed and the scarcity of empirical studies in this field, GAD was considered relatively unimportant both by clinicians and the public (Judd et al., 1998; Spitzer et al., 2006).

Various tools have been developed to diagnose depression (e.g., Beck, 1961; EuroQol Group, 1990; Kroenke et al., 2001; Kroenke, Strine, et al., 2009; Radloff, 1977; Rush et al., 2003). In today's societies, there is a need for reliable diagnostic tools that can be used for rapid screening in primary care health facilities as common mental disorders like depression become increasingly prevalent (Kroenke et al., 2003). This is

the purpose for which the PHQ-4 was developed (Löwe et al., 2010). Corson et al. (2004) reported that ultra-short tests consisting of two or three items identify approximately 80% of depression cases in depression screenings.

The first two items of the PHQ-4 assess depression and the other two items assess anxiety, making it a useful tool for detecting both common mental disorders. The 2-item Patient Health Questionnaire for Depression (PHQ-2) is a shorter version of the 9-item PHQ-9. PHQ-2 focuses solely on depressive mood and loss of interest, thus representing the basic diagnostic criteria of the DSM-IV (Kroenke et al., 2003; Löwe et al., 2005). The results of a prospective criterion standard study conducted on a sample of 520 outpatients indicate that PHQ-2 has good criterion validity, convergent validity, and sensitivity to change (Löwe et al., 2005). In measurements using PHQ-2, it was found that a cutoff of 3 or higher is comparable to the PHQ-9 diagnostic algorithm for any depressive disorder and major depressive disorder (Kroenke et al., 2003).

The 2-item Generalized Anxiety Disorder Scale (GAD-2), which measures anxiety, is a shorter version of the 7-item Generalized Anxiety Disorder Scale (GAD-7) (Kroenke et al., 2007; Löwe et al., 2008; Spitzer et al., 2006). A validation study conducted with a sample of 965 primary care patients who were diagnosed through criterion standard interviews for the most common four anxiety disorders found that GAD-2 has good criterion validity (Löwe et al., 2010).

The findings obtained from a study with over 5,000 subjects, conducted to determine the validity and reliability of PHQ-4, show that this extremely short 4-item measurement can reliably and validly measure depression and anxiety in the general population (Löwe et al., 2010).

Kroenke, Spitzer, et al. (2009) have classified PHQ-4 scores operationally as normal (0-2), mild (3-5), moderate (6-8), and severe (9-12). During the standardization process of the PHQ-4's two components, PHQ-2 and GAD-2, a cutoff of 3 was found to be the most appropriate (Kroenke et al., 2003; Löwe et al., 2005). The PHQ-4 serves as a rapid, valid, and reliable screening tool rather than a diagnostic tool (Kroenke, Spitzer, et al., 2009; Mitchell & Coyne, 2007).

The aims of the present study were (a) to develop a Farsi version of the Patient Health Questionnaire-4 (PHQ-4), (b) to explore correlations for anxiety and depression with psychological well-being, mental health, and happiness, and (c) to explore gender-related differences.

2. Methods

2.1. Participants

Using a cross-sectional study, a sample of 488 subjects from the Iranian general population (M_{age} = 29.5, *SD* = 11.5; 73.4% female) participated in the study. The participants were recruited via a self-selection method. They were provided with information regarding the nature and aim of the study, including the number and type of questions/content of the research. In the online questionnaire forms, the first question was "According to the above explanation, do you agree to participate in the study?". The scales completed by all the subjects who responded "yes" to this question were included in the study. This study protected participants' confidentiality.

2.2. Measures

2.2.1. The Patient Health Questionnaire-4 (PHQ-4).

The PHQ-4 consists of the first two items of Patient Health Questionnaire-9 (PHQ-9) and the first two items of Generalized Anxiety Disorder-7 (GAD-7). Each of the items is rated on a 4-point Likert scale: 0 (Not at all),

1 (Several days), 2 (More than half the days), and 3 (Nearly every day). Total scores are rated from 0 to 12. Scores are rated as normal (0-2), mild (3-5), moderate (6-8), and severe (9-12). Total scores ≥3 for the first 2 questions suggests anxiety, and ≥3 for the last 2 questions suggests depression (Kroenke et al., 2009). A higher score on the PHQ-4 indicates poorer mental health (Meng et al., 2024).

2.2.2. World Health Organization-five Well-Being Index (WHO-5).

The five items of the WHO-5 ask about positive mood, vitality, and general interests. Each of the items is rated on a 6-point Likert scale. Previous studies have found acceptable psychometric properties of the WHO-5 in different versions and in various clinical and non-clinical samples (Dadfar et al., 2018). The Farsi version of the WHO-5 was available on the WHO website in 2017 (WHO, 2017).

2.2.3. The Self–Rating Scale of Mental Health and the Self-Rating Scale of Happiness.

These scales consist of separate single-item questions: "What is your estimation of your mental health in general?" and “To what degree do you feel happy in general?” These questions are rated on a scale of 11 numbers from 0 to 10. The participants were requested (a) to respond according to their global estimation and general feeling (and not their present states), (b) to know that 0 is the minimum and that 10 is the maximum score, and (c) to circle a number which seems to them to describe their actual feeling accurately. High scores indicate a high level of mental health and happiness. The one-week test–retest reliabilities of the two self-rating scales ranged between .86 and .89, indicating high temporal stability and corroborating the trait-like nature of the scores (Abdel-Khalek, 2006; Abdel-Khalek & Lester, 2017; Dadfar et al., 2021, 2025).

2..3. Procedure

In the present study, the English version of the PHQ-4 was translated into the Farsi language by two native Iranian, Farsi-speaking bilingual individuals, and this version was back-translated into English by another person who was a native English speaker. Based on this, discrepancies were resolved and the final version of the Farsi PHQ-4 was developed. The PHQ-4, the WHO-5 and the self-rating scales were designed as a Google form and delivered to subjects living in various areas of Gonbad Kavos City, Golestan Province, Iran. The respondents were recruited by a self-selection method. Informed consent was electronically obtained from participants in accordance with the Declaration of Helsinki. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Data were collected between 30/07/2022 and 30/11/2022.

2.4. Data analysis

The data were analyzed with descriptive statistics, Pearson correlation coefficients, and a principal components analysis (PCA) with varimax rotation using SPSS version 26. Eigenvalue greater than 1.0 and the scree plot test were used to determine the number of factors to be retained. Factor loadings > .5 were considered adequate. Confirmatory factor analysis (CFA) was applied with the AMOS program to evaluate whether the factor structure obtained by exploratory factor analysis (EFA) was fit.

3. Results

The mean total score for the Farsi version of the PHQ-4 was 3.98 (SD = 3.22). The skewness and kurtosis indicate that distribution of the data is near-normal, with only mild positive skew (Table 1).

The Kaiser-Meyer-Olkin measure of sampling adequacy for the PHQ-4 was .803. The Bartlett's test of sphericity chi-square was 916.2 (df = 6, $p < .001$). The PHQ-4 is a unifactorial scale. One salient component with an Eigenvalue of 2.84 was extracted for the PHQ-4 labeled: "Anxious/Depressed", which explained 71% of the total variance (Table 1 and Figure 1).

Table 1. Means (M), standard deviations (SD), skewness, kurtosis, and loadings of the items on the unrotated first principal component.

PHQ-4	Item 1	Item 2	Item 3	Item 4
Mean	1.29	.93	.86	.908
SD	.94	.95	.95	.99
Skewness	.459	.744	.840	.827
Kurtosis	-.646	-.419	-.300	-.414
Factor loading	.813	.858	.840	.858

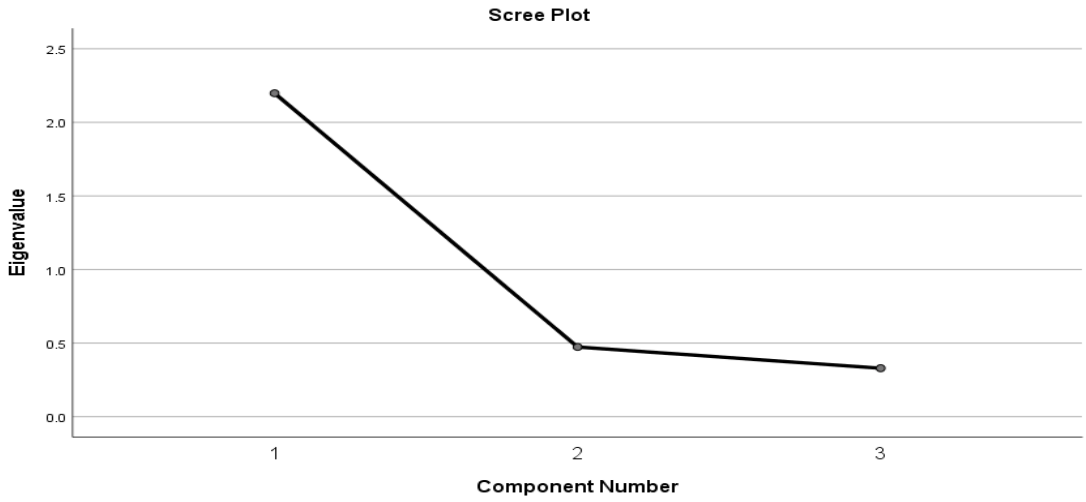

Figure 1. Scree plot of the Patient Health Questionnaire-4 (PHQ-4)

The single-factor solution presented in Figure 2 produced a good fit according to the fit indices. All estimated loadings were statistically significant ($p < .001$), the sign obtained was as expected, and the model was correctly identified (See Table 2).

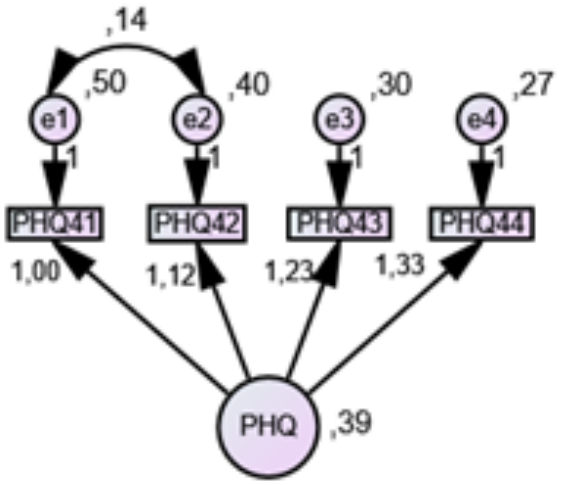

Figure 2. Measurement model with standardized loadings, PHQ-4 one-factor structure

Table 2. Model fit values of single factor and two factor structures.

	One factor	Two factors	Recommended
χ^2/df ratio	1.23	1.23	< 5
GFI	0.99	0.99	> 0.95
AGFI	0.98	0.98	> 0.90
CFI	1.00	1.00	> 0.90
NFI	0.99	0.99	> 0.90
RMSEA	0.022	0.022	< 0.08
SRMR	0.0054	0.0054	< 0.05

The two-factor solution had equally good fit according to the fit indices. All estimated loadings were statistically significant ($p < .001$), the covariance between PHQ-2 and GAD-2 was calculated as .59 ($p < .001$). The correlation coefficient between PHQ-2 and GAD-2 was .87. Standardized regression weights ranged from .76 to .85 (See Figure 3 and Table 3). Therefore, the internal consistency of the Farsi version of the PHQ-4 was high for both structures. The item-total correlations ranged from .503 to .689, all statistically significant at the .01 level. Cronbach's α was .86 (See Table 4).

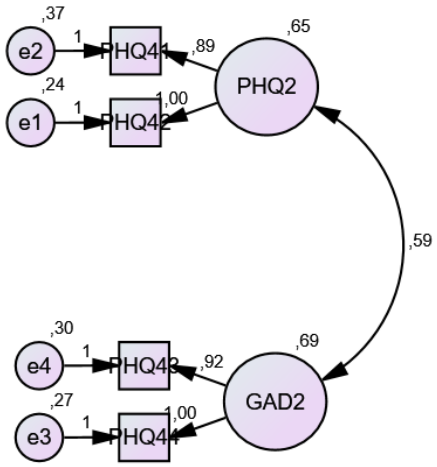

Figure 3. Measurement model with two-factors, standardized loadings

Table 3. PHQ-4 single- and two-factor standardized regression weights

Estimate (one factor)		Estimate (two factors)	
PHQ41 ← PHQ4	.666	PHQ41 ← PHQ2	.765
PHQ42 ← PHQ4	.742	PHQ42 ← PHQ2	.853
PHQ43 ← PHQ4	.814	PHQ43 ← GAD2	.814
PHQ44 ← PHQ4	.846	PHQ44 ← GAD2	.846

Table 4. Inter-item correlations, item-total correlations, and Cronbach's α of the PHQ-4. All correlations are statistically significant at $p < .01$.

Over the **last 2 weeks**, how often have you been bothered by any of the following problems?	1	2	3	4
1. Feeling nervous, anxious, or on edge	1			
2. Being unable to stop or control worrying	.653	1		
3. Little interest or pleasure in doing things	.530	.611	1	
4. Feeling down, depressed, or hopeless	.573	.623	.689	1
PHQ-4 total score	.816	.855	.840	.859
Cronbach's α	.86			

Table 5 shows that correlations of the PHQ-4 with WHO-5 and the two self-rating scales of mental health and happiness were largestatistically significant and negativ (-.71, -.53, and -.56, all statistically significant at the .01 level), indicating good validity.

Table 5. Means (M) and standard deviations (SD) of the other scales, and their correlations with the PHQ-4. All correlations are statistically significant at $p < .01$ (two-tailed).

Scales	M (SD)	*r* with PHQ-4 total score
World Health Organization-five Well-being Index (WHO-5)	14.98 (6.13)	-.719
Self-rating Scale of Mental Health	7.15 (2.34)	-.539
Self-rating Scale of Happiness	7.28 (2.27)	-.566

4. Discussion

The purpose of this study is to determine the validity and reliability of the Farsi version of PHQ-4, explore the relationships between anxiety and depression, psychological well-being, mental health, and happiness, and investigate gender differences in the general population of Iran. Exploratory factor analysis revealed a single-factor structure explaining 71% of the variance. The factor loadings ranged from .813 to .858 and item-total correlations were similarly high, indicating high validity of the Farsi version of PHQ-4. These results show good internal consistency of this 4-item scale. Demirci and Ekşi (2018) standardized PHQ-4 in Turkish and found correlations between each item and the total score ranging from .54 to .74.

Using confirmatory factor analysis, researchers obtained good fit values for both single- and two-factor structures. Materu et al. (2020) reported a single-factor structure in an adaptation study with 2426 participants in Tanzania. However, Kroenke et al. (2009) obtained a two-factor structure explaining 84% of the variance in their standardization study in the United States. Lowe et al. (2010) confirmed a one- and two-factor structure in a German population study. Kocalevent et al. (2014) standardized the Colombian version and confirmed a two-factor structure through confirmatory factor analysis. Stanhope (2016) found a two-factor structure in the confirmatory factor analysis of PHQ-4 in Australia. Ghaheri et al. (2020) confirmed the validity of the two-factor structure in an Iranian validation study. In a study conducted in Southeast Asia during the COVID-19 pandemic with 4,524 participants, a two-factor structure was found to be more consistent than a single-factor structure (Mendoza et al., 2022). Also, Havnen et al. (2023) reported a two-factor structure in Norwegian adults with attention deficit hyperactivity disorder (ADHD). Meng et al. (2024) obtained a two-factor structure in 512 Chinese healthcare students. The differences in the results among studies may arise from variations in statistical analyses and sample characteristics.

In this study, the internal consistency of the PHQ-4's Farsi version had high Cronbach's alpha (.88), indicating high reliability. Kroenke et al. (2009) reported an internal consistency coefficient of .85 for PHQ-4. Most other standardization studies also achieved Cronbach's alpha values greater than .80: Materu et al. (2020) .81; Lowe et al. (2010) .82; Kocalevent et al. (2014) .84; Demirci and Ekşi (2018) .83; Mendoza et al. (2022) .82; Stanhope (2016) .75, and Ghaheri et al. (2020) .76. A Cronbach's alpha value greater than .70 indicates that the scale is a reliable tool for measuring its construct (Adams & McGuire, 2023, p. 184).

Our results using confirmatory factor analysis indicate that the single- and two- factor structure of the scale both have good fit values. Previous studies support the validity of the single-factor structure. Standard regression factor loadings ranged from .66 to .84, all statistically significant ($p < 0.001$) The confirmatory factor analysis results support the validity of the single-factor structure. Additionally, when the two-factor structure of PHQ-2 and GAD-2 was considered in confirmatory factor analysis, similar fit values were obtained. The covariance between PHQ-2 and GAD-2 was calculated as .59, significant at $p < .001$, and the correlation coefficient between them was .87. Standard regression weights ranged from .76 to .85. The

internal consistency of the Farsi PHQ-4 was high for both the single-factor and two-factor structures (See Tables 2 and 3). With a correlation of .87 between PHQ-2 and GAD-2, the conclusion should be that there effectively are no two factors, that there is no separation between anxiety and depression.

Demirci and Ekşi (2018) reported CFI = 1.00, TLI = 1.00, SRMR = .008, RMSEA = .000 for both the single-factor and two-factor structures in their confirmatory factor analysis. Lowe et al. (2010) found good fit values in their confirmatory factor analysis, with CFI=0.984, TLI=0.988, RMSEA=0.027, and a 90% confidence interval for RMSEA of 0.023–0.032. Ghaheri et al. (2020) reported that the single-factor structure did not have robust fit values, but the two-factor structure did. Kocalevent et al. (2014) applied confirmatory factor analysis tor both two-factor and single-factor structures, stating that the two-factor structure had fit indices, except RMSEA, that indicated good fit (CMIN/DF = 32.31; GFI = 0.989; NFI = 0.987; TLI = 0.923; CFI = 0.987, RMSEA = 0.145). The single-factor structure showed poorer fit (CMIN/DF = 114.45; GFI = 0.964; NFI = 0.953; TLI = 0.861, CFI = 0.954, RMSEA = 0.194). The findings of the present study are in line with these results.

Our study also investigated the relationship between the well-being index accepted by the WHO (WHO-5) and PHQ-4. The results indicated a strong and highly significant ($p<.001$) negative correlation between the two scales. The negative correlation suggests that as scores from the WHO-5 (measuring well-being) increase, scores on the PHQ-4 (measuring anxiety and depression) decrease, confirming the validity of PHQ-4 as a measurement tool. Havnen et al. (2023) found a correlation of r =-.75 between PHQ-4 and WHO-5 in adults with ADHD. Ghaheri et al. (2020) reported correlations of -.51 between WHO-5 and PHQ-2, -.47 between WHO-5 and GAD-2, and -.55 between WHO-5 and PHQ-4. The relatively higher correlation in this study supports the findings in the literature.

In this study, a single-item "Self-rating Scale of Mental Health" and a single-item "Self-rating Scale of Happiness" were applied to determine the validity of PHQ-4. Both scales aim to assess an individual's positive mental state. Therefore, a negative relationship was expected between the scores on these scales and the PHQ-4 scores. The correlation analysis revealed a significant negative correlation between the PHQ-4 and both mental health ($r = -.53$) and happiness ($r = -.56$) at $p < .001$. These findings indicate that PHQ-4 is an effective and valid measure of depression and anxiety. Studies supporting this counclusion include Kroenke et al. (2009), who examined the correlation between the PHQ-4 and the Mental Health subscale of SF–20, reporting a strong correlation of .80. Kocalevent et al. (2014) found weak but significant correlations between PHQ-4 and the General Health Questionnaire-12 (GHQ-12), Hopelessness (BHS), Self-efficacy (GSES), and Life satisfaction (QLS). Also, Demirci and Ekşi (2018) reported a statistically significant negative relationship between the PHQ-4 and Life satisfaction ($r = -.38$, $p < .001$). All these results taken together show that much or most variation on these scales is caused by something we can call a “general factor of psychological well-being”. General factors are common in psychology. We find them for intelligence, well-being, and psychopathology. Our results support the importance of general factors, at least in well-being and psychopathology. In other words, when we discuss the correlations with other tests measuring similar constructs and we find high correlations, this supports the contention that to a large extent all these tests measure the same construct, which is general psychological well-being. It can be argued that any one of these well-being tests, including the PHQ-4, can be used to measure this general factor but also that it is redundant with the others.

In our Iranian study, the mean PHQ-4 scale score was determined to be 3.95. This score falls within the cutoff range of 3-5 (mild), indicating that the mental health of Iranian society is relatively good. Similar results have been found in studies conducted in various countries. For instance, Demirci and Ekşi (2018)

studied 450 healthy individuals from a Turkish sample and found that the average PHQ-4 score was 4.84. Similarly, Materu et al. (2020) reported a PHQ-4 score of 3.75 in their study on out-of-school adolescent girls and young women in Tanzania. However, in some societies, the total PHQ-4 score has been reported to fall within the 0-2 (normal) cutoff range. For example, Löwe et al. (2010) studied adults aged 14 and over in a German sample. They reported a PHQ-4 score of 1.76, while Kocalevent et al. (2014) reported a PHQ-4 score of 1.27 in a study conducted on the general population over 18 in Colombia. Studies conducted on younger groups outside the general population have also yielded noteworthy results. Khubchandani et al. (2016) found an average PHQ-4 score of 2.98 in a sample of university students in the United States. Other studies have measured PHQ-4 scores in patient groups. For example, Havnen et al. (2023) conducted a study on Norwegian-speaking individuals undergoing treatment for attention deficit hyperactivity disorder (ADHD) and found that the total PHQ-4 score among participants was 5.0. Kroenke et al. (2009) examined the averages for each item in a study conducted on a patient sample in the United States and found a general PHQ-4 score of 2.5. Considering the Iranian context's intense economic, social, and political factors, this Iranian score can be seen as quite significant. Comparing them, we see that Iranian “normal” people have more mental health problems than American patients.

The study found no statistically significant sex differences in PHQ-4 scores. In contrast, Ghaheri et al. (2020) reported statistically significant ($p < .001$) gender differences in PHQ-4 scores, with females ($M = 5.45$) scoring higher than males ($M = 3.67$). Kocalevent et al. (2014) also found that females ($M = 1.4$) had higher PHQ-4 averages than males ($M = 1.1$). The findings of this study do not align with these results. This is indeed a surprising finding, because it is well known that women are the more neurotic sex. At least our results are trending in the predicted direction.

The present study has some limitations. It is based on a convenience sample of Iran, and the findings were based only on Iranian community residents. Another limitation is the number of participants. This is a cross-sectional design. It needs to be highlighted that there is a possibility of translator bias or errors in this study. A further limitation is the lack of other similar measures as indicators used to confirm theoretical consistency (construct validity, or at least criterion validity). Based on Table 4 we can conclude that there is only slight evidence for a correlational separation between anxiety and depression because the highest inter-item correlations are between items 1 and 2 and between items 3 and 4. But the other correlations, between anxiety items and depression items, are only slightly lower. Therefore, for practical purposes, the PHQ-4 cannot be used to distinguish between anxiety and depression. This is a limitation of this scale, or a limitation of our diagnostic categories. In realitly PHQ-4 is a screening tool for general mental health. These limitations could be mentioned as possibilities for future works. A proposed next step would be to study the performance of different normal and abnormal groups on the PHQ-4. These are projects for future studies. One question that we could raise for future research is whether the correlations among the many alternative well-being measures are higher in poor countries than in rich countries, or higher in more educated than in less educated people in Iran. Perhaps the correlations among these alternative well-being scales are lower among those with higher education indicating that their emotional system is somehow more differentiated than that of less sophisticated individuals.

5. Conclusions

The PHQ-4 had good psychometric properties in the present sample from Iran. A principal components analysis (PCA) with a varimax rotation showed one factor labelled “Anxious/Depressed". Using exploratory factor analysis (EFA), the scale has both one-factor and a two-factor structure, consisting of a depression

factor and an anxiety factor. Confirmatory factor analysis (CFA) showed the PHQ-4 has good fit values for both structures. This study provides evidence for the usefulness of the Farsi version for assessing the anxiety and depression symptoms in Iranian community residents. The study demonstrates significant novelty by translating and validating an ultra-brief screening tool within this cultural context, providing a practical method to screen for depression and anxiety efficiently while reducing time and costs. The study findings suggest the use of the PHQ-4 in the national language of any country where it is used and where the national language is not English. Hence, we suggest that such trials be conducted in other non-English speaking countries.

Acknowledgements: We thank all the subjects for their participation in the study, and the research assistants for helping in the collection of data.
The Ethics Committee for Social and Human Sciences Research approval of Ordu University authorized the permission to conduct this study.

References

Abdel-Khalek, A. M. (2006). Measuring happiness with a single-item scale. *Social Behavior and Personality, 34,* 139-150.

Abdel-Khalek, A. M., & Lester, D. (2017). The association between religiosity, generalized self-efficacy, mental health, and happiness in Arab college students. *Personality and Individual Differences, 109,* 12-16.

Adams, K. A., & McGuire, E. K. (2023). *Student Study Guide with IBM® SPSS® Workbook for Research Methods, Statistics, and Applications* (3rd edition). Sage.

Beck, A. T. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*(6), 561. https://doi.org/10.1001/archpsyc.1961.01710120031004

Beck, A. T., & Alford, B. A. (2009). *Depression: Causes and Treatment* (2nd ed). University of Pennsylvania Press.

Beck, A. T., & Beck, R. W. (1972). Screening depressed patients in family practice: A rapid technic. *Postgraduate Medicine, 52*(6), 81-85.

Black, D. W., & Grant, J. E. (2014). *DSM-5 Guidebook: The Essential Companion to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition* (1st edition). American Psychiatric Publishing.

Corson, K., Gerrity, M. S., & Dobscha, S. K. (2004). Screening for depression and suicidality in a VA primary care setting: 2 items are better than 1 item. *American Journal of Managed Care, 10*(11 pt. 2), 839-845.

Dadfar, M., Lester, D., & Abdel-Khalek, A. (2021). Religiosity, happiness and suicidal behavior: A cross-sectional comparative study in Iran. *Mental Health, Religion & Culture, 24*(2), 128-141.

Dadfar, M., Momeni Safarabad, N., Asgharnejad Farid, A. A., Nemati Shirzy, M., & Ghazie Pour Abarghouie, F. (2018). Reliability, validity, and factorial structure of the World Health Organization-5 Well-Being Index (WHO-5) in Iranian psychiatric outpatients. *Trends in Psychiatry and Psychotherapy, 40*(2), 79-84.

Dadfar, M., Turan, Y., Sheybani, F., & Navari, N. (2025). Validation of the General Anxiety Disorder-7 (GAD-7): Its association with psychological factors in an Iranian sample. *Mankind Quarterly, 65*(3), 336-344.

Demirci, İ., & Ekşi, H. (2018). Don't bother your pretty little head otherwise you can't enjoy life. *ERPA International Congress on Education*, *2018*, 287-292.

Depressive disorder (depression) (2023). *World Health Organization*. https://www.who.int/news-room/fact-sheets/detail/depression

Dohrenwend, B. P. (1990). Socioeconomic status (SES) and psychiatric disorders: Are the issues still compelling? *Social Psychiatry and Psychiatric Epidemiology, 25*(1), 41-47.

EuroQol Group (1990). EuroQol-A new facility for the measurement of health-related quality of life. *Health Policy, 16*(3), 199-208.

Fryers, T., Melzer, D., Jenkins, R., & Brugha, T. (2005). The distribution of the common mental disorders: Social inequalities in Europe. *Clinical Practice and Epidemiology in Mental Health, 1*(1), 14. https://doi.org/10.1186/1745-0179-1-14

Ghaheri, A., Omani-Samani, R., Sepidarkish, M., Hosseini, M., & Maroufizadeh, S. (2020). The Four-Item Patient Health Questionnaire for Anxiety and Depression: A validation study in infertile patients. *International Journal of Fertility and Sterility, 14*(3), 234-239.

Gorman, J. M. (1996). Comorbid depression and anxiety spectrum disorders. *Depression and Anxiety, 4*(4), 160-168.

Haehner, P., Würtz, F., Kritzler, S., Kunna, M., Luhmann, M., & Woud, M. L. (2024). The relationship between the perception of major life events and depression: A systematic scoping review and meta-analysis. *Journal of Affective Disorders, 349,*145-157.

Havnen, A., Lydersen, S., Mandahl, A., & Lara-Cabrera, M. L. (2023). Factor structure of the Patient Health Questionnaire-4 in adults with attention-deficit/hyperactivity disorder. *Frontiers in Psychiatry, 14,* 1176298. https://doi.org/10.3389/fpsyt.2023.1176298

Institute for Health Metrics and Evaluation (IHME) (2024). Global Burden of Disease "Depressive disorders prevalence" [dataset], with major processing by Our World in Data. Retrieved January 10, 2025 from https://ourworldindata.org/grapher/depressive-disorders-prevalence-ihme

Judd, L. L., Kessler, R. C., Paulus, M. P., Zeller, P. V., Wittchen, H.-U., & Kunovac, J. L. (1998). Comorbidity as a fundamental feature of generalized anxiety disorders: Results from the National Comorbidity Study (NCS). *Acta Psychiatrica Scandinavica, 98*(s393), 6-11.

Khubchandani, J., Brey, R., Kotecki, J., Kleinfelder, J., & Anderson, J. (2016). The psychometric properties of PHQ-4 depression and anxiety screening scale among college students. *Archives of Psychiatric Nursing*, 30(4), 457-462.

Kocalevent, R.-D., Finck, C., Jimenez-Leal, W., Sautier, L., & Hinz, A. (2014). Standardization of the Colombian version of the PHQ-4 in the general population. *BMC Psychiatry, 14*(1), 205. https://doi.org/10.1186/1471-244X-14-205

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2003). The Patient Health Questionnaire-2: Validity of a two-Item depression screener. *Medical Care, 41*(11), 1284-1292.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., Monahan, P. O., & Löwe, B. (2007). Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. *Annals of Internal Medicine, 146*(5), 317. https://doi.org/10.7326/0003-4819-146-5-200703060-00004

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606-613.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: The PHQ-4. *Psychosomatics, 50*(6), 613-621.

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B. W., Berry, J. T., & Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders, 114*(1–3), 163-173.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical Care*, 46(3), 266–274.

Löwe, B., Kroenke, K., & Gräfe, K. (2005). Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *Journal of Psychosomatic Research, 58*(2), 163-171.

Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., Schneider, A., & Brähler, E. (2010). A 4-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders, 122*(1-2), 86-95.

Materu, J., Kuringe, E., Nyato, D., Galishi, A., Mwanamsangu, A., Katebalila, M., Shao, A., Changalucha, J., Nnko, S., & Wambura, M. (2020). The psychometric properties of PHQ-4 anxiety and depression screening scale among out of school adolescent girls and young women in Tanzania: A cross-sectional study. *BMC Psychiatry, 20*(1), 321. https://doi.org/10.1186/s12888-020-02735-5

Mendoza, N. B., Frondozo, C. E., Dizon, J., & Buenconsejo, J. U. (2022). The factor structure and measurement invariance of the PHQ-4 and the prevalence of depression and anxiety in a Southeast Asian context amid the COVID-19 pandemic. *Current Psychology*, *43*(25), 22019-22028.. https://doi.org/10.1007/s12144-022-02833-5

Meng, R., Jiang, C., Dzierzewski, J. M., Zhu, Y., Wang, M., Yang, N., Liu, X., ..., & Spruyt, K. (2024). Longitudinal measurement invariance and psychometric properties of the Patient Health Questionnaire-Four in China. *BMC Psychiatry, 24,* 517. https://doi.org/10.1186/s12888-024-05873-2

Mitchell, A. J., & Coyne, J. C. (2007). Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *British Journal of General Practice, 57*(535), 144-151.

Murray, C. J. L. (2020). GBD 2019 Diseases and Injuries Collaborators: Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Institute for Health Metrics and Evaluation, 396,* 1204-1222.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385-401.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., ..., & Keller, M. B. (2003). The 16-item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C),

and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573-583.

Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., …, & Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet, 398*(10312), 1700-1712.

Sartorius, N., Üstün, T. B., Lecrubier, Y., & Wittchen, H.-U. (1996). Depression comorbid with anxiety: Results from the WHO Study on Psychological Disorders in Primary Health Care. *British Journal of Psychiatry, 168*(S30), 38-43.

Sıncar, S., Sohbet, R., Bırımoğlu Okuyan, C., & Karasu, F. (2020). Üniversite Öğrencilerinde Yaşam Doyumu ve Depresyon: Kesitsel Bir Çalışma [Life satisfaction and depression in university students: A cross-sectional study]. *Balıkesir Sağlık Bilimleri Dergisi [Balıkesir Health Sciences Journal], 9*(2), 93-100.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*(10), 1092-1097.

Stanhope, J. (2016). Patient Health Questionnaire-4. *Occupational Medicine, 66*(9), 760-761.

Tiller, J. W. (2013). Depression and anxiety. *Medical Journal of Australia, 199*(S6), S28-S31.

Wetzler, S., & Katz, M. M. (1989). Problems with the differentiation of anxiety and depression. *Journal of Psychiatric Research, 23*(1), 1-12.

WHO (2017). *The WHO-5-Region Hovedstadend Psykiatri*. Hillerød: Psychiatric Research Unit, Mental Health Centre North Zealand.

WHO (2022). Mental health. https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response

WHO (2023). Anxiety disorders. https://www.who.int/news-room/fact-sheets/detail/anxiety-disorders.