Check for updates

Received: 6 June 2022 | Accepted: 22 January 2023

DOI: 10.1111/bjop.12637

the british psychological society
promoting excellence in psychology

ARTICLE

The structure of the Hospital Anxiety and Depression Scale: Theoretical and methodological considerations

Maddison Lloyd[1] | **Nicole Sugden**[1] | **Matt Thomas**[1,2,3] | **Andrew McGrath**[1] | **Clive Skilbeck**[4]

[1]School of Psychology, Charles Sturt University, Bathurst, New South Wales, Australia

[2]Western New South Wales Local Health District, Bathurst, New South Wales, Australia

[3]Marathon Health, Bathurst, New South Wales, Australia

[4]School of Psychology, University of Tasmania, Sandy Bay, Tasmania, Australia

Correspondence
Nicole Sugden, Panorama Avenue, Charles Sturt University, Bathurst, NSW, 2795, Australia.
Email: nisugden@csu.edu.au

Abstract

The Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983) is widely used; however, its factor structure is unclear, with studies reporting differing unidimensional, two-factor and three-factor models. We aimed to address some key theoretical and methodological issues contributing to inconsistencies in HADS structures across samples. We reviewed existing HADS models and compared their fit using confirmatory factor analysis (CFA). We also investigated methodological effects by comparing factor structures derived from Rasch and Principal Components Analysis (PCA) methods, as well as effects of a negative wording factor. An Australian community-dwelling sample consisting of 189 females and 158 males aged 17–86 (M = 35.73, SD = 17.41) completed the 14-item HADS. The Rasch Analysis, PCA and CFA all supported the original two-factor structure. Although some three-factor models had good fit, they had unacceptable reliability. In the CFA, a hierarchical bifactor model with a general distress factor and uncorrelated depression and anxiety subscales produced the best fit, but the general factor was not unidimensional. The addition of a negative wording factor improved model fit. These findings highlight the effects of differing methodologies in producing inconsistent HADS factor structures across studies. Further replication of model fit across samples and refinement of the HADS items is warranted.

KEYWORDS
anxiety, depression, factor analysis, Hospital Anxiety and Depression Scale, psychological assessment

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *British Journal of Psychology* published by John Wiley & Sons Ltd on behalf of The British Psychological Society.

Statement of contribution

What is already known on this subject?

- The Hospital Anxiety and Depression Scale (HADS) is widely used to assess anxiety and depression symptoms
- Various methodologies have been used to determine the factor structure of the HADS, including Rasch analysis, principal component analysis and confirmatory factor analysis, with different factor structures identified across and within factor extraction methods
- Differing factor structures have been obtained across clinical and community samples, with various one-, two- and three-factor models identified

What this study adds?

- A contemporary review of existing HADS factor structures and methodologies used in previous studies
- Use of robust estimation methods to overcome methodological limitations of previous factor analytic studies
- In-depth analysis of hierarchical bifactor models of the HADS and the suitability of a general psychological distress factor
- Guidelines for future clinical use of the HADS, examining the HADS latent structure and for refining the HADS items

BACKGROUND

Major depressive disorder and anxiety disorders were estimated in 2020 to have a global prevalence of 246 million and 374 people million respectively (Covid-19 Mental Disorders Collaborators, 2021). In 2019, these disorders were ranked 13 and 24 of 369 in the list of leading causes of global burden of diseases and injuries (GBD 2019 Diseases and Injuries Collaborators, 2020). It was predicted that the global costs associated with mental health would be over US$16 trillion during 2011–2030 (Trautmann et al., 2016). The World Health Organization Comprehensive Mental Health Action Plan 2013–2030 (World Health Organisation, 2021) aims to reduce the burden of mental health disorders through prevention, promotion of well-being and enhancing treatment. To achieve these aims, proper assessment of mental health conditions is vital so that appropriate interventions can be delivered and monitored. When assessing anxiety and depression symptoms, there is a large list of self-report measures that can be utilized by clinicians (Sunderland et al., 2019). The choice of which measure to use depends on the scale's clinical and psychometric suitability for the population of interest.

The Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983) is a widely used measure of depression and anxiety (Norton et al., 2013). The scale is popular due to its brevity of 14 items; being widely translated and having good reliability, specificity and sensitivity and concurrent validity in a variety of samples (Bjelland et al., 2002). As the comorbidity of depression and anxiety disorders is high (Kalin, 2020), the inclusion of separate seven-item anxiety and depression subscales makes the HADS clinically useful for identifying these often co-occurring conditions while facilitating the differentiation between symptoms within a single instrument (Bjelland et al., 2002).

The HADS anxiety subscale assesses generalized anxiety symptoms such as worry. The depression subscale focuses on anhedonia, or a loss of pleasure, which is argued to be the presentation most responsive to antidepressant medications and therefore clinically useful (Zigmond & Snaith, 1983). The HADS focuses on psychological states of anxiety and depression in an attempt to avoid being confounded by symptoms of associated physical illness (Zigmond & Snaith, 1983). This assessment of emotional states differs from other widely used scales such as the Depression Anxiety and Stress Scale (DASS), State Trait

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Anxiety Inventory, Beck Anxiety Inventory, Beck Depression Inventory-II and Hamilton Depression Rating Scale that measure somatic symptoms and suicidal ideation (Barth & Martin, 2005; Julian, 2011; Smarr & Keefer, 2011). The exclusion of suicidality and somatic symptoms makes it useful for the HADS to be used in nonpsychiatric settings (Schönberger & Ponsford, 2010).

The HADS has been widely used in clinical and nonclinical populations; however, multiple factor structures have been reported. Table 1 provides a summary of the various factor structures that have been identified in the literature to date. Factor structures have included the original two-factor model of anxiety and depression proposed by Zigmond and Snaith (1983), a unidimensional model where all items loaded on a single scale (Razavi et al., 1990) and various three-factor structures. Most of these three-factor models have comprised of anxiety and depression factors, with the third factor relating to either psychomotor (Friedman et al., 2001; Skilbeck et al., 2011) or restlessness symptoms (Brandberg et al., 1992; Caci et al., 2003).

Reviews of the factor structure of the HADS have found that there are inconsistencies in the latent structures obtained across samples. Bjelland et al. (2002) reviewed 19 studies and found that most studies (n = 11) supported a two-factor structure, while five reported a three-factor structure and two found a four-factor structure. In a more recent review of 50 studies, Cosco et al. (2012) also found that the majority of studies (n = 25) supported a two-factor structure. Saez-Flores et al. (2018) noted that most of the previous HADS studies had used community samples. They, therefore, reviewed 18 medical samples and found a three-factor structure had the best fit, suggesting that a three-factor structure may better account for the complexity of medical conditions such as traumatic brain injury (TBI), cystic fibrosis, cancer and cardiovascular disease.

Contributing to the inconsistencies discussed above is the finding that in many previous studies, models demonstrated misspecification. For example, Bjelland et al. (2002) reported that item 8 (a depression item) consistently had low loadings onto the depression scale. Similarly, Cosco et al. (2012) found item 7 (an anxiety item) to load anomalously in 20 of the 50 articles reviewed. Accordingly, Stott et al. (2017) reported improved model fit when items 7 and 8, which had the highest cross loadings and large standardized residual values, were removed. Thus, it remains unclear how many factors the HADS has, and which items should load onto these factors.

This uncertainty about item loadings and latent structure has implications for clinicians in using diagnostic cut-offs and for symptom monitoring, as recovery and deterioration rates for anxiety and depression may differ across clinical contexts (Saunders et al., 2019; Strauss & Smith, 2009). The lack of a clear HADS factor structure has led some to call for the abandonment of the use of the HADS (Coyne & van Sonderen, 2012). However, responding to this, Norton et al. (2012) argued that rather than abandoning the scale, the scale should be refined to eliminate inconsistencies in the factor structure of the HADS.

Norton et al. (2012) proposed that much of the inconsistency in the latent structure could be resolved by applying a hierarchical factor structure to the HADS. Hierarchical models have increasingly been used to explain the comorbidity of anxiety and depression (Watson, 2005). Such hierarchical models have produced good fit for measures such as the DASS-21 (Henry & Crawford, 2005). Dunbar et al. (2000) suggested a hierarchical HADS model based on the tripartite theory of anxiety and depression (Clark & Watson, 1991), where an overarching negative affectivity factor accounted for the overlap between anxiety and depression subscales. In this model, first-order factors of anxiety and depression loaded onto and mediated the higher order negative affectivity factor. However, Norton et al. (2013) criticized this model as it is difficult to delineate the effects of the higher order factor from those of the subfactors, making it unsuitable for clinical use. For this reason, this model is excluded from Table 1 and subsequent analyses in this study.

Following advancements in factor analytic methods and the introduction of bifactor models (Xie et al., 2012), Norton et al. (2013) proposed that a bifactor model may best explain the latent structure of the HADS. In bifactor models, items load onto uncorrelated grouping factors (i.e. anxiety, depression and restlessness) as well as a general distress factor that is at the same level, but uncorrelated with the grouping factors. Having orthogonal grouping and general factors allows for separate general distress and subscale scores to be calculated (Dunn & McCray, 2020). Norton et al.'s (2013) bifactor models included

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Factor structures identified in previous studies.

Study	Country (language)	Sample	Factor extraction method	Factor structure
One factor				
Razavi et al. (1990)	Belgium (French)	63 Cancer inpatients	Hotelling and Thurstone EFA using orthogonal rotation	Unidimensional: all items
Two-factors				
Zigmond and Snaith (1983)	UK (English)	Unspecified	Unspecified	Anxiety: 1, 3, 5, 7, 9, 11, 13 Depression: 2, 4, 6, 8, 10, 12, 14
Moorey et al. (1991)	UK (English)	568 cancer patients	PCA with oblique rotation	Anxiety: 1, 3, 5, 9, 11, 13 Depression: 2, 4, 6, 7, 8, 10, 12, 14
Stott et al. (2017)	UK (English)	268 dementia patients	CFA using maximum likelihood estimator with Satorra–Bentler correction for non-normal data	Anxiety: 1, 3, 5, 9, 11, 13 Depression: 2, 4, 6, 10, 12, 14 Items 7 and 8 excluded
Three-factors				
Brandberg et al. (1992)	Sweden (Swedish)	273 cancer patients	PCA with oblique rotation	Anxiety: 3, 5, 9, 13 Depression: 2, 4, 6, 8, 10, 12 Restlessness: 1, 7, 11, 14
Caci et al. (2003)	Ireland (English)	195 healthy students	CFA using Muthen's weighted least squares with mean and variance corrected estimate (WLSMV) and maximum likelihood estimator with Satorra–Bentler correction for non-normal data	Anxiety: 1, 3, 5, 9, 13 Depression: 2, 4, 6, 8, 10, 12 Restlessness: 7, 11, 14
Dunbar et al. (2000)	UK (English)	2547 healthy adults	CFA using maximum likelihood estimator with Satorra–Bentler correction for nonnormal data	Autonomic anxiety: 3, 9, 13 Negative affectivity: 1, 5, 7, 11 Anhedonic depression: 2, 4, 6, 8, 10, 12, 14
Friedman et al. (2001)	France (French)	2669 major depression clinic outpatients	PCA with varimax orthogonal rotation	Psychic Anxiety: 3, 5, 9, 13 Psychomotor Agitation: 1, 7, 11 Depression: 2, 4, 6, 8, 10, 12, 14
Lewis (1991)	UK (English)	173 adult dermatology clinic outpatients	PCA with varimax orthogonal rotation and CFA with maximum likelihood estimation	Factor 1: 3, 5, 8, 9, 13 Factor 2: 1, 7, 11, 12, 14 Factor 3: 2, 4, 6 *item 10 unspecified

2044835, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Study	Country (language)	Sample	Factor extraction method	Factor structure
Skilbeck et al. (2011)	Australia (English)	371 TBI patients	EFA using Principal axis factoring with varimax orthogonal rotation and CFA using maximum likelihood estimation	Anxiety: 3, 5, 9, 13 Psychomotor: 1, 6, 7, 8, 11,14 Depression: 2, 4, 10, 12
Stott et al. (2017)	UK (English)	268 dementia patients	CFA using maximum likelihood estimator with Satorra–Bentler correction for nonnormal data	Autonomic anxiety: 3, 9, 13 Negative affectivity: 1, 5, 11 Anhedonic depression: 2, 4, 6, 10, 12 Items 7 and 8 excluded
Four-factor				
Andersson (1993)	Sweden (Swedish)	163 general population adults	PCA with varimax orthogonal rotation	Well-being: 4, 10, 12 Momentary Anxiety: 3, 5, 7, 8, 9, 13 Power to Relax: 1, 6, 14 Undefined factor: 2, 11
Bi-factor models				
Norton et al. (2013)		21,820 (28 samples in meta analysis)	CFA using metaSEM in R	Anxiety: 1, 3, 5, 7, 9, 11, 13 Depression: 2, 4, 6, 8, 10, 12, 14 General factor: all items
Norton et al. (2013)		21,820 (28 samples in meta analysis)	CFA using metaSEM in R	Anxiety: 1, 3, 5, 9, 13 Depression: 2, 4, 6, 8, 10, 12 Restlessness: 7, 11, 14 General factor: all items

a two-factor (based on Zigmond and Snaith's original model) and a three-factor (anxiety, depression and restlessness) bifactor model (see Table 1). To test whether the bifactor models were superior to the 10 other factor structures identified in the literature, Norton et al. (2013) conducted a meta-confirmatory factor analysis (CFA) using inter-item correlation matrices from 28 samples from the review conducted by Cosco et al. (2012). These analyses found that the two-factor bifactor model had better fit than the other models tested. Other studies have also found a bifactor model to have better fit than unidimensional, bidimensional and tridimensional models for the HADS (Iani et al., 2014; Xie et al., 2012). Together, these findings support Norton et al.'s assertion that a bifactor model may better measure anxiety and depression.

In addition to their argument that a bifactor model may reduce inconsistency in HADS factor structures, Norton et al. (2012) suggested that methodological and statistical differences may also account for differing factor structures across samples. Bjelland et al. (2002) found that the majority of the early HADS studies had small sample sizes ($n < 250$) and provided limited detail about the analyses and psychometrics of the scale, which may have affected generalizability. In their review, Cosco et al. (2012)

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

found that two-factor models were typically derived from exploratory factor analyses (EFAs), including principal components analysis (PCA), where factors were determined by patterns of correlations between items (82% of studies reviewed). Alternatively, Rasch analysis, a statistical method based on item response theory (IRT) used to test whether items fit a unidimensional construct, has typically produced one-factor models (Cosco et al., 2012). However, in their Rasch analysis, Smith et al. (2006) found that unidimensional anxiety and depression factors were embedded within a unidimensional psychological distress factor, suggesting a bifactor or hierarchical structure. Studies employing CFAs have been used to compare models derived from EFAs, PCAs and Rasch methods. Contrary to EFA and PCA findings, Cosco et al. (2012) found that 67% of the CFA studies reviewed found superior fit with three-factor solutions, while only 29% reported two-factor solutions. In their meta-CFA, Norton et al. (2013) also found support for three-factor models but argued that the third factors reflected overextraction. Overextraction may somewhat explain differences in factor structures extracted from CFA and EFA/PCA, but it does not explain the differing solutions derived within CFA studies.

The variations in factor structures extracted across CFA studies may be due to the use of different estimation methods. Many studies have used maximum likelihood estimation which assumes that data are continuous and normally distributed. However, response categories for the HADS are ordinal (0–3), and in many studies data were non-normally distributed. For example, Norton et al. (2013) acknowledged that a limitation of their meta-CFA was that their dataset included data from previous studies that were based on Pearson correlation matrices. As these correlations assume continuous and normal data, this may have produced biased fit estimates. Therefore, consistent use of robust diagonal weighted least squares (DWLS) estimation methods that account for ordinal and non-normal data has been recommended to provide less biased estimates of model fit (Finney & DiStefano, 2006; Li, 2016; Shi et al., 2020).

Another potential reason for the inconsistencies in HADS factor structures is the inclusion of some negatively worded items. Coyne and van Sonderen (2012) argued that respondent inattention to changes in wording direction may have implications for how respondents approach HADS items across samples. Testing the effects of negative wording, Schönberger and Ponsford (2010) conducted a CFA comparing HADS models with and without negative wording factors and found that models with negative wording factors had better fit. Hence, while reversing item wording may reduce response biases, it would be worthwhile further investigating the effects of negative wording to determine whether item wording direction requires refinement.

The present study

As discussed above, our review of the literature revealed 14 possible factor structures for the HADS. Although Cosco et al. (2012) and Norton et al. (2013) have previously reviewed the factor structure of the HADS, an updated review is warranted to account for factor structures not examined in these past reviews. Expanding on Norton et al.'s (2013) meta-CFA of 10 factor structures, we reviewed the fit of 14 HADS structures identified in our literature review above (see Table 1). The additional four structures included in our study were Skilbeck et al.'s (2011) three-factor model for TBI samples, Stott et al.'s (2017) two and three-factor models that omitted mis-specified items 7 and 8, and Andersson's (1993) four-factor model. An updated review of the HADS factor structure was also needed in order to engage with current debates and incorporate contemporary methodological and theoretical approaches in determining factor structures. For example, there is evidence to suggest that varying factor structures across studies may be due to sample differences, a failure to consider hierarchical theories of anxiety and depression and methodological differences across factor analysis studies including the use of non-robust estimation methods and the effects of negative wording. The current study therefore aimed to provide an updated review of the structure of the HADS, with a particular focus on clarifying how methodological and theoretical considerations affect its latent structure.

From a methodological perspective, this study aimed to determine whether Cosco et al.'s (2012) finding of differing HADS latent structures across PCA, CFA and Rasch methods could be replicated by

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

comparing factor solutions derived from these three methods, using robust estimation methods. Similar to the findings of Cosco et al. (2012) we hypothesized that the Rasch analysis would support a unidimensional structure, PCA a two-factor solution and CFA three-factor models. Our study also aimed to investigate the impact of using robust estimation methods in CFA, based on previous studies, including Norton et al.'s (2013) meta-CFA, potentially including biased fit indices through the use of correlation matrices that assumed continuous and normal data. Therefore, to ascertain if these methodological differences impacted model fit, we used polychoric correlations and robust DWLS estimation methods that account for the ordinal and non-normal nature of HADS responses in our dataset. We also investigated the impact of negative wording on the interpretability of the factor structure of the HADS. Consistent with the findings of Schönberger and Ponsford (2010), we predicted that including a negative wording factor would improve model fit.

From a theoretical perspective, as anxiety and depression are believed to tap into an overarching general distress factor, we also aimed to assess the validity of a hierarchical theory of anxiety and depression by testing Norton et al.'s (2013) bifactor models. In line with Norton et al.'s (2013) bifactor model findings, we predicted that a large portion of the relationship between anxiety and depression factors would be absorbed by an overarching general factor, and that the two-factor bifactor model would have superior fit to other models.

The heterogeneity in HADS factor structures across samples, particularly across clinical and community samples, raises questions as to which factor structure (or structures) can be appropriately used to measure anxiety and depression. In this study, our investigation of the HADS factor structure employed a sample of Australian community-dwelling adults. To date, in Australian samples, the factor structure of the HADS has only been investigated in those with TBI (Skilbeck et al., 2011, 2019). It has been suggested that in TBI populations, HADS items may reflect brain injury-related cognitive issues (e.g. deficits in processing speed and focus) rather than anxiety or depression (Schönberger & Ponsford, 2010). In their TBI sample, Skilbeck et al. (2011) extracted a two-factor model using EFA, but CFA supported a three-factor model that included a psychomotor agitation factor, suggesting that the factor structure may differ in this sample due to TBI-specific patterns of symptoms. Internationally, Saez-Flores et al. (2018) also found three-factor models to obtain better fit in TBI and other medical samples. As Skilbeck et al.'s three-factor model was not examined in Norton et al.'s (2013) meta-CFA, the current study will compare this model to competing models. Thus, in addition to undertaking a more rigorous and updated analysis of the HADS factor structure for international use, the present paper advances previous work by examining the generalizability of Skilbeck et al.'s three-factor model to an Australian community context. In doing so, this study will provide greater clarity for clinicians regarding which HADS structure and scoring cut-off is most appropriate for use within TBI and community-dwelling samples.

METHOD

Participants

We recruited 351 Australian community-dwelling adults through social media, snowball and convenience sampling methods during 2016–2019. Participants were screened for clinical conditions using a health and demographic survey, with one case excluded due to a prior head injury. The Alcohol Use Disorders Identification Test, Section A (AUDIT-A; Barbor et al., 1992) screened for hazardous alcohol use, although no participants were excluded for excessive levels of alcohol use. Incomplete HADS responses of three participants were deleted. The final sample comprised of 347 healthy adults aged 17–86 (M = 35.73, SD = 17.41), 189 (54.5%) were female and 158 (45.4%) were male.

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Materials

A health and demographics survey was designed for the purpose of this study. This survey included questions on participants' age, education, occupation and health conditions (e.g. brain injury) that have been found to influence cognitive ability (Wechsler, 2009). The three-item AUDIT-A (Barbor et al., 1992) was used to measure the frequency, quantity and amount of hazardous drinking on a scale of 0–4, with higher scores indicating greater alcohol use.

Participants completed the 14-item HADS. The scale contains two 7-item subscales measuring anxiety and depression symptoms. These items are rated on a 4-point scale ranging from 0 to 3, with higher scores indicating greater levels of anxiety and depression. Bjelland et al. (2002) reviewed the psychometric properties of the HADS. Across 15 studies reviewed, internal consistency of the HADS ranged from 0.68 to 0.93 (M = 0.83) for the anxiety subscale and from 0.67 to 0.90 (M = 0.82) for the depression subscale. In terms of concurrent validity, six studies found moderate to strong correlations between the Beck Depression Inventory II and HADS depression (r = .62 to .73), anxiety (r = .61 to .83) and total scales (r = .73). Similarly, across five studies, moderate to strong correlations between the State Trait Anxiety Inventory and the HADS depression (r = .52 to .65), anxiety (r = .64 to .81) and total scale (r = .68 to .71) were reported. The HADS was also found to have good sensitivity and specificity across the 24 studies reviewed, with a cut off score of eight producing sensitivity and specificity scores between 0.70 and 0.90.

Procedure

Ethical approval was gained from institutional Human Research Ethics Committees (Protocols: 16055 and H7116). After giving their consent, participants completed the health and demographic survey, AUDIT-A and HADS. Items were self-administered, with assistance from the researcher when requested. Data were collected as part of a larger study, with participants also completing the Weschler Abbreviated Scale of Intelligence – Second edition (Wechsler, 2011), Test of Premorbid Functioning (Wechsler, 2009), National Adult Reading Test (Nelson, 1982) and Weschler Adult Reading Test (Wechsler, 2001). The findings relating to those measures are reported in [McGrath et al., 2022; Thomas et al., 2020, 2021].

Analyses

IBM SPSS Statistics 27 was used for the calculation of descriptive statistics for the 14 HADS items and total score to examine item distributions. This software was also used for the PCA to extract a simple structure of the HADS. For the PCA, a Kaiser criterion of eigenvalues greater than 1 was used to extract components.

The WINSTEPS software (Linacre, 2009) was used to conduct the Rasch analysis to investigate whether the HADS fits a unidimensional latent structure, similar to a one-factor or unidimensional model that would be obtained through CFA methods. The Rasch analysis utilized Masters' (1982) partial credit model for polytomous items, as it permits the structure of the rating scale to vary across items and allows unobserved response option categories.

R version 4.2.2 (R Core Team, 2022) was used for remaining analyses. Lavaan (Rosseel, 2012) was used to calculate polychoric correlations and compare models using CFA. In these analyses, DWLS estimation, which is based on polychoric correlations for non-normal and ordinal responses, was used to provide robust fit estimates (Finney & DiStefano, 2006; Li, 2016). Model fit was assessed using the chi-square test of exact fit; however, large sample sizes, ordinal data and non-normal data often lead to models being rejected (Shi et al., 2018). As such, we included additional fit indices including the Root Mean Squared Error of Approximation (RMSEA), standardized root mean squared residual (SRMR), comparative fit index (CFI) and Tucker–Lewis index (TLI). According to Hu and Bentler (1999), RMSEA scores below

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

0.06, SRMR scores below 0.08 and CFI and TLI scores above 0.95 suggest good fit. Caution needs to be exercised in using these cut-off scores to evaluate fit using ordinal data (Shi et al., 2020; Xia & Yang, 2019).

To assess the internal consistency of subscales within each of the models, omega coefficients were calculated using the R semTools package (Jorgensen et al., 2022). Omega is recommended when the assumption of tau equivalence is not met, that is, when items have varying factor loadings on a scale (McNeish, 2018). For the bifactor models, omega hierarchical coefficients and explained common variance (ECV) were calculated using Dueber's (2017) bifactor indices calculator. Omega hierarchical indices provide a reliability estimate for the general factor by removing variance produced by the subfactors (McNeish, 2018). Omega hierarchical subscale indices on the other hand remove variance from the general factor to provide a reliability estimate of the subscales (Rodriguez et al., 2016) ECV is used to test the multidimensionality of a model; it assesses strength of the general factor by measuring the proportion of variance explained by the general factor compared to the specific factors. An ECV score above 0.70 indicates that the scale is most likely unidimensional (Rodriguez et al., 2016).

RESULTS

The results are provided in three parts. First, distributions and inter-item correlations of the 14 HADS items and total score are reported. Next, a comparison of models produced by Rasch, PCA and CFA methodologies is presented. Finally, the CFA findings examining the fit of the 14 models identified in the literature review for our Australian community-dwelling sample are explored. Internal consistency and ECV for the subscales tested in the CFAs are also reported.

HADS distributions and relationships between items and scales

Table 2 shows average scores, range of scores and skew of the 14 HADS items and total score. As can be seen in the table, participants endorsed the full range of responses on all items except item 6. Inspection of histograms and z-skew statistics indicated that most items were positively skewed, with low levels of anxiety and depression reported in our nonclinical sample.

TABLE 2 Descriptive statistics for the HADS.

Item	*M*	*SD*	Range	*Z*-skew
1	1.12	0.66	0–3	6.27
2	0.49	0.60	0–3	6.85
3	0.93	0.89	0–3	4.34
4	0.23	0.50	0–3	17.29
5	1.27	0.83	0–3	2.28
6	0.34	0.54	0–2	9.68
7	0.88	0.73	0–3	4.24
8	0.84	0.76	0–3	5.36
9	0.79	0.66	0–3	4.70
10	0.57	0.74	0–3	7.74
11	1.20	0.88	0–3	3.24
12	0.32	0.62	0–3	15.57
13	0.77	0.77	0–3	6.42
14	0.32	0.68	0–3	18.26
Total	10.08	5.69	0–38	8.25

Note: Z-skew = skew/standard error of skew.

Polychoric correlations were calculated to investigate the relationship between the 14 HADS items. The inter-item correlations are shown in Table 3. As can be seen in the table, there were small to moderate correlations for the majority of items both within and between anxiety and depression subscales. The anxiety and depression subscales were also found to be moderately and significantly correlated (r = .47).

Comparison of factor extraction methodologies

To test whether different factor extraction techniques produced different factor structures, we investigated the factor structure of the HADS using PCA and Rasch analyses. As hypothesized, the PCA extracted two components with eigenvalues greater than 1, with the solution accounting for 43.94% of variance. However, possible third and fourth components that had eigenvalues above 0.9 and accounted for an extra 7% and 6.6% of variance were identified. A comparison of direct oblimin and varimax rotations revealed that the orthogonal varimax rotation provided a simpler solution. Loading of items onto the two components (as shown in Table 4) matched Zigmond and Snaith's (1983) model, but with small cross loadings of anxiety items 1 and 7 onto the depression subscale.

Finally, a Rasch analysis was conducted to investigate the unidimensionality of the HADS. A Principal Component Analysis of Residuals (Linacre, 1998) found that the HADS was most likely bi-dimensional. The raw variance explained by the latent trait was 45.5% and variance explained by the first contrast in the residuals was 8% (equating to an eigenvalue of 2.1). An examination of this first contrast showed the HADS items clearly loading onto the anxiety and depression factors proposed in Zigmond and Snaith's (1983) original HADS model. Thus, contrary to our expectations that the Rasch analysis would support a unidimensional structure, these findings instead supported the original two-factor model of the HADS.

Factor structure and reliability of the HADS subscales

To explore the factor structure of the HADS, we conducted CFAs to measure the fit of each of the 14-factor structures identified in Table 1 to our Australian adult community-dwelling sample. Table 5 shows the CFA fit indices for the 14 models as well as omega and omega hierarchical reliability indices.

TABLE 3 HADS interitem correlations (polychoric).

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	–													
2	.32	–												
3	.35	.25	–											
4	.28	.23	.21	–										
5	.53	.31	.52	.27	–									
6	.30	.39	.31	.39	.40	–								
7	.44	.31	.32	.25	.40	.34	–							
8	.35	.28	.24	.26	.31	.28	.18	–						
9	.31	.20	.38	.19	.41	.18	.32	.12	–					
10	.21	.21	.09	.23	.15	.17	.14	.27	.07	–				
11	.34	.16	.32	.22	.44	.17	.36	.26	.20	.19	–			
12	.26	.35	.26	.34	.29	.39	.25	.29	.09	.22	.17	–		
13	.48	.27	.52	.22	.60	.25	.41	.26	.50	.15	.36	.24	–	
14	.24	.21	.24	.18	.25	.21	.26	.11	.09	.19	.18	.23	.26	–

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 PCA item loadings.

Item	Zigmond and Snaith (1983) Original two-factor model loading	Factor 1 Anxiety	Factor 2 Depression
1	Anxiety	**0.60**	**0.37**
2	Depression	0.24	**0.58**
3	Anxiety	**0.69**	0.18
4	Depression	0.17	**0.61**
5	Anxiety	**0.76**	0.28
6	Depression	0.25	**0.63**
7	Anxiety	**0.56**	**0.31**
8	Depression	0.21	**0.55**
9	Anxiety	**0.71**	−0.05
10	Depression	−0.00	**0.57**
11	Anxiety	**0.52**	0.23
12	Depression	0.11	**0.69**
13	Anxiety	**0.80**	0.15
14	Depression	0.26	**0.36**

Note: Significant factor loadings (>.30) highlighted in bold.

As can be seen in Table 5, the unidimensional model had good reliability, but the model had the poorest fit of all models tested.

Of the two-factor models, Zigmond and Snaith's (1983) model had good fit as well as acceptable reliability indices for the anxiety and depression scales respectively. In Moorey et al.'s (1991) model, moving item 7 from the anxiety to depression subscale slightly improved reliability for the depression scale but reduced RMSEA levels. Whereas Stott et al.'s (2017) model that eliminated mis-specified items 7 and 8 produced good fit, but unacceptable reliability for the depression scale. Thus, the original two-factor model was the most appropriate, as modifications to this model compromised either reliability or model fit.

Fit of the three-factor models was comparable to that of the two-factor models, with the exception of Lewis' (1991) model, that fell below CFI and TLI thresholds. In most three-factor models, the anxiety subscale had acceptable to good reliability, but the addition of the third factor in these models resulted in reliability that was on the borderline of acceptability for the depression subscale, and unacceptable reliability for the third factor. Dunbar et al.'s (2000) model on the other hand had lower reliability on the anxiety subscale, but all three subscales were at acceptable levels. In our Australian community-dwelling sample, Skilbeck et al.'s (2011) model for Australians with TBI had good fit that was equivalent fit to the original two-factor model. However, the depression and restlessness subscales had poor reliability in this sample. Therefore, the movement of some anxiety and depression items to the third factor had little impact on model fit but came at the expense of reliability.

The bifactor models that comprised of an overarching general factor and orthogonal specific factors were tested next. The three-factor model indicated a Heywood case with the restlessness subscale producing non-significant factor loadings and negative variance (−0.016, $p = .791$) that precluded the calculation of standardized estimates. This suggested that this model was mis-specified and that the variance in the restlessness subscale was not specific and therefore better attributed to the general factor.

Norton et al.'s (2013) two-factor bifactor model had the best fit statistics of the 14 models, but item 14 on the depression subscale did not load significantly. On the anxiety subscale, loadings for all seven items were non-significant with particularly low loadings for items 1, 7 and 11 (0.12, 0.08 and 0.14), and variance was low (0.02). All items, except item 10 loaded more strongly onto the general factor than the specific factors. Supporting this, the anxiety subscale explained only 26.9% of common variance, compared to 37.2% for the depression scale, and 68.7% for the general factor. Omega hierarchical reliability indices in Table 5 also showed that the anxiety subscale accounted for only a small portion of reliable variance,

TABLE 5 Summary of model fit and internal consistency of subscales.

	Model fit							Internal consistency (ω)				
Model	χ^2	*df*	*p*	RMSEA [90% CI]	SRMR	CFI	TLI	General	Anxiety	Depression	Factor 3	Factor 4
1-factor models												
Razavi et al. (1990)	278.395	77	<.001	0.087 [0.076, 0.098]	0.082	0.905	0.888	0.85				
2-factor models												
Zigmond and Snaith (1983)	154.101	76	<.001	***0.054 [0.042, 0.067]***	***0.058***	***0.963***	***0.956***		0.83	0.71		
Moorey et al. (1991)	177.931	76	<.001	0.062 [0.050, 0.074]	***0.063***	***0.952***	0.943		0.82	0.73		
Stott et al. (2017)	92.752	53	.001	***0.047 [0.030, 0.062]***	***0.053***	***0.978***	***0.972***		0.82	0.68		
3-factor models												
Brandberg et al. (1992)	115. 298	74	.002	***0.040 [0.025, 0.054]***	***0.050***	***0.981***	***0.976***		0.80	0.70	0.64	
Caci et al. (2003)	138.337	74	<.001	***0.050 [0.037, 0.063]***	***0.054***	***0.970***	***0.963***		0.82	0.70	0.55	
Dunbar et al. (2000)	116.207	74	.001	***0.041 [0.026, 0.054]***	***0.051***	***0.980***	***0.976***		0.74	0.71	0.74	
Friedman et al. (2001)	119.363	74	.001	***0.042 [0.028, 0.056]***	***0.051***	***0.979***	***0.974***		0.80	0.71	0.63	
Lewis (1991)	175.226	62	<.001	0.073 [0.060, 0.085]	***0.067***	0.946	0.932		[a]0.78	[a]0.63	[a]0.66	
Skilbeck et al. (2011)	155.376	74	<.001	***0.056 [0.044, 0.069]***	***0.059***	***0.962***	***0.953***		0.80	0.59	0.69	
Stott et al. (2017)	67.430	51	.061	***0.031 [0.000, 0.049]***	***0.046***	***0.991***	***0.988***		0.74	0.68	0.71	
4-factor models												
Andersson (1993)	2221 .679	91	<.001	0.078 [0.066, 0.089]	***0.070***	0.931	0.911		0.51	[b]0.80	0.51	0.27
									ω_H	ω_{HS}	ω_{HS}	ω_{HS}
Bifactor models												
Norton et al. (2012) 2 bifactor	85.040	63	.034	***0.032 [0.009, 0.048]***	***0.041***	***0.990***	***0.985***		0.76	0. 19	0.29	
Norton et al. (2012) 3 bifactor	model not indentified											
Negative wording factor												
Norton et al. (2012) 2 bifactor	72.233	57	***.084***	***0.028 [0.000, 0.046]***	***0.037***	***0.993***	***0.989***		0.75	0.15	0.24	0.19[c]

Abbreviations: 90% CI, 90% confidence interval; CFI, comparative fit index; H = Omega heirachical; HS = Omega heirachical subscale; RMSEA, root mean squared error of approximation; SRMR = squared root mean residual, TLI = Tucker Lewis index; H = Omega heirachical, HS = Omega heirachical subscale; Fit indices showing good fit are highlighted in bold and italics.

[a]Undefined factors.

[b]Well-Being.

[c]Negative wording factor.

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

with most variance explained by the general factor. Although, the variance explained by the general factor was below the threshold of 0.80 for unidimensionality (Reise et al., 2013). Given the low anxiety subscale loadings, as recommended by Chen et al. (2006) and Brown (2015), we tested an incomplete bifactor model whereby anxiety items loaded onto the general factor only. However, this model was not identified. Together, these findings partially confirm our hypothesis that the general factor would explain a large portion of variance in anxiety and depression subscales, with superior fit to competing models. Nevertheless, although the general factor explained most of the variance in anxiety and depression, particularly anxiety, the general factor was not entirely unidimensional.

To test the effects of negative wording, a CFA of the best fitting model (i.e. Norton et al.'s two-factor bifactor model) that included a negative wording factor including items 2, 4, 6, 7, 12 and 14, orthogonal anxiety and depression factors and an overarching general factor. As can be seen in Table 5, the addition of the negative wording factor slightly improved model fit. Including the negative wording factor, which explained 10% of common variance, reduced the ECV of the anxiety subscale to 12.7%, depression subscale to 13.1% and general factor to 64.2%. However, item 11 and 14 on the negative wording factor did not significantly load onto the specific factors. In summary, as predicted, some of the variance in some, but not all, HADS items can be attributed to the effects of negative wording.

DISCUSSION

The aim of this study was to address theoretical and methodological problems listed in the introduction pertaining to the factor structure of the HADS by investigating the latent structure of the HADS in an Australian community-dwelling sample. We predicted that the HADS structure would vary depending on the type of analysis used; specifically, we hypothesized that Rasch analysis would support a unidimensional model, PCA a two-factor model and CFA a three-factor model. Supporting a hierarchical theory of anxiety and depression, we also hypothesized that Norton et al.'s two-factor bifactor model with an overarching general factor and separate anxiety and depression factors would have superior fit to competing models. Moreover, we hypothesized that including a negative wording factor would improve model fit. These predictions were only partially supported. The Rasch analysis, PCA and CFA all primarily supported Zigmond and Snaith's (1983) original two-factor model. While there was some support for three-factor models in PCA and CFA, these models resulted in unacceptable reliability for some of the subscales. Norton et al.'s two-factor bifactor model had better fit than single-level models and some variance in anxiety and depression was accounted for by the general factor; however, the general factor was not found to fully capture this variance. Finally, as expected, the inclusion of a negative wording factor improved model fit.

Consistent with previous reviews of HADS factor structures (Bjelland et al., 2002; Cosco et al., 2012), our findings indicated that differing statistical methods across studies may explain some of the inconsistencies in factor structures that have been reported in the literature. In our study, all three extraction methods (Rasch, PCA and CFA) found Zigmond and Snaith's original two-factor model was suitable for our Australian community-dwelling sample. However, the PCA identified possible third and fourth factors. Some of these inconsistencies in factor extraction may be due to the use of arbitrary criterion in EFA and PCA studies. For example, visual inspection of scree plots or fixing extraction parameters is highly subjective. In our study, we used a conservative Kaiser criterion of eigenvalues above 1 (Nunnally & Bernstein, 1994) which may have excluded potentially viable third and fourth factors with eigenvalues above 0.9. As suggested by Cosco et al. (2012), the use of more robust and flexible CFA and IRT methods is recommended to eliminate inconsistencies in factor structures found in EFA and PCA.

In our study, multiple two and three-factor models were found to have good fit. However, in these models, the reduction of the number of items or rearrangement of items across factors tended to reduce internal consistency of the factors. For example, items 7 and 8 which have previously been identified as problematic due to anomalous loadings (Bjelland et al., 2002; Cosco et al., 2012) were excluded in Stott et al.'s (2017) models, but in these models, the depression subscale had unacceptable reliability.

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Similarly, in the three-factor models, moving items from anxiety or depression subscales to create the third subscale came at the cost of reliability, particularly in the depression and third factors, with only Dunbar et al.'s (2000) model maintaining acceptable reliability across all three subscales. Thus, in addition to providing difficulties in using existing diagnostic cut-offs for the original two-factor model, the use of these reduced-item scales or three-factor models does not appear to be suitable in this community-based sample. However, given the inconsistency in factor structures identified across samples (Cosco et al., 2012; Saez-Flores et al., 2018), further replication is needed to determine whether these findings can be generalized to other samples.

In the Australian context, our study was the first to investigate the latent structure of the HADS in a community-dwelling sample. In this sample, the original two-factor HADS model (Zigmond & Snaith, 1983) was found to be the most appropriate solution in Rasch, PCA and CFA. This differed from the three-factor structure found in Australian adults with TBI (Skilbeck et al., 2011). Saez-Flores et al. (2018) also found clinical samples to produce three-factor structures as opposed to community-dwelling samples where two-factor structures were more common. Thus, when using the HADS, researchers and clinicians should be mindful of potential differences in factor structures across samples and appropriate normative data should be used.

In our study, we also considered whether a hierarchical theory of anxiety and depression may better explain the factor structure of the HADS. Norton et al. (2013) suggested that the superiority of three-factor models over two-factor models in previous CFA studies may be a consequence of anxiety and depression sharing variance with a third general psychological distress factor. They posited that this general factor explains much of the common variance in the HADS. As expected, we found Norton et al.'s (2013) two-factor bifactor model, which had orthogonal anxiety and depression factors, as well as a separate general psychological distress factor, had the best fit of the 14 models tested in our CFA. However, these findings need to be considered with caution, as fit indices typically favour bifactor models over single-level correlated factor models because these indices are less likely to penalize misspecifications in complex models with more parameters (Greene et al., 2019). As bifactor models are designed to absorb as much variance into the general factor as possible, it remains unclear whether this factor reflects general psychological distress or is a composite of unexplained common variance (Greene et al., 2019). Due to these issues, Bornovalova et al. (2020) argue that rather than focusing on global fit, bifactor models may be better used to determine the presence, strength and reliability of a unidimensional factor through analysis of factor loadings. Supporting the presence of a bifactor model with general distress factor, we found significant interitem correlations between the majority of anxiety and depression items, indicating shared variance. In our PCA, although two factors were extracted using the Kaiser criterion, the first factor reflected a general factor, explaining 33.7% of variance. As noted by Norton et al. (2013) in previous EFA and PCA studies, our unrotated solution had all items except item 10 loading onto a general factor, but when rotation was applied to achieve simple structure, additional factors emerged. Moreover, in our Rasch analysis, it is possible that a second factor was over-extracted, as the variance explained by the first contrast in the residuals was only slightly above the criteria of 2.0 needed to extract this additional factor. Yet, the ECV and hierarchical reliability indices were slightly below the threshold used to determine unidimensionality of this general factor. Issues with low variance and non-significant loadings on the anxiety scale also suggest that the bifactor model may have been somewhat mis-specified. Thus, while there is some support for a hierarchical model of anxiety and depression with an overarching general distress factor, anomalous loadings, particularly of anxiety items and the lack of unidimensionality of the general distress factor suggests that the HADS in its current state is mis-specified and should be interpreted with caution. Our findings also support the claims of Coyne and van Sonderen (2012) that the HADS may be mis-specified because of its inclusion of negatively worded items. Similar to the findings of Schönberger and Ponsford (2010), we found that negative wording impacted upon interpretability of the HADS. Specifically, we found a small improvement in model fit of the two-factor bifactor model when negative wording was controlled for with a negative wording factor. However, not all negatively worded items loaded significantly onto the negative wording factor, suggesting that negative wording may only affect variance in some of the HADS items.

2044835, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Limitations

As with all factor analysis studies, our analyses were subject to arbitrary criteria when extracting factors. This resulted in some ambiguity in interpreting factors, as in some cases, differences between extracted and non-extracted factors were marginal. To reduce bias in estimating model fit, we used robust DWLS CFA methods. In addition, to minimize subjectivity in factor extraction, we employed commonly accepted cut-offs for Rasch (Linacre, 1998), PCA (Nunnally & Bernstein, 1994) and CFA (Hu & Bentler, 1999). The CFA cut-offs used were developed for normally distributed, continuous data. As our data were skewed and ordinal, interpretation of fit, especially close to cut-off values, was carried out with caution with cut-offs used as guidelines rather than clear delineators of fit (Shi et al., 2020; Xia & Yang, 2019).

We used a convenience sample of community-dwelling Australian adults. We attempted to create a representative sample by stratifying recruitment to achieve a balance of age and gender. We also screened for health conditions and hazardous alcohol use to reduce potential clinical confounds. The positively skewed distribution of HADS responses, with low levels of anxiety and depression, indicated that our sample did reflect a nonclinical population. However, without a fully stratified, representative sample, there are limits on the generalizability of our findings. Future research is therefore required to determine whether the findings in our sample can be generalized.

Implications and future directions

In our community-dwelling sample, the original two-factor model (Zigmond & Snaith, 1983) was found to have good fit and reliability compared to unidimensional, modified two-factor models, three-factor models and a four-factor model. There was also some support for a two-factor bifactor model (Norton et al., 2013) based on a hierarchical theory of anxiety and depression with a general psychological distress factor (Clark & Watson, 1991; Watson, 2005). In this model, the lack of clear separation between the three factors reflected the overlap of symptoms seen with the comorbidity of anxiety and depression (Dunn & McCray, 2020). Interpreting this bifactor model, with all but one depression item loading significantly onto the depression factor, there was evidence for a separate depression factor. However, items on the anxiety subscale could not be separated from the general factor, suggesting that this subscale measures a more multifaceted, general distress aspect of anxiety. Due to this overlap, it is recommended that the HADS should not be used in clinical settings to distinguish between anxiety and depression (Norton et al., 2013). Instead, the HADS appears to be most suitable for screening purposes, with responses to individual items providing information to facilitate further investigation with other assessment methods more closely aligned with clinical diagnostic taxonomies.

We have presented evidence to suggest that the use of different psychometric cut-offs and extraction procedures may result in different factor structures being identified. Therefore, it is recommended that guidelines for factor extraction in HADS studies be developed to improve consistency across studies. This includes the use of estimation methods that account for the ordinal and often non-normal nature of HADS data (Finney & DiStefano, 2006; Li, 2016; Shi et al., 2020). When comparing the latent structure of the HADS, most studies have focused on fit statistics; however, consideration of the theoretical and practical suitability of the model needs to be a priority (Decker, 2021). This is particularly pertinent when investigating bifactor models; in these cases, the presence, strength and reliability of the general factor needs to be considered (Bornovalova et al., 2020; Decker, 2021).

Mis-specified and negatively worded items appear to be contributing to some of the inconsistencies in factor structures that have been derived across studies. Rather than abandon the HADS, as proposed by Coyne and van Sonderen (2012) though, we agree with Norton et al. (2012) that the HADS requires refinement. This refinement should include psychometric investigations of the HADS, using contemporary psychometric methods to elucidate whether items can be refined to improve scale properties (Xie et al., 2012). For example, item banking using IRT methods may tailor items to better suit testing

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

populations (Sunderland et al., 2019). If such refinements are undertaken, these new versions of the scales will need to be validated in the samples in which they are intended to be used.

In conclusion, our findings support separate, yet comorbid and overlapping anxiety and depression constructs that may be explained by a hierarchical psychological distress factor. Due to this overlap, we suggest that the anxiety and depression HADS subscales should not be used to differentiate between these comorbid conditions. As we found variations in factor structures due to the use of differing methodologies, item misspecification and item wording to influence model fit, further research and psychometric investigations using contemporary methods a required to refine the HADS item wording and structure to better assess anxiety and depression symptoms.

AUTHOR CONTRIBUTIONS

M. Lloyd: Conceptualization; data curation; formal analysis; investigation; validation; visualization; writing – original draft; writing – review and editing. **Nicole Sugden:** Conceptualization; data curation; formal analysis; methodology; project administration; resources; software; supervision; validation; visualization; writing – original draft; writing – review and editing. **Matt Thomas:** Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; supervision; validation; writing – review and editing. **Andrew McGrath:** Conceptualization; data curation; investigation; methodology; project administration; resources; supervision; validation; writing – review and editing. **Clive Skilbeck:** Conceptualization; data curation; funding acquisition; investigation; methodology; project administration; resources; validation; writing – review and editing.

ACKNOWLEDGEMENTS

The authors acknowledge Michael Magee, Catherine Richardson and Michelle Vondgara for their assistance with collecting data for this project. Open access publishing facilitated by Charles Sturt University, as part of the Wiley - Charles Sturt University agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

All authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

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

ORCID

Nicole Sugden https://orcid.org/0000-0002-7010-1136
Matt Thomas https://orcid.org/0000-0002-1622-2092
Andrew McGrath https://orcid.org/0000-0001-9249-0219

REFERENCES

Andersson, E. (1993). The hospital anxiety and depression scale: Homogeneity of the subscales. *Social Behavior and Personality: An International Journal*, *21*(3), 197–204. https://doi.org/10.2224/sbp.1993.21.3.197

Barbor, T., La Fuente, J., Saunders, J., & Grant, M. (1992). *AUDIT: The alcohol use disorders identification test: Guidelines for use in primary health care*. World Health Organization.

Barth, J., & Martin, C. R. (2005). Factor structure of the hospital anxiety and depression scale (HADS) in German coronary heart disease patients. *Health and Quality of Life Outcomes*, *3*(1), 1–9. https://doi.org/10.1186/1477-7525-3-15

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the hospital anxiety and depression scale: An updated literature review. *Journal of Psychosomatic Research*, *52*(2), 69–77. https://doi.org/10.1016/s0022-3999(01)00296-3

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M. (2020). Appropriate use of bifactor analysis in psychopathology research: Appreciating benefits and limitations. *Biological Psychiatry*, *88*(1), 18–27. https://doi.org/10.1016/j.biopsych.2020.01.013

Brandberg, Y., Bolund, C., Sigurdardottir, V., Sjödén, P. O., & Sullivan, M. (1992). Anxiety and depressive symptoms at different stages of malignant melanoma. *Psycho-Oncology*, *1*(2), 71–78. https://doi.org/10.1002/pon.2960010204

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

Caci, H., Baylé, F. J., Mattei, V., Dossios, C., Robert, P., & Boyer, P. (2003). How does the hospital and anxiety and depression scale measure anxiety and depression in healthy subjects? *Psychiatry Research*, *118*(1), 89–99. https://doi.org/10.1016/s0165-1781(03)00044-1

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189–225.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*(3), 316–336.

Cosco, T. D., Doyle, F., Ward, M., & McGee, H. (2012). Latent structure of the hospital anxiety and depression scale: A 10-year systematic review. *Journal of Psychosomatic Research*, *72*(3), 180–184. https://doi.org/10.1016/j.jpsychores.2011.06.008

Covid-19 Mental Disorders Collaborators. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, *398*(10312), 1700–1712. https://doi.org/10.1016/S0140-6736(21)02143-7

Coyne, J. C., & van Sonderen, E. (2012). No further research needed: Abandoning the hospital and anxiety depression scale (HADS). *Journal of Psychosomatic Research*, *72*(3), 173–174. https://doi.org/10.1016/j.jpsychores.2011.12.003

Decker, S. L. (2021). Don't use a bifactor model unless you believe the true structure is bifactor. *Journal of Psychoeducational Assessment*, *39*(1), 39–49. https://doi.org/10.1177/0734282920977718

Dueber, D. M. (2017). Bifactor indices calculator: A Microsoft excel-based tool to calculate various indices relevant to bifactor CFA models. https://doi.org/10.13023/edp.tool.01

Dunbar, M., Ford, G., Hunt, K., & Der, G. (2000). A confirmatory factor analysis of the hospital anxiety and depression scale: Comparing empirically and theoretically derived structures. *British Journal of Clinical Psychology*, *39*(1), 79–94. https://doi.org/10.1348/014466500163121

Dunn, K. J., & McCray, G. (2020). The place of the Bifactor model in confirmatory factor analysis investigations into construct dimensionality in language testing. *Frontiers in Psychology*, *11*(1357), 1–16. https://doi.org/10.3389/fpsyg.2020.01357

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. *Structural Equation Modeling: A Second Course*, *10*(6), 269–314.

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

GBD 2019 Diseases and Injuries Collaborators. (2020). 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. *The Lancet*, *396*(10258), 1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9

Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., Waldman, I. D., Cicero, D. C., Conway, C. C., Docherty, A. R., Fried, E. I., Ivanova, M. Y., Jonas, K. G., Latzman, R. D., Patrick, C. J., Reininghaus, U., Tackett, J. L., Wright, A. G. C., & Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of Abnormal Psychology*, *128*(7), 740–764. https://doi.org/10.1037/abn0000434

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the depression anxiety stress scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *44*(2), 227–239. https://doi.org/10.1348/014466505x29657

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Iani, L., Lauriola, M., & Costantini, M. (2014). A confirmatory bifactor analysis of the hospital anxiety and depression scale in an Italian community sample. *Health and Quality of Life Outcomes*, *12*(1), 84. https://doi.org/10.1186/1477-7525-12-84

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). semTools: Useful tools for structural equation modeling. R package version 0.5-6. Retrieved from https://CRAN.R-project.org/package=semTools

Julian, L. J. (2011). Measures of anxiety: State-trait anxiety inventory (STAI), Beck anxiety inventory (BAI), and hospital anxiety and depression scale-anxiety (HADS-A). *Arthritis Care & Research*, *63*(S11), S467–S472. https://doi.org/10.1002/acr.20561

Kalin, N. H. (2020). The critical relationship between anxiety and depression. *American Journal of Psychiatry*, *177*(5), 365–367. https://doi.org/10.1176/appi.ajp.2020.20030305

Lewis, G. (1991). Observer bias in the assessment of anxiety and depression. *Social Psychiatry and Psychiatric Epidemiology*, *26*(6), 265–272.

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Linacre, J. M. (1998). Structure in Rasch residuals: Why principal components analysis (PCA). *Rasch Measurement Transactions*, *12*(2), 636.

Linacre, J. M. (2009). Winsteps Rasch measurement. In (Version 3.70.0.2) [Computer Program]. Retrieved from http://www.winsteps.com/winsteps.htm

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174. https://doi.org/10.1007/BF02296272

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

McGrath, A., Thomas, M., Sugden, N., & Skilbeck, C. (2022). The Flynn effect in estimates of premorbid intellectual functioning in an Australian sample. *Australian Journal of Psychology*, 2001297. https://doi.org/10.1080/00049530.2021.2001297

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Moorey, S., Greer, S., Watson, M., Gorman, C., Rowden, L., Tunmore, R., Robertson, B., & Bliss, J. (1991). The factor structure and factor stability of the hospital anxiety and depression scale in patients with cancer. *The British Journal of Psychiatry*, *158*(2), 255–259.

Nelson, H. E. (1982). *National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual.* Nfer-Nelson.

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

Norton, S., Sacker, A., & Done, J. (2012). Further research needed: A comment on Coyne and van Sonderen's call to abandon the hospital anxiety and depression scale. *Journal of Psychosomatic Research*, *73*(1), 75–76. https://doi.org/10.1016/j.jpsychores.2012.04.005

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Razavi, D., Delvaux, N., Farvacques, C., & Robaye, E. (1990). Screening for adjustment disorders and major depressive disorders in cancer in-patients. *The British Journal of Psychiatry*, *156*(1), 79–83. https://doi.org/10.1192/bjp.156.1.79

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rosseel, Y. (2012). Lavaan: An R package for structural equation modelling. *Journal of Statistical Software*, *48*, 1–36.

Saez-Flores, E., Tonarely, N. A., Barker, D. H., & Quittner, A. L. (2018). Examining the stability of the hospital anxiety and depression scale factor structure in adolescents and young adults with cystic fibrosis: A confirmatory factor analysis. *Journal of Pediatric Psychology*, *43*(6), 625–635. https://doi.org/10.1093/jpepsy/jsx155

Saunders, R., Buckman, J. E. J., Cape, J., Fearon, P., Leibowitz, J., & Pilling, S. (2019). Trajectories of depression and anxiety symptom change during psychological therapy. *Journal of Affective Disorders*, *249*, 327–335. https://doi.org/10.1016/j.jad.2019.02.043

Schönberger, M., & Ponsford, J. (2010). The factor structure of the hospital anxiety and depression scale in individuals with traumatic brain injury. *Psychiatry Research*, *179*(3), 342–349. https://doi.org/10.1016/j.psychres.2009.07.003

Shi, D., DiStefano, C., McDaniel, H. L., & Jiang, Z. (2018). Examining chi-square test statistics under conditions of large model size and ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 924–945.

Shi, D., Maydeu-Olivares, A., & Rosseel, Y. (2020). Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 1–15.

Skilbeck, C., Holm, K., Slatyer, M., Thomas, M., & Bell, T. (2011). The factor structure of the hospital anxiety and depression scale (HADS) in a traumatic brain injury (TBI) population. *Brain Impairment*, *12*(1), 22–32. https://doi.org/10.1375/brim.12.1.22

Skilbeck, C., Thomas, M., & Holm, K. (2019). Predicting mood outcome following traumatic brain injury (TBI): PTA & demographic variables. *Brain Impairment*, *21*(1), 65–85. https://doi.org/10.1017/BrImp.2019.13

Smarr, K. L., & Keefer, A. L. (2011). Measures of depression and depressive symptoms: Beck depression inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), geriatric depression scale (GDS), hospital anxiety and depression scale (HADS), and patient health Questionnaire-9 (PHQ-9). *Arthritis Care & Research*, *63*(S11), S454–S466. https://doi.org/10.1002/acr.20556

Smith, A. B., Wright, E., Rush, R., Stark, D., Velikova, G., & Selby, P. (2006). Rasch analysis of the dimensional structure of the hospital anxiety and depression scale. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer*, *15*(9), 817–827. https://doi.org/10.1002/pon.1015

Stott, J., Spector, A., Orrell, M., Scior, K., Sweeney, J., & Charlesworth, G. (2017). Limited validity of the hospital anxiety and depression scale (HADS) in dementia: Evidence from a confirmatory factor analysis: Limited validity of HADS in dementia. *International Journal of Geriatric Psychiatry*, *32*(7), 805–813. https://doi.org/10.1002/gps.4530

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, *5*(1), 1–25. https://doi.org/10.1146/annurev.clinpsy.032408.153639

Sunderland, M., Batterham, P., Calear, A., & Carragher, N. (2019). Self-report scales for common mental disorders: An overview of current and emerging methods. In J. A. Suhr & M. Sellbom (Eds.), *The Cambridge handbook of clinical assessment and diagnosis* (pp. 263–277). Cambridge University Press. https://doi.org/10.1017/9781108235433.019

Thomas, M., McGrath, A., Sugden, N., Weekes, C., & Skilbeck, C. (2021). Investigating the National Adult Reading Test (NART-2) and Wechsler Test of Adult Reading Test (WTAR) in predicting Wechsler Abbreviated Scale of Intelligence–Second edition (WASI-II) scores in an Australian sample. *Australian Psychologist*, 1–10. https://doi.org/10.1080/00050067.2021.1937923

Thomas, M., Sugden, N., McGrath, A., Rohr, P., Weekes, C., & Skilbeck, C. (2020). Investigating the Test of Premorbid Functioning (TOPF) in predicting Wechsler Abbreviated Scale of Intelligence–Second edition (WASI-II) scores in an Australian sample. *Neuropsychological Rehabilitation*, 1–25. https://doi.org/10.1080/09602011.2020.1842213

Trautmann, S., Rehm, J., & Wittchen, H.-U. (2016). The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders? *EMBO Reports*, *17*(9), 1245–1249. https://doi.org/10.15252/embr.201642951

Watson, D. (2005). Rethinking the mood and anxiety disorders: A quantitative hierarchical model for DSM-V. *Journal of Abnormal Psychology*, *114*(4), 522–536. https://doi.org/10.1037/0021-843x.114.4.522

Wechsler, D. (2001). *Wechsler test of adult reading*. Pearson.

Wechsler, D. (2009). *Test of premorbid functioning: Advanced clinical solutions for use with WAIS-IV and WMS-IV*. Pearson.

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Wechsler, D. (2011). *Weschler abbreviated scale of Intelligencen– Second edition (WAIS-II)*. Pearson.

World Health Organisation. (2021). Comprehensive mental health action plan 2013-2030. *World Health Organisation*, 1–40. https://www.who.int/initiatives/mental-health-action-plan-2013-2030

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409–428.

Xie, J., Bi, Q., Li, W., Shang, W., Yan, M., Yang, Y., Miao, D., & Zhang, H. (2012). Positive and negative relationship between anxiety and depression of patients in pain: A bifactor model analysis. *PLoS One*, *7*(10), e47577. https://doi.org/10.1371/journal.pone.0047577

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370.

How to cite this article: Lloyd, M., Sugden, N., Thomas, M., McGrath, A., & Skilbeck, C (2023). The structure of the Hospital Anxiety and Depression Scale: Theoretical and methodological considerations. *British Journal of Psychology*, *114*, 457–475. https://doi.org/10.1111/bjop.12637

20448295, 2023, 2, Downloaded from https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjop.12637 by Technical University Ostrava, Wiley Online Library on [30/04/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License