anales de psicología / annals of psychology
2023, vol. 39, nº 3 (october), 395-404
https://doi.org/10.6018/analesps.550551

ISSN print: 0212-9728. ISSN online: 1695-2294.
https://revistas.um.es/analesps

Rasch analysis of the Hospital Anxiety and Depression Scale (HADS) in patients with major depression

Xiao-Jie Huang[1], Hai-Yan Ma[2], Xue-Mei Wang[3], Jing Zhong[1], Yu-Qi Chen[1], and Ming-Zhi Xu[1,*]

[1] *Guangdong Mental Health Center, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, Guangdong, People's Republic of China*
[2] *Guangdong Second Provincial General Hospital, Guangzhou, Guangdong, People's Republic of China*
[3] *Guangdong Provincial Hospital of Traditional Chinese Medicine, Guangdong Provincial Academy of Chinese Medical Sciences, and The Second Clinical College, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, People's Republic of China*

Título: Análisis de Rasch de la Escala Hospitalaria de Ansiedad y Depresión (Hospital Anxiety and Depression Scale, HADS) en pacientes con depresión mayor.
Resumen: La Hospital Anxiety and Depression Scale (HADS) [Escala Hospitalaria de Ansiedad y Depresión] es una herramienta ampliamente utilizada para detección de la depresión y la ansiedad en pacientes con enfermedades médicas. Este estudio tuvo como objetivo explorar las propiedades psicométricas de la HADS en la detección de pacientes con depresión mayor utilizando el método de la teoría de respuesta al ítem. Un total de 460 pacientes con depresión mayor completaron el HADS. Se utilizó el análisis de Rasch para examinar la unidimensionalidad, el ajuste de los ítems, la dependencia local, la confiabilidad, el orden de las categorías, el funcionamiento diferencial de los ítems (DIF) y la focalización. La HADS mostró una construcción bidimensional. Todos los ítems se ajustaban al modelo de Rasch. Tres pares de ítems mostraron una dependencia local menor pero desconsiderada. Ambas subescalas tuvieron una confiabilidad aceptable. Ninguno de los ítems mostró categorías desordenadas o DIF. Todos los ítems estaban bien dirigidos y los participantes con niveles altos y bajos de angustia fueron menos objetivo que aquellos con niveles moderados de angustia. Finalmente, se generó una tabla de conversión para transformar las puntuaciones brutas en medidas de intervalo. El HADS demostró propiedades psicométricas adecuadas para evaluar la depresión y la ansiedad en pacientes con depresión mayor. Fue más apropiado para evaluar niveles de angustia moderados que altos o bajos. La tabla de conversión se puede utilizar para una medición más precisa. Estos resultados pueden allanar el camino para métodos eficientes y sensibles para analizar la respuesta a los síntomas de depresión en la investigación y en la práctica clínica.
Palabras clave: Salud mental. Depresión mayor. HADS. Psicometría. Análisis de Rasch

Abstract: The Hospital Anxiety and Depression Scale (HADS) is a widely used screening tool for depression and anxiety in patients with medical illnesses. This study aimed to explore the psychometric properties of the HADS in screening for patients with major depression using item response theory method. A total of 460 patients with major depression completed the HADS. Rasch analyses were used to examine unidimensionality, item fit, local dependency, reliability, ordering of categories, differential item functioning (DIF) and targeting. The HADS showed a two-dimensional construct. All items fit the Rasch model. Three pairs of items showed minor but inconsiderate local dependency. Both subscales had acceptable reliability. None of the items displayed disordered categories or DIF. All items were well targeted, and participants with high and low levels of distress were less targeted than those with moderate levels of distress. Finally, a conversion table to transform the raw scores into interval measures was generated. The HADS demonstrated adequate psychometric properties in assessing depression and anxiety in patients with major depression. It was more appropriate for assessing moderate than high or low levels of distress. The conversion table can be used for more precise measurement. These results may pave the way for efficient and sensitive methods of analyzing depression symptom response in research and in clinical practice.
Keywords: Mental health. Major depression. HADS. Psychometrics. Rasch analysis.

Introduction

Major depression is a common mental disorder that severely impairs psychosocial functioning and diminishes quality of life. In China, the 12-month prevalence of depressive disorders among adults is 3.6%, and that of major depression is 2.1% (Lu et al., 2021). Symptoms of depression vary and involve the emotional, somatic, cognitive, and behavior domains. Yet none of the symptoms are pathognomonic. Some symptoms, such as emotional symptoms like depressed mood and anhedonia, are more specific to depression. Other symptoms, such as somatic symptoms like fatigue and insomnia, are also prevalent in many physical diseases (Malhi & Mann, 2018). This may leads to challenges in rating depressive symptoms, which overlap with the symptoms of other physical diseases. Similarly, individuals with depression often have features of anxiety. Approximately two-thirds of patients diagnosed with major depression have clinical anxiety (Malhi & Mann, 2018). The Diagnostic and Statistical Manual of Mental Disorders Fifth Edition (DSM-5) criteria describes anxious depression (comorbid anxiety and depression) as an anxious distress specifier in major depressive disorder section (American Psychiatric Association, 2013). Therefore, rating anxiety symptoms in patients with major depression is practical and valuable.

There is convincing evidence that self-assessment tools contribute to accurate diagnoses, effective treatment, and appropriate follow-up in clinical practice. Commonly used tools for depression include the Patient Health Questionnaire (PHQ), Back Depression Inventory (BDI) and the Hospital Anxiety and Depression Scale (HADS) for adults

*** Correspondence address [Dirección para correspondencia]:**
Ming-Zhi Xu. Guangdong Mental Health Center, Guangdong Provincial People's Hospital Huifu Branch, No. 123 Huifu Road West, Guangzhou, 510120, Guangdong, People's Republic of China.
E-mail: xumingzhi@gdph.org.cn
(Article received: 09-12-2022; revised: 09-01-2023; accepted: 26-02-2023)

(Siu et al., 2016). The HADS was initially designed to screen for depression and anxiety for patients coping with a physical illness (Zigmond & Snaith, 1983). To avoid confounding physical disorder with mood disorder, the HADS does not contain somatic symptoms. Therefore, compared to other popularly used depression scales, the HADS may have advantages in assessing cognitive processes associated with depressive symptoms and anxiety (Beekman & Verhagen, 2018).

Numerous studies have explored the ability of the HADS to screen for anxiety and depression. Some meta-analyses have reported that the HADS demonstrates optimal accuracy in screening for major depression (Vodermaier & Millman, 2011; Thombs et al., 2016; Wu et al., 2021). Notably, most of these studies were based on populations dealing with medical illnesses, such as cancer and strokes. Nevertheless, some studies have suggested the HADS can be a useful tool for assessing people with depression (Dagnan et al., 2000; Hung et al., 2011); moreover, some of them (Friedman et al., 2001; Flint & Rifat, 2002) explored the factor structure of the HADS in populations experiencing major depression. To the best of our knowledge, no study has examined the performance of the HADS in a single sample of patients with major depression. For example, how each item performs in differentiating patients in different distress levels, whether the category settings are reasonable, or whether the responses vary across different subgroups have not been adequately explored. This is possibly because of a limitation in measurement used in these studies, a traditional psychometric paradigm known as the Classical Test Theory (CTT).

Although CTT has been a dominant paradigm for several decades, it has some weaknesses that obstruct the objectivity of measurement (Petrillo et al., 2015). In CTT, data are based on ordered counts, which makes it incapable of supporting mathematical calculations. Applying ordinal scores on sophisticated parametric analyses could lead to a misinference of the findings (da Rocha et al., 2013). Rasch analysis is considered a modern psychometric paradigm that has been utilized to supplement the CTT. It provides certain methods to obtain scientific measurements (Tennant & Conaghan, 2007): (a) separately estimating the person and item measure and comparing them in an interval scale; (b) assessing the internal construct validity of the scale for unidimensionality; (c) testing the invariance of items and respondents; and (d) examining the appropriate category settings. Rasch analysis has been used for exploring psychometric properties of existing instruments (Zehirlioglu & Mert, 2019; Vindbjerg et al., 2020). Furthermore, although Rasch analysis has also been used in other studies to assess the HADS, these studies were based on populations with physical illnesses (Smith et al., 2006; Tang et al., 2007; Pallant & Tennant, 2007; Lin & Pakpour, 2017).

This study aimed to explore the psychometric properties of the HADS and its anxiety and depression subscales in a cohort of Chinese patients with major depression by applying the Rasch model. We explored the dimensional structure, item fit, reliability, category scheme, item–person targeting, and the differences in subgroups like gender, age, and education.

Materials and Methods

Participants

Data were collected from a mental health center in China from 2017 to 2021. All participants were adults with a diagnosis of major depressive disorder. They had the ability and willingness to provide informed consent to participate in the study; accordingly, patients meeting DSM-5 criteria for other severe or complex mental disorders (e.g., schizoaffective disorder and schizophrenia) and those with difficulty in comprehending or answering the questionnaire were excluded. The principles from the Declaration of Helsinki were followed when conducting this study.

Measures

The Hospital Anxiety and Depression Scale (HADS)

The HADS comprises 14 items equally divided into two subscales: anxiety (HADS-A) and depression (HADS-D). Respondents evaluate how they feel in the recent month; their responses are rated on a 4-point Likert scale ranging from 0 to 3, with higher scores indicating a higher severity of distress. The ratings are summed for each subscale (ranging from 0 to 21 in 7 items). The score of >11 indicates a high probability of suffering from a mood disorder, and the score of <8 indicates a low probability (Zigmond & Snaith, 1983). The reliability and validity of the Chinese version of the HADS has been established (Zheng et al., 2003).

Statistical Analysis

The demographics of the participants were described using IBM SPSS 25.0. The Rasch measurement was conducted using the Winsteps 5.1.7 program (Linacre, 2020).

Unidimensionality is a fundamental assumption of the Rasch model that indicates the items summed together form a single latent dimension,which contributes to the verification of the construct validity and helps validate the interpretation of the total scores (Kendel et al., 2010). The Winsteps program provides a principal components analysis (PCA) of the item residuals to examine unidimensionality. The eigenvalue of unexplained variance (the variance in the residuals) in the first contrast (factor) being smaller than 2.00 is not a reason for concern (Raîche, 2005). This analysis also provides the residual correlation between items and the first contrast using the positive and negative correlation to define two subsets of items. Unidimensionality holds true as long as the item responses are affected in the same manner (Tennant & Conaghan, 2007; Omara et al., 2019).

Local independency is another fundamental assumption that indicates each response is imagined to be independent among items; in other words, a response to one item should not influence a response to other items (Bond, 2015). Locally dependent items may violate unidimensionality and threaten estimations of reliability and validity (Omara et al., 2019). Thus, a much-used criterion to identify locally dependent items is a residual correlation of the paired items that exceeds 0.20 or 0.30 (Christensen et al., 2016).

Item fit summarizes the difference between observed responses and those expected by the Rasch model to an item across all participants. Items with such significant differences are considered to be misfits, which can potentially disrupt the unidimensional constructs and contribute to error when estimating person measures (Bond, 2015). The Winsteps program provides infit and outfit statistics. The infit is an information-weighted indicator that gives greater weight to responses to items with the measure located closer to the person; in contrast, the outfit is not weighted and therefore remains relatively more sensitive to the influence of outlying scores (Tennant & Conaghan, 2007). Both the infit and outfit mean square (MnSq) values are suggested to range from 0.5 to 1.5 for a sufficient fit (Wright & Linacre, 1994). Finally, the point-measure correlation demonstrates a correlation between observed responses and corresponding expected responses to an item, where the coefficient larger than 0.3 is considered acceptable (Linacre, 2020).

The person separation index (PSI) demonstrates how much we can rely on an order of person estimates to be replicated for a set of items; moreover, the item separation index (ISI) portrays how reproducible an item's difficulty is ordered for the sample. Finally, Cronbach's alpha is used to reflect the internal consistency and is required to exceed 0.80 (Linacre, 2020).

The HADS has a rating structure with responses to each item ranging from 0 to 3, which can be evaluated based on three criteria (Linacre, 1999). First, the categories should be ordered properly; in other words, the logit measure of categories should follow an increasing level of the latent trait. Therefore, category disordering is a big problem that contradicts the Rasch axiom that a higher measure presents a higher score on the rating scale. Second, the outfit MNSQ value for each rating category is less than 2.00 (Linacre, 2020).

Population invariance is an important requirement of the Rasch measurement. Differential item functioning (DIF) may exist when subgroups of individuals differ in how they respond to an item despite their equal levels of a latent trait (Rodriguez et al., 2019). Assessing DIF may ascertain whether the difference observed reflects a form of item bias in the instrument or a case of genuine diversity in prevalence (Rouquette et al., 2016). DIF contrast refers to the difference between the difficulty measures of an item for two specific subgroups. As such, the DIF with a contrast size above 0.5 logits is considered substantial (Gothwal et al., 2014). In this study, DIF analyses were performed for gender, age, and education (median split).

An item-person map is used to plot person ability (distress) along the item difficulty (severity) continuum on an interval scale. This scale is centered on zero logit, representing the item of average difficulty for the scale (Bond, 2015). Persons with higher value have higher levels of latent traits, and items with higher value are more difficult (Linacre, 2020). Well-targeted items demonstrate that items are appropriate for these participants, where the mean of person ability locates around zero logit (it is generally considered good to be within one logit). Therefore, well-targeted items provide more precise estimates of person ability than poorly targeted items and vice versa (Tennant & Conaghan, 2007; Bond, 2015).

A conversion table can be used to transform raw scores into the logit-based interval scales. Any given raw total score (ordinal-level data) could be converted to a corresponding person measure (interval-level data), which may increase accuracy of assessment.

Results

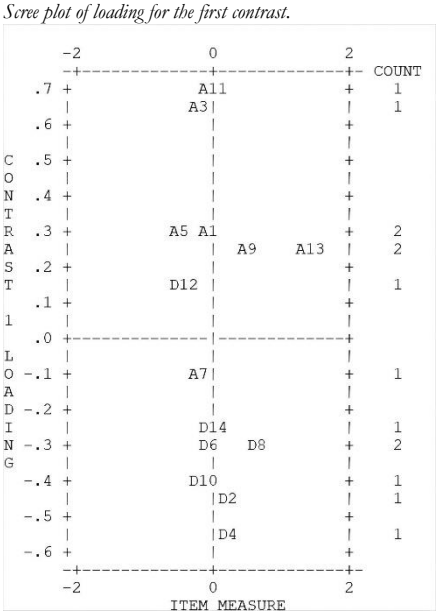
Demographic Information

Patients in different stages of the MDD course, whether they had received antidepressant treatments, were enrolled to gain a wide range of distress severity in order to increase the stability and precision of item estimates. In total, 481 patients completed the self-rating questionnaires; 21 were excluded due to missing values in the HADS. The final study sample consisted of 460 patients, of whom 67.0% were women and 37% were men. Participants were aged between 18 and 63 years (mean age 28.5 ± 9.0 years), and the majority (55.7%) were aged 18-27 years. Years of formal education ranged from 9 to 23 years (mean 14.3 ± 2.5 years), and most participants (60.4%) were educated for 13-16 years, which corresponded to an undergraduate education. For the HADS-D (mean raw score 10.4 ± 4.7), 230 (50.0%) patients were experiencing significant depression symptoms and 108 (23.5%) were not. For the HADS-A (mean raw score 10.1 ± 4.6), 235 (51.1%) patients were experiencing significant anxiety symptoms and 139 (30.2%) were not.

Unidimensionality

The PCA reported the eigenvalue of the first contrast of the HADS was 2.20, which suggested a potential subdimension. Figure 1 depicts a plot where the x-axis refers to the difficulty of items and the y-axis refers to the loading (correlation of individual items for the first contrast). Table 1 displays the values of loading. The items were divided into two subsets by the positive and negative loadings. This distribution was nearly consistent with that of the HADS-A and HADS-D subscales—except for item 7 and 12. The two subscales were examined to be unidimensional with the eigenvalues of 1.62 and 1.44, respectively.

Figure 1
Scree plot of loading for the first contrast.



Note: Items above the zero horizontal line indicate positive loading and items below indicate negative loading; The A items belong to anxiety subscale and The D items belong to depression subscale.

Table 1
Description, PCA loading and difficulty and local dependency of individual items.

Item description	PCA loading	Difficulty (logit)	Local dependency (paired items)
HADS-A			
1. I feel tense or wound up	0.31	-0.35	1&5;
3. I get a sort of frightened feeling as if something awful is about to happen	0.64	-0.21	3&11
5. Worrying thoughts go through my mind	0.32	-0.78	
7. I can sit as ease and feel relaxed	-0.10	-0.46	
9. I feel restless as if I have to be on the move	0.26	0.52	
11. I get sudden feelings of panic	0.70	-0.05	
13. I get a sort of frightened feeling like "butterfly" in the stomach	0.23	1.33	
HADS-D			
2. I still enjoy the things I used to enjoy	-0.45	0.19	6&12
4. I can laugh and see the funny side of things	-0.57	0.13	
6. I feel cheerful	-0.32	-0.13	
8. I have lost interest in my appearance	-0.31	0.51	
10. Look forward with enjoyment to things	-0.42	-0.10	
12. I feel as if I am slowed down	0.13	-0.53	
14. I can enjoy a good book or TV program	-0.25	-0.07	

Abbreviations: PCA, principal component analysis; HADS-A, anxiety subscale; HADS-D, depression subscale.

Local Dependency

There were three pairs of items showed evidence of local dependency: item 1 and 5 (r = .21), item 3 and 11 (r = .28), item 6 and 12 (r = .20). However, these correlation coefficients did not exceed the .30 threshold and were therefore insignificant.

Item Fit

As shown in Table 2, all items of HADS-A/HADS-D fulfilled the criteria for model fit with an infit/outfit MnSq between 0.66 and 1.49, as well as the point-measure correlations of greater than 0.55. This indicated that the HADS-A/HADS-D items contributed to a unidimensional construct and provided precision for estimating person measure.

Table 2
Item fit and differential item functioning statistics of individual items.

Item description	Item fit			DIF contrast (logit)		
	INFIT MnSq	OUFIT MnSq	r_{PM}	Gender	Age	Education
HADS-A						
1. I feel tense or wound up	0.66	0.71	0.74	0.00	0.00	0.23
3. I get a sort of frightened feeling as if something awful is about to happen	0.92	0.91	0.74	-0.06	0.10	-0.07
5. Worrying thoughts go through my mind	0.66	0.71	0.75	-0.16	-0.43	0.19
7. I can sit as ease and feel relaxed	0.83	0.89	0.68	0.24	0.21	0.00
9. I feel restless as if I have to be on the move	1.30	1.34	0.60	0.23	0.05	-0.13
11. I get sudden feelings of panic	0.87	0.84	0.76	-0.03	-0.09	-0.13
13. I get a sort of frightened feeling like "butterfly" in the stomach	1.28	1.33	0.55	0.28	0.00	-0.40
HADS-D						
2. I still enjoy the things I used to enjoy	0.94	0.95	0.67	-0.28	0.05	0.00
4. I can laugh and see the funny side of things	1.02	1.02	0.65	0.44	0.26	0.15
6. I feel cheerful	0.75	0.81	0.70	0.33	0.32	-0.08
8. I have lost interest in my appearance	1.25	1.24	0.60	0.17	0.00	0.30
10. Look forward with enjoyment to things	0.86	0.86	0.71	-0.06	-0.19	0.22
12. I feel as if I am slowed down	0.69	0.74	0.73	0.17	0.00	0.00
14. I can enjoy a good book or TV program	1.49	1.44	0.64	0.07	0.08	-0.12

Abbreviations: INFIT MnSq, mean square of information-weighted fit statistics; OUTFIT MnSq, mean square of outlier-sensitive fit statistics; r_{PM}, point-measure correlation; DIF, Differential Item Functioning; HADS-A, anxiety subscale; HADS-D, depression subscale.
Note: DIF contrast (Gender) = DIF measure (female) -DIF measure (male); DIF contrast (age) = DIF measure (≤ 25 years old) -DIF measure (>25 years old); DIF contrast (education) = DIF measure (≤15 years) -DIF measure (>15 years).

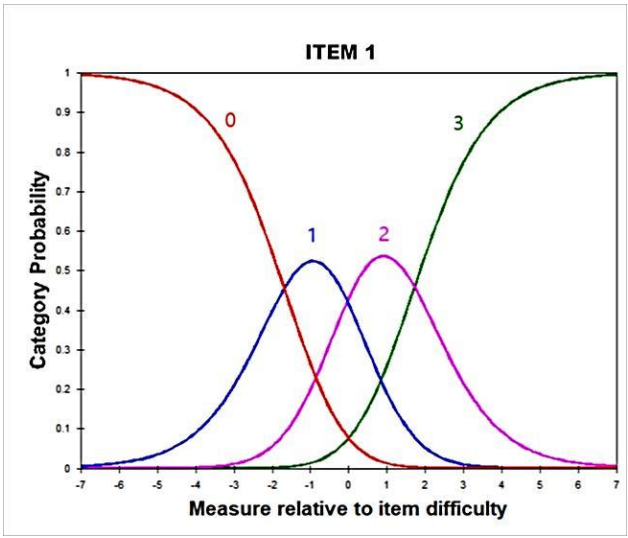
Reliability

The internal consistency of the HADS-A and HADS-D was generally sufficient with the Cronbach's alpha of 0.83 and 0.81, respectively. The ISI were 8.98 and 4.23, respectively, thereby indicating a good item reliability of either subscale. The PSI were 1.96 and 1.94, respectively, which were close to the criterion. Furthermore, the Winsteps program also reported the "MODEL estimates." Notably, the MODEL estimates were computed on the basis that the data fit the model, and all misfits in the data were merely a reflection of the stochastic nature of the model, which presented the best-case situation (Linacre, 2020). The MODEL PSI of the two subscales reached 2.20 and 2.13, respectively.

Ordering of Response Categories

Each of the rating categories had an outfit MNSQ value of less than 2.00. Figure 2 depicts the category probability curves of item 1 as an example. The x-axis represents the item difficulty, and the y-axis represents the probability of endorsing the response categories to this item. The different colored curves respectively corresponded with the 0–3 response categories. These curves were located in a monotonic pattern from left to right as their related ratings increased with item difficulty, thereby demonstrating the properly ordered categories and thresholds of item 1. We examined the category probability curves for the rest of the items and none of them showed disordering.

Figure 2
Category probability curves of item1.



Note: The red, blue, pink and green curves on the graph represent the 0, 1, 2 and 3 rating categories respectively.

Differential Item Functioning (DIF)

The DIF analyses were performed for gender (female vs. male), age (≤ 25 vs. > 25 years old), and education (≤ 15 vs. > 15 years). We found no significant DIF regarding gender, age, or education, which provided evidence of adequate population invariance for the items. Table 2 shows the values of DIF contrast.

Targeting

The item–person map of either the depression or anxiety subscale showed the distribution of person ability (distress) was nearly normal, and most patients showed a moderate level of distress. Figure 3(A) showed the HADS-A items clustered on the middle (moderate) region and were well-targeted by person ability (mean = -0.17 ± 1.52 logits). Consequently, patients with high and low levels of anxiety were poorly targeted, indicating that the items provided less precise estimates for patients with high and low levels of anxiety than those with moderate levels of anxiety. The item difficulty (severity) of HADS-A ranged -0.78–1.33 logits, and the item 13 and 5 were the most and least difficult. Similarly, Figure 3(B) shows the HADS-D items were targeted by patients with a moderate level of depression (mean = -0.02 ± 1.45 logits). The item difficulty of HADS-D ranged -0.53–0.51 logits, and the item 12 and 8 were the most and least difficult to endorse for these participants.

Transformation to Interval Scales

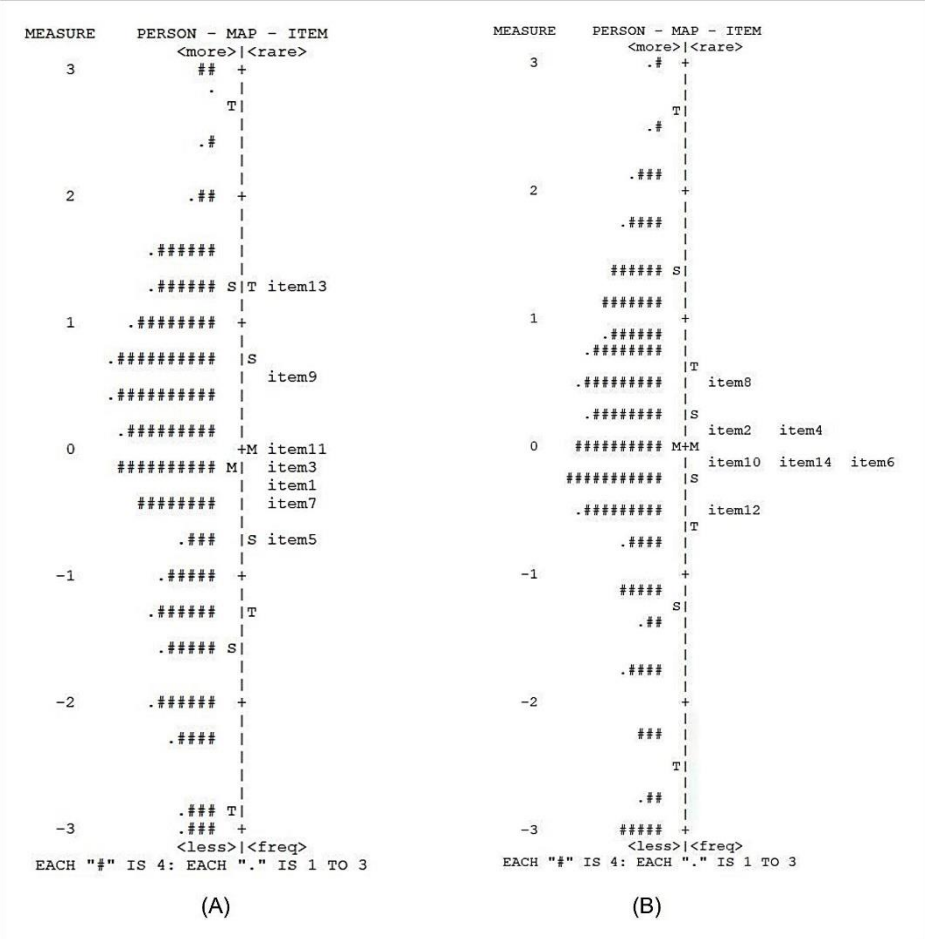
We transformed the total raw scores of the HADS-A/HADS-D subscales into interval scales respectively based on person estimates of the Rasch model (Table 3).

Table 3
Conversion table of the raw scores and interval measures.

HADS-A			HADS-D		
Raw score	Measure	SE	Raw score	Measure	SE
0	-4.97	1.86	0	-4.85	1.86
1	-3.69	1.06	1	-3.56	1.06
2	-2.88	0.79	2	-2.74	0.79
3	-2.36	0.68	3	-2.20	0.69
4	-1.94	0.62	4	-1.78	0.63
5	-1.59	0.58	5	-1.41	0.59
6	-1.27	0.55	6	-1.08	0.56
7	-0.97	0.54	7	-0.78	0.54
8	-0.69	0.53	8	-0.50	0.52
9	-0.42	0.52	9	-0.25	0.50
10	-0.15	0.52	10	0.00	0.49
11	0.11	0.52	11	0.23	0.48
12	0.38	0.52	12	0.46	0.48
13	0.66	0.53	13	0.69	0.48
14	0.94	0.54	14	0.92	0.49
15	1.25	0.56	15	1.16	0.50
16	1.57	0.59	16	1.43	0.52
17	1.94	0.63	17	1.72	0.56
18	2.37	0.69	18	2.07	0.63
19	2.92	0.80	19	2.53	0.74
20	3.76	1.07	20	3.27	1.02
21	5.08	1.87	21	4.50	1.84

Abbreviations: HADS-A, anxiety subscale; HADS-D, depression subscale; SE, standard error.

Figure 3
Item-person map of the (A) HADS-A, (B) HADS-D.



Note: In each map, person measures are depicted on the left side and item measures are on the right side. More distressed persons/more difficult items locate at the top of the map and less distressed persons/less difficult items locate at the bottom. Each "#" represents 4 persons and each "." represents 1~3 persons.
Abbreviations: M, mean; S, standard deviation; T, two standard deviations; HADS-A, anxiety subscale; HADS-D, depression subscale.

Discussion

Main Findings

This study aimed to explore the psychometric properties of the HADS for screening patients with major depression via the Rasch measurement model. We subsequently investigated unidimensionality, local independence, item fit, reliability, ordering of the category threshold, DIF, and targeting. Finally, we generated a conversion table to transform the raw scores into an interval scale for both subscales.

The HADS displayed a potentially two-dimensional construct with items explicitly related to anxiety and depression respectively. The exceptions were item 7 *(I can sit as ease and feel relaxed)*, which was related to anxiety but fell on the depression dimension, and item 12 *(I feel as if I am slowed down)* was on the contrary that related to depression but fell on the anxiety dimension. This was possibly because these two items were loaded too small to be apparently distributed to either dimension. Thus, we subsequently verified the unidimensional construct of the HADS-A/HADS-D. Some previous studies have suggested a similar two-dimensional construct of the HADS (Norton et al., 2013; Jerković et al., 2021), whereas others have reported unidimensional (Lin & Pakpour, 2017) or multidimensional constructs (Schouten et al., 2020).

All items displayed an adequate fit to the model. The following three pairs of items presented minor but inconsiderate dependency: item 1 *(I feel tense or wound up)* and item 5 *(Worrying thoughts go through my mind)*; item 3 *(I get a sort of frightened feeling as if something awful is about to happen)* and item 11 *(I get sudden feelings of panic)*; item 6 *(I feel cheerful, negatively-scored)* and item 12 *(I feel as if I am slowed down)*. This was possibly because the descriptions of the involved items were similar; for example, the opposite meaning of feeling cheerful was similar to feeling down. However, they were insufficient to

threaten the unidimensional construct of the subscales ($r <$ .30).

The item separation and Cronbach's alpha of the HADS-A/HADS-D were satisfactory; however, their person separation was insufficient (PSI ranged 1.94–1.96), which indicated that a wider difficulty range was needed on the instrument. The Winsteps program provided ideal estimates that indicated the person separation could be optimal if the data fit the model as a best-case situation (MODEL PSI ranged 2.13–2.20). Thus, we considered the reliability of the two subscales to be acceptable.

The Rasch rating scale model was used to examine the category structure of the subscales as they were polytomous scales, and the result showed the category settings were reasonable for HADS-A/HADS-D. DIF was explored for gender, age, and education and no significant DIF was found, suggesting that the HADS-A/HADS-D had a valid internal structure. This result was similar to that of Lin's (Lin & Pakpour, 2017) study, which reported this in patients across gender and types of epilepsy. However, some other studies reported the existence of DIF in some items (Pallant & Tennant, 2007; Cameron et al., 2013).

Targeting was analyzed by comparing person ability (distress) and item difficulty (severity) in a similar logit scale. The result indicated the HADS-A/HADS-D items were all well-targeted by corresponding person ability and obtained precise item estimates. Nevertheless, the items had a narrow range to target persons in high and low levels of distress, thereby indicating the items provided less precise estimates for high and low levels of distress than those for the moderate level. The HADS required better discriminating items in the segment of persons with high and low levels of distress. Moreover, the least and the most challenging items of HADS-A were item 5 *(Worrying thoughts go through my mind)* and 13 *(I get a sort of frightened feeling like a "butterfly" in the stomach)*. For HADS-D, these were item 12 *(I feel as if I am slowed down)* and 8 *(I have lost interest in my appearance)*. Kendel et al. (Kendel et al., 2010) reported a similar result for the HADS-D items. Moreover, Pallent and Tennant (Pallant & Tennant, 2007) also reported a similar item that was least challenging, but the most challenging item was item 6 *(I feel cheerful)*. As for the HADS-A, Lin et al. (Lin & Pakpour, 2017) suggested that item 13 *(I get a sort of frightened feeling like "butterfly" in the stomach)* was one of the most challenging items.

A conversion table was generated to transform raw scores to logit-based interval data. The raw scores are based on ordered counts, wherein distance between scores is unequal. For example, it can be said only that a score of 10 represents a higher level of distress than a 9, but this doesn't specify how high the stress is. When transforming to an interval measurement, we can posit that a score of 10 is 0.27 logits higher than a score of 9 in HADS-A. This result may provide more information for clinicians or researchers to effectively apply the HADS in practical settings. Moreover, it was noted that the conversion table was valuable as long as all the items were completed because the raw scores referred to the composite score across items.

Strengths and Limitations

As the HADS was initially developed for patients with physical illnesses within the general population, it has rarely been studied in populations with mental disorders. As such, this study was based on a sample of patients with depression to explore the psychometric properties of the HADS. Moreover, the Rasch analysis applied in this study is regarded as a modern test theory that is superior to the classical methods.

This study has some limitations. First, the participants were recruited came from a single mental health center. Studies that recruit participants from multiple centers should be conducted to examine the validity in other depression populations. Second, the participants were mostly young and well educated, as older adult and poorly educated samples were scarce. This may have caused bias in the measurement. Hence, further studies should consider stratified sampling into consideration.

Conclusions

The HADS presented satisfactory psychometric properties for assessing depression and anxiety in patients with major depression, thereby presenting a two-dimensional construct with items related to anxiety and depression. The two subscales displayed unidimensional and acceptable reliability. None of the items showed significant DIF or disordered categories. The HADS was more appropriate for persons with a moderate level of distress than those with high or low distress levels. The conversion table can be used for more precise measurements. Finally, further studies should consider stratified sampling.

Conflicts of interest.- The authors declare no potential conflicts of interest with respect to the authorship or publication of this paper.

Funding.- This work was supported by the Science and Technology Planning Project of Guangdong Province (Project NO: 2017A02015095).

Acknowledgements.- This work was completed in Guangdong Mental Health Center, Guangdong Provincial People's Hospital and the authors appreciate all the colleagues for their support and encouragement. Besides, the authors wish to express their gratitude to John Michael Linacre, Ph.D., the developer of Winsteps software for the operating counseling.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: American Psychiatric Association.

Beekman, E., & Verhagen, A. (2018). Clinimetrics: Hospital Anxiety and Depression Scale. *Journal of Physiotherapy*, *64*(3), 198. https://doi.org/10.1016/j.jphys.2018.04.003

Bond, T. (2015). *Applying the Rasch Model Fundamental Measurement in the Human Sciences, Third Edition*. Routledge.

Cameron, I. M., Crawford, J. R., Lawton, K., & Reid, I. C. (2013). Differential item functioning of the HADS and PHQ-9: An investigation of age, gender and educational background in a clinical UK primary care sample. *Journal of Affective Disorders*, *147*(1-3), 262–268. https://doi.org/10.1016/j.jad.2012.11.015

Christensen, K. B., Makransky, G., & Horton, M. (2016). Critical Values for Yen's Q_3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. *Applied Psychological Measurement*, *41*(3), 178–194. https://doi.org/10.1177/0146621616677520

da Rocha, N. S., Chachamovich, E., de Almeida Fleck, M. P., & Tennant, A. (2013). An introduction to Rasch analysis for Psychiatric practice and research. *Journal of Psychiatric Research*, *47*(2), 141–148. https://doi.org/10.1016/j.jpsychires.2012.09.014

Dagnan, D., Chadwick, P., & Trower, P. (2000). Psychometric properties of the Hospital Anxiety and Depression Scale with a population of members of a depression self-help group. *British Journal of Medical Psychology*, *73*(1), 129–137. https://doi.org/10.1348/000711200160255

Flint, A. J., & Rifat, S. L. (2002). Factor structure of the hospital anxiety and depression scale in older patients with major depression. *International Journal of Geriatric Psychiatry*, *17*(2), 117–123. https://doi.org/10.1002/gps.535

Friedman, S., Samuelian, J.-C., Lancrenon, S., Even, C., & Chiarelli, P. (2001). Three-dimensional structure of the Hospital Anxiety and Depression Scale in a large French primary care population suffering from major depression. *Psychiatry Research*, *104*(3), 247–257. https://doi.org/10.1016/s0165-1781(01)00309-2

Gothwal, V. K., Bagga, D. K., & Sumalini, R. (2014). Rasch validation of the PHQ-9 in people with visual impairment in South India. *Journal of Affective Disorders*, *167*, 171–177. https://doi.org/10.1016/j.jad.2014.06.019

Hung, Ching-I., Liu, C.-Y., Wang, S.-J., Yao, Y.-C., & Yang, C.-H. (2011). The cut-off points of the Depression and Somatic Symptoms Scale and the Hospital Anxiety and Depression Scale in detecting non-full remission and a current major depressive episode. *International Journal of Psychiatry in Clinical Practice*, *16*(1), 33–40. https://doi.org/10.3109/13651501.2011.617456

Jerković, A., Proroković, A., Matijaca, M., Vuko, J., Poljičanin, A., Mastelić, A., … Rogić Vidaković, M. (2021). Psychometric Properties of the HADS Measure of Anxiety and Depression Among Multiple Sclerosis Patients in Croatia. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.794353

Kendel, F., Wirtz, M., Dunkel, A., Lehmkuhl, E., Hetzer, R., & Regitz-Zagrosek, V. (2010). Screening for depression: Rasch analysis of the dimensional structure of the PHQ-9 and the HADS-D. *Journal of Affective Disorders*, *122*(3), 241–246. https://doi.org/10.1016/j.jad.2009.07.004

Lin, C.-Y., & Pakpour, A. H. (2017). Using Hospital Anxiety and Depression Scale (HADS) on patients with epilepsy: Confirmatory factor analysis and Rasch models. *Seizure*, *45*, 42–46. https://doi.org/10.1016/j.seizure.2016.11.019

Linacre JM.Investigating rating scale category utility. J Outcome Meas (1999) 3:103-122. https://pubmed.ncbi.nlm.nih.gov/10204322/

Linacre, J. (2020). *User's Guide to Winsteps*. Retrieved from https://www.winsteps.com

Lu, J., Xu, X., Huang, Y., Li, T., Ma, C., Xu, G., … Yan, J. (2021). Prevalence of depressive disorders and treatment in China: a cross-sectional epidemiological study. *The Lancet Psychiatry*, *8*(11), 981–990. https://doi.org/10.1016/s2215-0366(21)00251-0

Malhi, G. S., & Mann, J. J. (2018). Depression. *The Lancet*, *392*(10161), 2299–2312. https://doi.org/10.1016/s0140-6736(18)31948-2

Norton, S., Cosco, T., Doyle, F., Done, J., & Sacker, A. (2013). The Hospital Anxiety and Depression Scale: A meta confirmatory factor analysis. *Journal of Psychosomatic Research*, *74*(1), 74–81. https://doi.org/10.1016/j.jpsychores.2012.10.010

Omara, M., Stamm, T., Boecker, M., Ritschl, V., Mosor, E., Salzberger, T., … Bekes, K. (2019). Rasch model of the Child Perceptions Questionnaire for oral health–related quality of life. *The Journal of the American Dental Association*, *150*(5), 352-361.e7. https://doi.org/10.1016/j.adaj.2018.12.001

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, *46*(1), 1–18. https://doi.org/10.1348/014466506x96931

Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using Classical Test Theory, Item Response Theory, and Rasch Measurement Theory to Evaluate Patient-Reported Outcome Measures: A Comparison of Worked Examples. *Value in Health*, *18*(1), 25–34. https://doi.org/10.1016/j.jval.2014.10.005

Raîche, G. (2005). Critical eigenvalue sizes (variances) in standardized residual principal components analysis. *Rasch Meas Trans*,19,1012. Retrieved from https://www.rasch.org/rmt/rmt191h.htm

Rodriguez, V. J., Shaffer, A., Are, F., Madden, A., Jones, D. L., & Kumar, M. (2019). Identification of differential item functioning by race and ethnicity in the Childhood Trauma Questionnaire. *Child Abuse & Neglect*, *94*, 104030. https://doi.org/10.1016/j.chiabu.2019.104030

Rouquette, A., Hardouin, J.B., & Coste, J. (2016). Differential Item Functioning (DIF) and Subsequent Bias in Group Comparisons using a Composite Measurement Scale: A Simulation Study. *J Appl Meas*, 17:312–334. https://pubmed.ncbi.nlm.nih.gov/28027055/

Schouten, R. W., Nadort, E., van Ballegooijen, W., Loosman, W. L., Honig, A., Siegert, C. E. H., … Broekman, B. F. P. (2020). General distress and symptoms of anxiety and depression: A factor analysis in two cohorts of dialysis patients. *General Hospital Psychiatry*, *65*, 91–99. https://doi.org/10.1016/j.genhosppsych.2020.04.004

Siu, A. L., Bibbins-Domingo, K., Grossman, D. C., Baumann, L. C., Davidson, K. W., Ebell, M., … Pignone, M. P. (2016). Screening for Depression in Adults. *JAMA*, *315*(4), 380. https://doi.org/10.1001/jama.2015.18392

Smith, A. B., Wright, E. P., Rush, R., Stark, D. P., Velikova, G., & Selby, P. J. (2006). Rasch analysis of the dimensional structure of the Hospital Anxiety and Depression Scale. *Psycho-Oncology*, *15*(9), 817–827. https://doi.org/10.1002/pon.1015

Tang, W. K., Wong, E., Chiu, H. F. K., & Ungvari, G. S. (2007). Rasch analysis of the scoring scheme of the HADS Depression subscale in Chinese stroke patients. *Psychiatry Research*, *150*(1), 97–103. https://doi.org/10.1016/j.psychres.2006.01.015

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism*, *57*(8), 1358–1362. https://doi.org/10.1002/art.23108

Thombs, B. D., Benedetti, A., Kloda, L. A., Levis, B., Azar, M., & Riehm, K. E. (2016). Diagnostic accuracy of the Depression subscale of the Hospital Anxiety and Depression Scale (HADS-D) for detecting major depression: protocol for a systematic review and individual patient data meta-analyses. *BMJ Open*, *6*(4), e011913. https://doi.org/10.1136/bmjopen-2016-011913

Vindbjerg, E., Carlsson, J., Mortensen, E. L., Makransky, G., & Nielsen, T. (2020). A Rasch-based validity study of the Harvard Trauma Questionnaire. *Journal of Affective Disorders*, *277*, 697–705. https://doi.org/10.1016/j.jad.2020.08.071

Vodermaier, A., & Millman, R. D. (2011). Accuracy of the Hospital Anxiety and Depression Scale as a screening tool in cancer patients: a systematic review and meta-analysis. *Supportive Care in Cancer*, *19*(12), 1899–1908. https://doi.org/10.1007/s00520-011-1251-4

Wright, B.D., & Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch Meas Trans,* 370–371. Retrieved from https://www.rasch.org/rmt/rmt83b.htm

Wu, Y., Levis, B., Sun, Y., He, C., Krishnan, A., & Neupane, D. (2021). Accuracy of the Hospital Anxiety and Depression Scale Depression subscale (HADS-D) to screen for major depression: systematic review and individual participant data meta-analysis. *BMJ*, *373*, n972. https://doi.org/10.1136/bmj.n972

Zehirlioglu, L., & Mert, H. (2019). Validity and reliability of the Heart Disease Fact Questionnaire (HDFQ): a Rasch measurement model approach. *Primary Care Diabetes*, *14*(2). https://doi.org/10.1016/j.pcd.2019.06.006

Zheng, L., Wang, Y., & Li, H. (2003). Application of Hospital Anxiety and Depression Scale in general hospital: an analysis in validity and reliability. *Shanghai Archives of Psychiatry*, 15, 264-266. https://doi.org/10.3969/j.issn.1002-0829.2003.05.003

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x