Arabic Translation, Validation and Cultural Adaptation of the 7-Item Hamilton Depression Rating Scale in Two Community Samples

Ahmad N. Alhadi,1 *Mohammed A. Alarabi,1 Abdulaziz T. Alshomrani,2 Raafat M. Shuqdar,1
Mohammad T. Alsuwaidan,4 Roger S. McIntyre5

Abstract: Objectives: Depression is a common mental disorder, the severity of which is frequently assessed via interview-based clinical scales such as the 7-item Hamilton Depression Rating Scale (HAM-D-7). The current study aimed to translate and examine the validity of an Arabic version of the HAMD-7 scale. Methods: This study took place between February and March 2016 in the Psychiatry Department of King Saud University, Riyadh, Saudi Arabia. The HAMD-7 scale was translated into Arabic using forward and backward translation methods. A total of 153 Arabic speakers were recruited to test the translated scale, including 57 medical students and 96 members of the general public. The Arabic version of the HAMD-7 scale was completed by trained investigators during face-to-face interviews with the participants. In order to assess convergent validity, participants also completed an Arabic version of the self-assessed Patient Health Questionnaire-9 (PHQ-9) scale. Subsequently, the test-retest reliability of the translated HAMD-7 scale was evaluated two weeks later during a second interview. Results: Overall, HAMD-7 scores were positively correlated with PHQ-9 scores (r = 0.633–0.749). Moreover, the translated HAMD-7 scale proved to be reliable in terms of test-retest reliability (intra-class correlation coefficient: 0.807; P < 0.001). With regards to internal consistency, the Cronbach's a values ranged between 0.607–0.756. Conclusion: The Arabic HAMD-7 scale was found to be reliable and valid among two samples of Arabic speakers in Saudi Arabia. However, further research among Arabic-speaking patients diagnosed with depression is needed in order to establish its usefulness in assessing the severity of depressive symptoms.

Keywords: Psychiatry; Depression; Psychometrics; Validity and Reliability; Translation; Questionnaire Design; Saudi Arabia.

Advances in Knowledge
- To the best of the authors' knowledge, this study is the first to translate the 7-item Hamilton Depression Rating Scale (HAM-D-7) into Arabic using forward and backward translation methods.

References

1Department of Psychiatry, King Saud University, Riyadh, Saudi Arabia; 2College of Medicine, Al-Imam University, Riyadh, Saudi Arabia; 3College of Medicine, Taibah University, Madinah, Saudi Arabia; 4Department of Psychiatry, Faculty of Medicine, Kuwait University, Kuwait; 5Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

*Corresponding Author’s e-mail: malarabi@ksu.edu.sa

المصطلحات: الهدف: بعد الانتقادات من الإحترامات النفسية الشائعة الذي يتم تقديم علاج للعظام باستخدام مقاييس صغيرة مثل مقياس هاميلتون للإكتئاب، تم اجراء هذه الدراسة لإثبات موثوقية المقياس للإكتئاب لدى مرضى الاكتئاب في المملكة العربية السعودية من خلال مختبر التدريبيات. تم إجراء الدروس في كلية الطب - جامعة الملك سعود في الرياض، بالسعودية باستخدام معالجات إلكترونية. وتم إعادة تجربة المقياس مرات عديدة لضمان توازي النتائج. والنتائج: بعد فحص التأكد من صحة الترجمة، تم استخدام مقياس هاميلتون للإكتئاب لدى المرضى باللغة العربية ومن ثم إعادة تجربته باستخدام تقنيات المختبر الرئيسي. الهدف: بعد الانتقادات من الإحترامات النفسية الشائعة الذي يتم تقديم علاج للعظام باستخدام مقاييس صغيرة مثل مقياس هاميلتون للإكتئاب، تم اجراء هذه الدراسة لإثبات موثوقية المقياس للإكتئاب لدى مرضى الاكتئاب في المملكة العربية السعودية من خلال مختبر التدريبيات. تم إجراء الدروس في كلية الطب - جامعة الملك سعود في الرياض، بالسعودية باستخدام معالجات إلكترونية. وتم إعادة تجربة المقياس مرات عديدة لضمان توازي النتائج. والنتائج: بعد فحص التأكد من صحة الترجمة، تم استخدام مقياس هاميلتون للإكتئاب لدى المرضى باللغة العربية ومن ثم إعادة تجربته باستخدام تقنيات المختبر الرئيسي.
Depression is a frequent problem in the general population and the most common mental health condition among patients seen at primary care facilities. Moreover, it is currently the leading cause of disability and premature death among those aged 18–43 years old. However, unlike many other chronic medical disorders, there are currently no clinically relevant biological markers for depression, meaning that the assessment of depressive symptoms depends on standardised rating tools. Patient-administered self-assessment screening tools are a quick and reliable option in the initial assessment of depression or for treatment monitoring purposes. Such scales frequently used in clinical practice and research include the Beck Depression Inventory (BDI), Patient Health Questionnaire-9 (PHQ-9), Center for Epidemiologic Studies Depression (CES-D) scale and the Zung Self-Rating Depression Scale.

The Hamilton Depression Rating Scale (HAMD) is one of the most widely used scales to objectively rate depression severity worldwide. The original scale contains 17 items, of which four do not contribute to the final score and instead provide additional clinical information. The other items are scored on a 3–5 point scale, with the final score indicative of the severity of the patient’s depression. Since its development in 1960, structured interview guides, self-report forms and computerised versions of the scale have been created in an effort to standardise administration of the tool and improve the reliability and validity of individual items. Researchers have also attempted to shorten the scale while preserving its psychometric properties; accordingly, a 7-item version was developed in 2003 to estimate depression severity based on criteria in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders. The 7-item HAMD (HAMD-7) scale utilises the same scoring system as the original version, but can also be used to evaluate treatment efficacy by distinguishing between a clinical response to treatment and full symptomatic remission. McIntyre et al. reported that the HAMD-7 scale takes significantly less time to complete than the original version, while offering the same level of accuracy.

Overall, the HAMD-7 scale focuses on a subset of symptoms covered in the original HAM-D (i.e. depressed mood, feelings of guilt, interest, pleasure and level of activity, psychological anxiety, somatic anxiety, energy level and suicidality) and differs from other depression rating scales like the BDI or the CES-D by omitting various indicators of depression, such as altered sleep patterns and changes in appetite. In contrast, the 6-item version of the HAMD scale has shown good psychometric properties and has been translated into several languages, including Arabic. However, this version does not score for suicidality, which may limit its clinical usefulness. Moreover, a review of the literature did not reveal any other brief validated structured Arabic-language interview-based scale to measure depression severity and response to treatment in clinical settings. Therefore, this study aimed to translate the HAMD-7 scale into Arabic and determine the validity and reliability of the final Arabic version.

**Methods**

This study took place between February and March 2016 in the Psychiatry Department of King Saud University, Riyadh, Saudi Arabia. A multi-step process was undertaken to translate the HAMD-7 scale into Arabic and validate it in an Arab population. The original English-language version of the HAMD-7 scale was translated using established forwards and backwards translation methods. First, two independent translators each translated the English-language version of the scale into Arabic. A committee of Arabic-speaking researchers then compared these translations both with each other and with the original English version in order to create a first draft. Next, the two independent translators translated the draft back into English and the committee again compared both back-translated versions together and with the first draft to ensure the accuracy of the content. A second draft was then created and examined to resolve any ambiguities.

The second draft, along with the original English-language version of the HAMD-7 scale, was submitted to five mental health experts to check face validity, with minor modifications made based on their feedback. The final Arabic version of the scale was deemed to reflect an accurate translation of the items in the original HAMD-7 scale. A pilot study was subsequently conducted among 10 healthy bilingual adults (six males and four females) to assess the results of this study indicate that the translated scale had acceptable reliability and convergent validity among two samples of Arabic-speaking participants in Saudi Arabia.

**Application to Patient Care**

- The validated Arabic version of the HAMD-7 scale may be a potentially useful psychiatric tool for the assessment of depression among Arabic-speaking patients, due to its brevity and focus on major aspects of depressive disorders.
Ahmad N. Alhadi, Mohammed A. Alarabi, Abdulaziz T. Alshomrani, Raafat M. Shuqdar, Mohammad T. Alsuwaidan and Roger S. McIntyre

Clinical and Basic Research | e169

the clarity of the items in the final Arabic version of the HAMD-7 scale, estimate the time required for the interview and carry out a preliminary investigation of the psychometric properties of the scale. The results indicated that the Arabic HAMD-7 scale was time-efficient, taking approximately 4 minutes to administer. Moreover, feedback from the participants in the pilot study refined the translation of the sixth item in that "energy level" was instead translated to "activity level" due to the latter term being more comprehensible in Arabic.

A total of 153 Arabic-speaking participants were recruited to complete the Arabic HAMD-7 scale in two separate samples. The first sample (n = 57) was selected randomly using online randomisation software (Research Randomizer, Version 4.0, Wesleyan University, Middletown, Connecticut, USA) from a list of fourth-year medical students at King Saud University. The second sample (n = 96) was recruited as a convenience sample from members of the general Arabic-speaking population at two public parks in Riyadh. The inclusion criteria encompassed individuals with Arabic literacy and no pronounced cognitive deficits. The sample size was calculated using a subject-to-item ratio with at least 10 subjects per item, resulting in a total required sample size of 70 participants.20 Each participant was interviewed by trained investigators who administered the Arabic HAMD-7 scale while following a structured interview guide.13 In order to measure the test-retest reliability of the scale, the translated HAMD-7 scale was administered to each subject twice, including once at enrolment (day one) and then again two weeks after the initial interview (day 14). Each participant was assigned a code number so that their scores in the first and second interviews could be paired for analysis. Subjects also completed an Arabic version of the self-administered PHQ-9 during the first interview so as to establish convergent validity.5

The Statistical Package for the Social Sciences (SPSS), Version 21.0 (IBM Corp., Armonk, New York, USA), was used to carry out the analysis of the data. Due to the skewed distribution of the scale scores, Spearman's rank correlation coefficient with a two-tailed test of significance was used to test agreement between the Arabic HAMD-7 and PHQ-9 scores and establish convergent validity. Internal consistency of the Arabic HAMD-7 scale was deemed satisfactory at a Cronbach's α value of >0.600.22 The test-retest reliability was quantified using intra-class correlation coefficient (ICC) values for absolute agreement. A P value of <0.050 was deemed statistically significant.

Ethical approval for this study was granted by the King Saud University Institutional Review Board prior to data collection (#E-15-1691). Each participant provided written informed consent before being included in the study. The identity of each participant was codified for the purposes of anonymity and no incentives or rewards were given for participation.

Results
A total of 153 Arabic-speaking individuals took part in the study, including 57 medical students and 96 members of the general public [Figure 1]. Among the first sample, 38 were female (66.7%) and 19 were male (33.3%). The mean age was 22.12 ± 0.44 years old.
Arabic Translation, Validation and Cultural Adaptation of the 7-Item Hamilton Depression Rating Scale in Two Community Samples

The second sample consisted of 74 female (77.1%) and 22 male (22.9%) members of the general public. The mean age was 27.42 ± 8.28 years (range: 19–63 years). Among the first sample, mean HAMD-7 scores were 4.3 ± 2.9 (range: 0–16) and 3.5 ± 3.4 (range: 0–13) at days one and 14, respectively. The mean PHQ-9 score was 5.3 ± 4.3 (range: 0–17).

Among the second sample, the initial mean HAMD-7 score was 4.2 ± 3.5 (range: 0–16), while the mean score two weeks later was 3.2 ± 3.4 (range: 0–13). The mean PHQ-9 score was 6.3 ± 4.3 (range: 0–13). For both samples combined, the mean HAMD-7 scores were 4.2 ± 3.3 and 3.6 ± 3.4 at days one and 14, respectively, while the mean PHQ-9 score was 5.7 ± 4.3 [Table 1]. In general, female participants had higher HAMD-7 scores, with a mean difference of 1.83 at day one and 1.54 at day 14 (P <0.010 each). The PHQ-9 scores differed as well, with females scoring an average of 1.5 points higher than males (P = 0.920).

The internal consistency of the Arabic HAMD-7 scale was acceptable among both the first and second samples during the initial interview (Cronbach’s α: 0.607 and 0.660, respectively) and two weeks later (Cronbach’s α: 0.756 and 0.719, respectively). Moreover, the internal consistency remained acceptable after the scores of both samples were combined (Cronbach’s α: 0.641 and 0.728, respectively). Furthermore, the test-retest reliability of the translated scale was maintained among both samples, with ICC values of 0.836 for the first sample and 0.793 for the second sample [Table 2]. For convergent validity, the initial scores of the Arabic HAMD-7 scale were significantly positively correlated with scores from the PHQ-9 for both samples (r = 0.668 and 0.749, respectively; P <0.001 each). The second HAMD-7 scores also correlated with the initial PHQ-9 scores for both samples (r = 0.714 and 0.648; P <0.001 each) [Table 3].

In terms of individual items on the Arabic HAMD-7 scale, the lowest measure of inter-item correlation was between the suicidality and psychological anxiety items (0.046), while the highest was between the energy level and somatic anxiety items (0.428) [Table 4]. In addition, an analysis of the psychometric properties of the individual items revealed that the internal consistency of the translated scale did not improve with the deletion of any of the seven items [Table 5]. All of the individual items of the HAMD-7 scale correlated positively with the PHQ-9 scores (P <0.050).

Discussion

This study aimed to validate an Arabic version of the HAMD-7 scale to ensure conceptual and performance equivalence with the original scale.14 During the interviews, the participants noted that they found the individual items in the scale easy to understand; however, in the event of any ambiguity, the interviewer provided additional clarification, as recommended in the structured interview guide for the original HAMD

<table>
<thead>
<tr>
<th>Scale</th>
<th>Total</th>
<th>First sample (n = 57)</th>
<th>Second sample (n = 96)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAMD-7</td>
<td>Day one</td>
<td>Mean ± SD</td>
<td>4.2 ± 3.3</td>
</tr>
<tr>
<td></td>
<td>Median (IQR)</td>
<td>4 (6–2)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Day 14</td>
<td>Mean ± SD</td>
<td>3.6 ± 3.4</td>
</tr>
<tr>
<td></td>
<td>Median (IQR)</td>
<td>3 (6–1)</td>
<td>-</td>
</tr>
<tr>
<td>PHQ-9</td>
<td>Mean ± SD</td>
<td>5.7 ± 4.3</td>
<td>6.3 ± 4.3</td>
</tr>
<tr>
<td></td>
<td>Median (IQR)</td>
<td>5 (8–2)</td>
<td>-</td>
</tr>
</tbody>
</table>

HAMD-7 = 7-Item Hamilton Depression Rating; PHQ-9 = Patient Health Questionnaire-9; SD = standard deviation; IQR = interquartile range.

Table 2: Internal consistency and test-retest reliability of an Arabic version of the 7-Item Hamilton Depression Rating scale14 among two samples in Riyadh, Saudi Arabia (N = 153)

<table>
<thead>
<tr>
<th>Cronbach’s α</th>
<th>ICC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>First sample (n = 57)</td>
<td>0.607</td>
</tr>
<tr>
<td>Second sample (n = 96)</td>
<td>0.660</td>
</tr>
<tr>
<td>Total</td>
<td>0.641</td>
</tr>
</tbody>
</table>

ICC = intraclass correlation coefficient.
*Tested for absolute agreement

Table 3: Convergent validity between Arabic versions of the 7-Item Hamilton Depression Rating scale14 and Patient Health Questionnaire-9 among two samples in Riyadh, Saudi Arabia (N = 153)

<table>
<thead>
<tr>
<th>Rho</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day one</td>
<td>Day 14</td>
</tr>
<tr>
<td>First sample (n = 57)</td>
<td>0.668</td>
</tr>
<tr>
<td>Second sample (n = 96)</td>
<td>0.749</td>
</tr>
<tr>
<td>Total</td>
<td>0.717</td>
</tr>
</tbody>
</table>

HAMD-7 = 7-Item Hamilton Depression Rating Scale; PHQ = Patient Health Questionnaire.
*Calculated using Spearman’s rank correlation coefficient with a two-tailed test of significance.

Table 4: Convergent validity between Arabic versions of the 7-Item Hamilton Depression Rating scale14 and Patient Health Questionnaire-9 among two samples in Riyadh, Saudi Arabia (N = 153)

Table 5: The psychometric properties of the individual items of the HAMD-7 scale (N = 153)

Discussion

This study aimed to validate an Arabic version of the HAMD-7 scale to ensure conceptual and performance equivalence with the original scale.14 During the interviews, the participants noted that they found the individual items in the scale easy to understand; however, in the event of any ambiguity, the interviewer provided additional clarification, as recommended in the structured interview guide for the original HAMD
The average time taken to complete the Arabic HAMD-7 scale was approximately 4 minutes, although administration of the scale in a clinical setting may take more or less time depending on whether the scale is incorporated as part of a psychiatric interview. In the present study, neither the mean HAMD-7 nor PHQ-9 scores of the two samples differed significantly. Females, however, scored higher than males; this finding may be a reflection of the higher incidence of depressive disorders in the female population. In addition, subjects of the first student sample scored slightly higher in the first interview, possibly because the students had undertaken an examination very shortly beforehand.

Overall, the current study found that the Arabic HAMD-7 scale had acceptable internal consistency, with Cronbach’s α values ranging between 0.607—0.756. However, this does not rule out the possibility that the scale measures a different, albeit related, construct to depression. Moreover, these values likely do not represent a true measure of the internal consistency of the scale as both samples consisted of individuals without clinically-diagnosed depression. Hence, in order to more accurately measure internal consistency, the scale should be tested among a sample of depressed patients. The present study also revealed good convergent validity between the Arabic HAMD-7 and PHQ-9 scales. This indicates that, despite the different depressive symptoms included in each scale, both the HAMD-7 and PHQ-9 scales measure similar constructs, with all items of the HAMD-7 scale correlating with the PHQ-9 scores. Finally, an analysis of the test-retest reliability of the Arabic HAMD-7 scale showed that the test was reliable over a 14-day period, despite being administered at different times and by different investigators.

The current study was subject to certain limitations. Due to difficulties in testing the newly translated scale on patients diagnosed with depression, two community samples of college students and members of the general population, respectively, were utilised. This limits the generalisability of the results, especially as the HAMD-7 scale is not intended to be a diagnostic tool, but instead measure a severity of certain depressive symptoms among patients already diagnosed with clinical depression. Further studies are therefore recommended to fully explore the psychometric properties of the Arabic version of the scale by testing it on patients diagnosed with clinical depression. In addition, future research should

<table>
<thead>
<tr>
<th>Item</th>
<th>Scale mean*</th>
<th>Scale variance*</th>
<th>ITC</th>
<th>Cronbach’s α*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressed mood</td>
<td>3.14</td>
<td>6.848</td>
<td>0.517</td>
<td>0.541</td>
</tr>
<tr>
<td>Feelings of guilt</td>
<td>3.75</td>
<td>8.941</td>
<td>0.307</td>
<td>0.618</td>
</tr>
<tr>
<td>Interest, pleasure and level of activity</td>
<td>3.76</td>
<td>8.316</td>
<td>0.365</td>
<td>0.601</td>
</tr>
<tr>
<td>Psychological anxiety</td>
<td>3.28</td>
<td>7.901</td>
<td>0.322</td>
<td>0.620</td>
</tr>
<tr>
<td>Somatic anxiety</td>
<td>3.49</td>
<td>8.173</td>
<td>0.310</td>
<td>0.621</td>
</tr>
<tr>
<td>Energy level</td>
<td>3.80</td>
<td>8.939</td>
<td>0.452</td>
<td>0.590</td>
</tr>
<tr>
<td>Suicidality</td>
<td>4.08</td>
<td>9.757</td>
<td>0.277</td>
<td>0.629</td>
</tr>
</tbody>
</table>

*ITC = item-total correlation.

*After elimination of the item in question from the scale.
be conducted to determine the efficacy of the Arabic HAMD-7 scale in assessing the response to treatment with antidepressants.

Conclusion

In this study, an Arabic version of the interview-based HAMD-7 scale was validated among two community-based samples of Arabic speakers in Riyadh. The scale showed good convergent validity when compared to a previously validated Arabic version of the self-reported PHQ-9 scale. In addition, it also showed acceptable test-retest reliability over a period of two weeks, even when the scale was completed by different interviewers. Overall, the translated Arabic HAMD-7 scale may be a potentially useful clinical tool for clinicians and researchers seeking to assess depression severity among Arab patients. However, further research is required to assess the validity of the scale among patients diagnosed with depression.

ACKNOWLEDGEMENTS

The authors acknowledge the efforts of Abdulrahman Bahhari, Amani Al-Rafei, Bodoor Al-Enazi, Dalal Al-Roais, Haifa Al-Katheri, Hannin Khurmi, Manal Al-Anzi, Mohamad Al-Dkheel, Nujud Al-Otaibi, Omar Al-Bader, Rawand Al-Otaibi, Reem Al-Shehri, Reemah Al-Aifan, Sara Al-Bakr and Sultan Al-Otaibi, whose work as trained interviewers was indispensable in the data collection process.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This research was financially supported by the SABIC Psychological Health Research and Applications Chair of King Saud University.

References