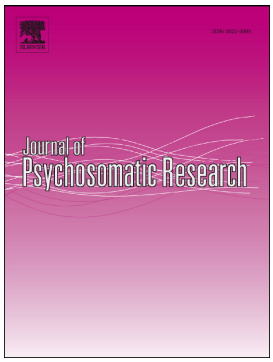

Rasch analysis of the hospital anxiety and depression scale in patients with chronic fatigue syndrome

Emerson J. Bartholomew, Oleg N. Medvedev, Keith J. Petrie, Trudie Chalder

PII: S0022-3999(25)00334-4

DOI: https://doi.org/10.1016/j.jpsychores.2025.112370

Reference: PSR 112370

To appear in: *Journal of Psychosomatic Research*

Received date: 11 July 2025

Revised date: 31 August 2025

Accepted date: 1 September 2025

Please cite this article as: E.J. Bartholomew, O.N. Medvedev, K.J. Petrie, et al., Rasch analysis of the hospital anxiety and depression scale in patients with chronic fatigue syndrome, *Journal of Psychosomatic Research* (2024), https://doi.org/10.1016/j.jpsychores.2025.112370

Rasch Analysis of the Hospital Anxiety and Depression Scale in Patients with Chronic Fatigue Syndrome

Emerson J. Bartholomew[1], Oleg N. Medvedev[2], Keith J. Petrie[1] & Trudie Chalder*[3].

[1]Department of Psychological Medicine, University of Auckland, Auckland, NZ

[2]School of Psychology, University of Waikato, Hamilton, NZ

[3]Department of Psychological Medicine, King's College London, London, UK

*Corresponding author

Trudie Chalder, Department of Psychological Medicine, King's College London, London, UK.

Email: trudie.chalder@kcl.ac.uk

Orcid ID: https://orcid.org/0000-0003-0775-1045

Abstract

Background: The Hospital Anxiety and Depression Scale (HADS) is widely utilized for assessing psychological distress in medical populations, yet its clinimetric properties in chronic fatigue conditions remain underexplored. Given the complex symptom presentation in chronic fatigue syndrome (CFS), rigorous clinimetric validation is essential for accurate clinical assessment.

Objective: This study aimed to evaluate the clinimetric properties of the HADS using Rasch methodology in patients with CFS, with particular emphasis on dimensionality, item functioning, and measurement precision.

Methods: Rasch analysis was conducted on HADS responses from 286 participants diagnosed with CFS. The Partial Credit Rasch model was applied to assess overall model fit, item performance, unidimensionality and differential item functioning.

Results: Initial analysis revealed suboptimal model fit, necessitating subtest modifications to address local response dependence. The subtest solution demonstrated acceptable fit to the Rasch model with evidence of strict unidimensionality, high reliability (PSI = .87), and no differential item functioning

by demographic variables. Rasch-converted interval scores showed improved measurement precision compared to ordinal scoring. Interval scoring yielding a significantly higher mean (M = 22.55, SD = 3.78) compared to unconverted scoring (M = 20.30, SD = 6.87), $t(275) = -19.54$, $p < .001$, indicating that ordinal scoring systematically underestimates the latent trait level. Interval scores showed a 45% reduction in measurement error demonstrated by the substantial reduction in standard error.

Conclusions: The HADS demonstrated acceptable measurement properties in patients with CFS. The development of ordinal-to-interval conversion tables enhances the scale's precision, supporting its continued use in clinical and research contexts.

Keywords: Chronic fatigue syndrome, Invariance, Measurement, Rasch analysis

1. Introduction

Chronic fatigue syndrome (CFS) is a debilitating condition characterized by persistent, unexplained fatigue lasting at least six months, accompanied by cognitive difficulties, post-exertional malaise, and unrefreshing sleep. Psychological comorbidities are notably prevalent, with a substantial number of CFS patients experiencing clinically significant anxiety and/or depression. This high comorbidity rate, ~~combined with substantial symptom overlap between CFS and mood disorders,~~ creates significant assessment challenges. Accurate identification of psychological distress in CFS is crucial for treatment planning, yet the overlapping symptomatology can confound traditional assessment approaches.

The Hospital Anxiety and Depression Scale (HADS) remains one of the most extensively utilized self-report instruments for assessing psychological distress in medical populations [1]. Originally developed to circumvent the confounding effects of somatic symptoms in physically ill patients, the HADS excludes items related to physical manifestations of distress, focusing instead on cognitive and emotional components of anxiety and depression [2, 3]. This has made it particularly

attractive for use in populations where physical symptoms are prominent, including chronic fatigue conditions.

Despite its widespread application, the clinimetric properties of the HADS have been subject to considerable scrutiny and debate. One study documented acceptable reliability coefficients ($\alpha > .80$) and moderate correlations between anxiety and depression subscales ($r = .53 - .56$), suggesting both discriminant validity and shared variance indicative of general psychological distress [4]. However, subsequent investigations have suggested complexities in the scale's dimensional structure, with some studies supporting the original two-factor model while others have identified three-factor solutions or bifactor structures incorporating a general distress factor [5, 6, 7].

The application of modern psychometric approaches, particularly Item Response Theory (IRT) and Rasch analysis, has provided deeper insights into the HADS's measurement properties [8, 9]. While IRT and Rasch methods were developed within psychometric theory, their application to clinical measurement aligns with the clinimetric framework established by Alvan Feinstein, who emphasized the importance of practical clinical utility in measurement instruments [10]. Rasch analysis offers several advantages over classical test theory approaches, including the ability to transform ordinal data to interval-level measurement, assess item-level functioning, and evaluate measurement invariance across different populations [11, 12]. Rasch analyses of the HADS have identified specific items that may function suboptimally, with items 7 ("I can sit at ease and feel relaxed") and 8 ("I feel slowed down") frequently emerging as problematic across different clinical populations [13, 14].

Chronic fatigue syndrome presents unique challenges for psychological assessment [15, 16, 17]. The condition is often comorbid with mood disturbances, creating potential overlap between illness-related symptoms and psychological distress measures [18]. Epidemiological evidence suggests approximately 50% of patients with CFS experience clinically significant anxiety and/or depression, with rates substantially exceeding those in healthy populations and often surpassing other chronic medical conditions [19]. ~~These psychological comorbidities impact treatment outcomes, with systematic reviews suggesting that baseline depression severity moderates therapeutic response to interventions, while comorbid anxiety and depression consistently predict poor functional prognosis alongside prolonged illness duration.~~ Substantial symptom overlap between CFS and psychiatric

conditions creates diagnostic challenges, some studies suggest misdiagnosis rates exceeding 80% when clinicians assume depression without thorough evaluation [20]. The high prevalence of psychological distress, complex symptom presentations, and substantial assessment challenges emphasize the need for clinimetrically sound evaluation tools.

However, empirical evidence regarding the HADS's performance in chronic fatigue syndrome remains unclear. McCue et al. [18] conducted factor analysis of the HADS in a sample of people with CFS and identified an unexpected three-factor structure rather than the intended two-factor solution, leading the authors to question the scale's construct validity in this population. Findings raise important questions about whether the HADS functions as intended when applied to individuals with CFS, particularly given the complex symptom presentation that may blur traditional boundaries between physical and psychological illnesses.

The importance of rigorous clinimetric evaluation in chronic fatigue syndrome is underscored by the clinical significance of accurately identifying and monitoring psychological distress in this condition. Anxiety and depression are commonly comorbid with chronic fatigue syndromes and can significantly impact quality of life, treatment response, and functional recovery [21]. Inaccurate assessment tools may lead to inappropriate clinical decisions, including under-recognition of psychological conditions or over-pathologizing normal adjustment responses to chronic illness. Rasch methodology has demonstrated the potential for enhancing measurement precision through ordinal-to-interval conversion, which can improve the accuracy of statistical analyses and enable more valid comparisons with other interval-level measures [22].

This study aimed to conduct a comprehensive Rasch analysis of the HADS in a sample of patients with CFS. The primary objectives were to: (1) evaluate the overall fit of the HADS to the Rasch measurement model; (2) assess the dimensionality and item-level functioning of the scale; (3) examine potential differential item functioning across demographic variables; and (4) if fit to the Rasch model is evident, develop an ordinal-to-interval conversion algorithm to enhance measurement precision for future clinical and research applications.

2. Method

2.1 Participants

The current sample comprised 286 participants with confirmed CFS diagnosis resulting from medical assessments conducted in an out-patient clinic in South London. Those who fulfilled NICE criteria for CFS were included in this study. Participants were included if they were >18 years of age. Additional inclusion criteria required participants to have a documented diagnosis of chronic fatigue syndrome by a qualified healthcare professional, possess adequate English language proficiency to complete self-report measures independently, and provide informed written consent for data to be collected routinely. Exclusion criteria included active psychosis or severe cognitive impairment that would preclude a diagnosis of CFS, as assessed through clinical interview. Participants completed questionnaires after the specialist assessment by a clinician and prior to starting treatment.

2.2 Procedure

Data collection was conducted in accordance with institutional ethical guidelines. Participants completed the HADS as part of a comprehensive clinical assessment in the outpatient clinic setting. Clinical and demographic information was provided by participants. Collection of data was approved by the audit committee of the South London and Maudsley NHS Foundation Trust. Data was collected routinely and anonymized. This data was collected from November 2011 to March 2013.

2.3 Measures

The Hospital Anxiety and Depression Scale (HADS) is a 14-item self-report questionnaire comprising two 7-item subscales designed to assess anxiety (HADS-A) and depression (HADS-D) symptoms over the past week [2]. Each item is scored on a 4-point Likert scale (0-3), yielding subscale scores ranging from 0-21, with higher scores indicating greater symptom severity. The HADS was specifically designed for use in medical settings and deliberately excludes somatic symptoms that might be confounded with physical illness, focusing instead on cognitive and emotional manifestations of anxiety and depression.

The anxiety subscale includes items assessing worry, tension, panic, and restlessness, while the depression subscale focuses on anhedonia, loss of interest, and reduced positive affect. Several items

are reverse-scored to minimize response bias. Conventional interpretation guidelines suggest scores of 0-7 indicate normal levels, 8-10 indicate mild symptoms, 11-14 indicate moderate symptoms, and 15-21 indicate severe symptoms, though these cut-points have been subject to considerable debate [4].

The HADS has demonstrated acceptable clinimetric properties across diverse populations, with reported internal consistency coefficients typically exceeding .80 for both subscales and test-retest reliability coefficients generally above .75 [23]. However, its performance in specific clinical populations, particularly those with chronic fatigue conditions, has received limited empirical attention despite its frequent use in such contexts.

2.4 Data Analyses

Data analysis was conducted using IBM SPSS version 29.0 for preliminary descriptive statistics and RUMM2030plus software for Rasch analysis [24]. As indicated by significant likelihood ratio test rejecting the Rating Scale Model [25], the Partial Credit Rasch model [26] was the most appropriate for the polytomous response structure of the HADS. This model allows threshold parameters to vary freely across items rather than constraining them to be equal as required by assumptions of the Rating Scale model.

Rasch analysis follows an iterative process of model evaluation and refinement, guided by established fit criteria and theoretical considerations through seven key stages:

(1) Overall model fit was assessed using the item-trait interaction chi-square statistic, with non-significant values ($p > .05$) indicating an acceptable fit (no significant deviation from the model expectations). An acceptable fit enables transformation of raw (ordinal) scores into interval-level measurement.

(2) Individual item fit was evaluated using standardized residuals, with values between ±2.5 considered acceptable.

(3) Smith's t-test was employed to assess unidimensionality, with values below 5% providing evidence for strict unidimensionality [27].

(4) Response category functioning was examined through inspection of category probability curves to identify potential threshold disorders that might indicate problematic response category performance.

(5) Local response dependence was assessed through examination of residual correlations between items, with values exceeding .20 considered indicative of potential dependency that could threaten the assumption of local independence fundamental to Rasch modeling [28]. Items are considered locally dependent when they covary for reasons unrelated to the latent trait, distorting dimensionality and artificially inflating reliability indices [29]. When local dependency is identified, items may be considered for removal or joined into subtests which may resolve this issue while preserving item content [30, 31].

(6) Differential item functioning (DIF) analysis was conducted to assess measurement invariance across demographic subgroups, including sex (male/female) and age categories. Age data were stratified into three balanced categories (18-32, 33-46, and 47+). The presence of DIF would indicate that individuals with equivalent levels of the latent trait, but from different subgroups, have different probabilities of endorsing particular items, potentially compromising the fairness and validity of comparisons across groups.

(7) Reliability was assessed using multiple indices, including Cronbach's alpha, McDonald's omega, and the Person Separation Index (PSI). PSI reflects the scale's ability to differentiate between individuals with different levels of the latent trait, with values above .70 considered acceptable for group-level comparisons and values above .85 preferred for individual-level assessment. A satisfactory fit to the Rasch model permits the creation of an ordinal to interval conversion algorithm. A paired-samples *t*-test is used to evaluate the benefits of conversion particularly in standard error reduction and correlation coefficients between scoring methods.

3. Results

The study sample comprised 286 participants with chronic fatigue syndrome (27.4% male, 62.7% female) aged 18-73 years (M = 39.48, SD = 12.04). Initial analysis of the 14-item HADS revealed poor overall model fit with a significant item-trait interaction chi-square value (χ^2 = 125.47, *df* = 42, $p < .01$) (Table 1). Individual item analysis showed item 8 ("I still enjoy the things I used to enjoy") (fit

residual = 3.23, $p < .01$) and item 11 ("I feel as if I am slowed down") had substantial misfit (fit residual = 4.94, $p < .01$) (Table 2). To investigate whether the observed item misfit was related to symptom overlap with fatigue, correlational analyses were conducted between the two misfitting items and total fatigue severity (as measured by the Chalder Fatigue Questionnaire). Item 8 ("I still enjoy the things I used to enjoy") demonstrated a significant moderate correlation with fatigue severity ($r = .488$, $p < .001$), while item 11 ("I feel as if I am slowed down") showed no significant association ($r = .034$, $p = .556$).

Examination of the residual correlation matrix revealed evidence of local response dependency, with correlations exceeding the .20 threshold between eight different items for a total of 14 problematic pairs. Initial Smith's [27] *t*-test for unidimensionality yielded 5.24% of significant t-tests, failing to meet the criterion for strict unidimensionality (< 5%). Reliability was excellent, with a PSI of .87, Cronbach's alpha of .87, and McDonald's omega of .88 observed. Person fit statistics showed acceptable functioning (mean = -.12, SD = 1.07), suggesting that participants generally responded in patterns consistent with the underlying latent trait.

To address the identified local response dependency while preserving item content and maintaining construct validity, subtest methodology was implemented following established protocols [30]. Items demonstrating high residual correlations and conceptual similarity were combined into subtests based on both empirical evidence and theoretical considerations. The final subtest configuration comprised: Subtest 1, incorporating items 3, 5, 9, 11, and 12, representing a broader range of anxiety and distress-related symptoms; Subtest 2, combining items 4, 6, and 8, focusing primarily on positive affect and relaxation capacity; and six individual items (1, 2, 7, 10, 13, 14) that remained separate due to their distinct content domains and acceptable individual functioning. Prior to this final subtest solution being chosen, a different solution was tested based on the original two-factor conceptualization of the HADS. Here, we first investigated whether the observed misfit could be resolved by creating two subtests corresponding to the theoretical anxiety and depression subscales. This configuration, with Subtest 1 containing all seven anxiety items (1, 3, 5, 7, 9, 11, 13) and Subtest 2 containing all seven depression items (2, 4, 6, 8, 10, 12, 14), achieved good model fit ($\chi^2 = 4.09$, $df = 6$, $p = .664$) and met the criterion for strict unidimensionality. However, this solution yielded poor

reliability (PSI = .65), indicating insufficient ability to differentiate between individuals with varying latent trait levels. Due to the poor reliability observed, we proceeded with the aforementioned approach based on residual correlations to optimize both model fit and reliability.

Our final subtest configuration (ST1: 3, 5, 9, 11, and 12, ST2: 1, 2, 7, 10, 13, 14) successfully resolved the local dependence issues while maintaining the clinical interpretability of the scale (Table 1). Following subtest modifications, the HADS demonstrated improved clinimetric properties across a range of parameters, such as an acceptable overall fit to the Rasch model evidenced by the item-trait interaction chi-square indicating no significant deviations from the model expectations (χ^2 = 58.35, *df* = 48, *p* = .146). Individual subtest and item fit statistics showed all subtests and individual items demonstrating fit residuals within acceptable ranges (M = .44, SD = 1.18) and no items or subtests showing significant misfit in the final configuration (Table 2).

Smith's test of the final subtest configuration yielded significant *t*-tests below 1% providing evidence of strict unidimensionality, indicating that a single latent trait adequately explained the observed response patterns (Table 1). Findings support the conceptualization of the HADS as measuring a unified construct of psychological distress. Thus, principal component analysis of residuals confirmed the unidimensional structure, with the first residual component explaining minimal additional variance beyond that accounted for by the primary Rasch dimension.

DIF analysis was conducted across sex and age categories (18-32, *n* = 95; 33-46, *n* = 98; 47+, *n* = 93). No evidence of uniform or non-uniform DIF was detected for any subtests or individual items across sex groups, indicating that the HADS functions equivalently for male and female participants with CFS. Similarly, age-based DIF analysis showed no significant differential functioning, supporting the measurement invariance of the scale by age. Observed score differences between demographic subgroups can therefore be attributed to genuine differences in the underlying psychological distress levels rather than measurement bias. The final HADS configuration demonstrated no changes in the robust reliability observed in the initial analysis (PSI = .87).

Figure 1 shows the person-item threshold distribution for the final testlet analysis. A test information function is over-laid on the figure showing the scale provides optimal measurement precision at .17 logits (test information function = 9.05, equivalent reliability = .90). High measurement

precision is seen across a range of -1.1 to +1.8 logits, encompassing approximately 78 % of the sample and suggesting the HADS provides particularly high-precision measurement at moderate levels of psychological distress. The distribution demonstrates excellent sample targeting with 96 % of participants falling within the effective measurement range of the instrument, with no floor or ceiling effects observed.

Analysis of response categories revealed ordered thresholds for the majority of items, indicating that respondents could generally distinguish between the four response options (0-3). Both subtests showed properly ordered response categories observed in figure 2. However, disordering was identified in both initial and final analyses for response option 2/category 3 in items 2 and 14 (figure 3). Dysfunction for item 2 ('I still enjoy things I used to enjoy') option 'Only a little' suggested respondents experienced difficulty distinguishing between this option and its adjacent categories 'Not quite so much' and 'Hardly at all,' likely due to insufficient semantic distinction between these adjacent response options. Similarly, Item 14 ('I can enjoy a good book or radio or TV program') exhibited disordering for 'Not often' in the progression from 'Sometimes' to 'Not often' to 'Very seldom', suggesting respondents struggled to differentiate between these categories likely due to their semantic similarity.

Despite these localized threshold issues, category probability curves demonstrated adequate separation for most items, with each response category maintaining a distinct range where it was the most probable response. The average category measure advanced monotonically from -2.14 logits (category 0) to 2.87 logits (category 3), supporting the overall functionality of the response scale structure. We decided against collapsing disordered threshold to retain the original four-category response format as the threshold disordering was limited to only two categories out of 56 categories and the overall response scale demonstrated acceptable functionality as evidenced by the monotonic progression of average category measures.

Successful model fit enabled the development of comprehensive ordinal-to-interval conversion tables, facilitating the transformation of raw ordinal HADS scores to interval-level data. A paired-samples t-test was conducted to evaluate the precision benefits of this conversion approach, comparing unconverted ordinal scores with Rasch-converted interval scores. The analysis revealed significant differences between scoring approaches, with interval scoring yielding a significantly higher mean (M

= 22.55, SD = 3.78) compared to unconverted scoring (M = 20.30, SD = 6.87), t(275) = -19.54, p < .001. The mean difference of -2.25 (SE = 0.12, 95% CI [-2.48, -2.02]) indicates that ordinal scoring systematically underestimates the latent trait level. The correlation between scoring methods was high (r = .993, p < .001), demonstrating strong concordance while highlighting the systematic bias inherent in ordinal approaches. The standard error associated with interval scores (SE = 3.78) was substantially lower than that of ordinal scores (SE = 6.87), representing a 45% reduction in measurement error. The provided ordinal-to-interval conversion table enables researchers and clinicians to transform traditional HADS total scores (ranging 0-42) into interval-level measurements that better reflect the underlying psychological distress construct (Table 3).

4. Discussion

The current study provides evidence for the clinimetric viability of the HADS in patients with CFS by achieving Rasch model fit while preserving construct validity. Initial analysis revealed substantial local response dependence across multiple item pairs, threatening the fundamental assumptions required for valid measurement. A large number of these locally dependent item pairs were observed between reverse scored items, suggesting a method factor independent of the intended construct. Respondents may be processing these items differently, potentially due to the cognitive burden of switching between positively and negatively framed statements. This finding is in line with recent clinimetric research that has suggested reversed items may paradoxically increase measurement error rather than reduce it [32]. Regarding observed item disfunction, previous investigations have resorted to item deletion, or complete abandonment of the instrument in the face of poor clinimetric indices. The former approach, while expedient, fundamentally compromises construct validity by removing theoretically essential content domains. The latter approach, eliminates a valuable clinical tool, as McCue et al. [18] recommended following their identification of aberrant factor structure that "fundamentally compromised" the HADS's clinical utility. Our investigation employed subtest methodology as a solution that strategically combined locally dependent items into coherent measurement units based on empirical evidence and theoretical considerations. This approach successfully resolved the statistical violations driving initial model misfit while preserving all 14

original items and their complete clinical content. The resulting configuration demonstrated superior clinimetric properties, facilitating the development of ordinal-to-interval conversion tables that transform ordinal HADS scores into precise interval-level data with substantially reduced measurement error and enhanced statistical power.

Initial findings of poor model fit and problematic item functioning align with previous Rasch investigations across diverse clinical populations, particularly the consistent identification of items 8 and 11 as problematic [13, 14]. Correlational analysis showed item 8 was associated with fatigue severity ($r = .488$), while item 11 was not, suggesting item 8's misfit reflected symptom overlap, where patients may report reduced enjoyment due to fatigue-related activity limitations rather than depressive anhedonia. This suggests that some HADS items may conflate illness-related functional limitations with psychological distress, particularly for anhedonia-related content where fatigue-induced inactivity could masquerade as mood symptoms.

The achievement of strict unidimensionality has implications for understanding psychological distress in CFS. While the HADS was designed to measure distinct anxiety and depression constructs, our findings suggest that a single latent trait adequately explains response patterns in this population, indicating these symptoms may be less differentiated than in general medical samples. This convergence potentially reflects the complex interplay between physical symptoms and psychological distress that characterizes chronic fatigue syndromes. This interpretation aligns with contemporary models of chronic illness that emphasize the bidirectional relationships between somatic and psychological symptoms [33]. Further, this convergence of typically distinct psychological constructs may reflect central sensitization mechanisms characteristic of chronic fatigue syndromes. Central sensitization, defined as an amplification of neural signaling within the CNS that elicits pain hypersensitivity, has been documented in CFS and represents a plausible neurobiological explanation for the altered processing of both physical and psychological symptoms [34]. Within this framework, the amplified neural signaling could diminish the traditional boundaries between anxiety and depression symptoms, resulting in a more unified psychological distress presentation. For targeted assessment of psychological distress, shorter instruments designed explicitly for distress measurement might be preferable. The choice should be guided by the specific clinical or research objectives, recognizing that

in CFS populations, the HADS appears to function most effectively as a measure of overall psychological distress rather than capturing distinct anxiety and depression dimensions.

The absence of differential item functioning across demographic variables provides crucial evidence for measurement fairness and supports the equitable application of the HADS across diverse subgroups within chronic fatigue populations. This finding is particularly significant given the heterogeneous nature of chronic fatigue presentations and the potential for symptom expression to vary across age and sex categories. The demonstrated measurement invariance allows clinicians and researchers to have confidence in comparisons.

The substantial improvement in measurement precision achieved through ordinal-to-interval conversion represents a significant methodological contribution with implications for chronic fatigue research. The 45% reduction in standard error enhances statistical power and enables more sensitive detection of treatment effects and more accurate characterization of symptom trajectories over time. This precision enhancement is particularly valuable in chronic fatigue populations, where subtle changes in psychological distress may have profound clinical implications for functional recovery and quality of life. Raw scores are converted through a logit-based transformation that accounts for the non-linear relationship between ordinal response patterns and the underlying latent trait. For practical application, users can reference the conversion table by matching their obtained raw score to the corresponding interval value (Table 3). For example, a raw score of 20 converts to an interval score of 22.61, while a raw score of 25 converts to 24.94. This transformation not only improves measurement precision but also enables valid parametric statistical analyses and meaningful comparisons with other interval-level psychological measures.

From a clinical perspective, the refined HADS offers healthcare providers a clinimetrically sound instrument for assessing psychological distress in chronic fatigue patients. The availability of interval-level scoring facilitates the investigation of relationships between psychological distress and physiological associations of fatigue and sleep in chronic fatigue research. Interval converted data are necessary to make comparisons between psychological constructs and physiological markers valid, as ordinal data cannot be meaningfully added and subtracted, thus making it unsuitable for comparisons with physical measures and inappropriate for use in parametric statistical tests.

Future research directions emerging from these findings include longitudinal validation of the refined HADS to assess stability of the clinimetric properties over time, investigation of responsiveness to treatment interventions, and exploration of the relationship between interval-converted HADS scores and objective physiological markers of distress. Additionally, cross-validation in independent chronic fatigue samples and examination of the generalizability of the subtest approach to other medical populations would strengthen the evidence base for these methodological innovations.

The clinical implications of this work extend to treatment outcome research, where precise measurement of psychological distress is essential for evaluating intervention efficacy. The enhanced measurement precision achieved through Rasch analysis may facilitate detection of treatment effects that would be obscured by the measurement error inherent in traditional ordinal scoring approaches. This capability is particularly relevant for chronic fatigue populations, where treatment effects may be clinically meaningful.

5. Limitations of this study

Several limitations warrant consideration when interpreting these findings. The cross-sectional design precludes examination of the temporal stability of the scale and responsiveness to clinical change over time. Longitudinal validation would strengthen confidence in the reliability and clinical utility of the refined HADS configuration across different stages of chronic fatigue conditions and treatment trajectories. The sample composition, while representative of clinical chronic fatigue syndrome, demonstrates diagnostic heterogeneity that may limit generalizability to more homogeneous diagnostic groups. Future research should examine the stability of the HADS properties in fatigue across diagnostic categories and in samples with different demographic characteristics.

Compliance with ethical standards

Collection of data was approved by the audit committee of the South London and Maudsley NHS Foundation Trust. Data is collected routinely and anonymized.

Fundings

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. TC receives salary support from the National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre at South London and Maudsley NHS Foundation Trust

and King's College London. The views expressed in this article are those of the authors and not necessarily those of the NIHR or the NHS.

References

1. L. Zhang, T. Fu, R. Yin, Q. Zhang, B. Shen, Prevalence of Depression and Anxiety in Systemic Lupus Erythematosus: A Systematic Review and MetaAnalysis, BMC Psychiatry 17 (2017) 70.
2. A.S. Zigmond, R.P. Snaith, The Hospital Anxiety and Depression Scale, Acta Psychiatr. Scand. 67 (1983) 361-370.
3. R.P. Snaith, The Hospital Anxiety and Depression Scale, Health Qual. Life Outcomes 1 (2003) 29.
4. I. Bjelland, A.A. Dahl, T.T. Haug, D. Neckelmann, The validity of the Hospital Anxiety and Depression Scale: An updated literature review, J. Psychosom. Res. 52 (2002) 69-77.
5. T.D. Cosco, F. Doyle, M. Ward, H. McGee, Latent structure of the Hospital Anxiety and Depression Scale: A 10-year systematic review, J. Psychosom. Res. 72 (2012) 180-184.
6. L. Iani, M. Lauriola, M. Costantini, A confirmatory bifactor analysis of the Hospital Anxiety and Depression Scale in an Italian community sample, Health Qual. Life Outcomes 12 (2014) 84.
7. S. Norton, T. Cosco, F. Doyle, J. Done, A. Sacker, The Hospital Anxiety and Depression Scale: A meta confirmatory factor analysis, J. Psychosom. Res. 74 (2013) 74-81.
8. X.J. Huang, H.Y. Ma, X.M. Wang, J. Zhong, Y.Q. Chen, M.Z. Xu, Rasch analysis of the Hospital Anxiety and Depression Scale (HADS) in patients with major depression, An. Psicol. 39 (2023) 395-404.
9. S.D. Lambert, J.F. Pallant, A.W. Boyes, M.T. King, B. Britton, A. Girgis, A Rasch analysis of the Hospital Anxiety and Depression Scale (HADS) among cancer survivors, Psychol. Assess. 25 (2013) 379-390.
10. A.R. Feinstein, Clinimetric perspectives, J. Chronic Dis. 40 (1987) 635-640.
11. G. Rasch, Probabilistic models for some intelligence and attainment tests, Danish Institute for Educational Research, Copenhagen, 1960.
12. A. Tennant, P.G. Conaghan, 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 Rheum. 57 (2007) 1358-1362.

13. S.D. Lambert, J.F. Pallant, A. Girgis, Rasch analysis of the Hospital Anxiety and Depression Scale among caregivers of cancer survivors: Implications for its use in psycho-oncology, Psychooncology 20 (2011) 919-925.
14. F. Pallant, A. Tennant, An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS), Br. J. Clin. Psychol. 46 (2007) 1-18.
15. K. Fukuda, S.E. Straus, I. Hickie, M.C. Sharpe, J.G. Dobbins, A. Komaroff, The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group, Ann. Intern. Med. 121 (1994) 953-959.
16. National Institute for Health and Care Excellence, Chronic fatigue syndrome/myalgic encephalomyelitis (or encephalopathy): Diagnosis and management, NICE Clinical Guideline CG53, NICE, London, 2007.
17. M.C. Sharpe, L.C. Archard, J.E. Banatvala, L.K. Borysiewicz, A.W. Clare, A. David, et al., A report--chronic fatigue syndrome: guidelines for research, J. R. Soc. Med. 84 (1991) 118-121.
18. P. McCue, C.R. Martin, T. Buchanan, J. Rodgers, A.B. Scholey, An investigation into the psychometric properties of the Hospital Anxiety and Depression Scale in individuals with chronic fatigue syndrome, Psychol. Health Med. 8 (2003) 425-439.
19. K.A. Janssens, W.L. Zijlema, M.L. Joustra, J.G. Rosmalen, Mood and Anxiety Disorders in Chronic Fatigue Syndrome, Fibromyalgia, and Irritable Bowel Syndrome: Results From the LifeLines Cohort Study, Psychosom. Med. 77 (2015) 449-457.
20. J.P. Griffith, F.A. Zarrouf, A systematic review of chronic fatigue syndrome: Don't assume it's depression, Prim. Care Companion J. Clin. Psychiatry 10 (2008) 120-128.
21. J. Daniels, A. Brigden, A. Kacorova, Anxiety and depression in chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): Examining the incidence of health anxiety in CFS/ME, Psychol. Psychother. 90 (2017) 502-509.
22. O.N. Medvedev, C.U. Krägeloh, Rasch measurement model, in: O.N. Medvedev, C.U. Krägeloh, R.J. Siegert, N.N. Singh (Eds.), Handbook of assessment in mindfulness research, Springer, Cham, 2022, pp. 81-109.
23. C. Herrmann, International experiences with the Hospital Anxiety and Depression Scale: A review of validation data and clinical results, J. Psychosom. Res. 42 (1997) 17-41.
24. D. Andrich, B. Sheridan, G. Luo, RUMM 2030, RUMM Laboratory, Perth, 2009.
25. D. Andrich, A rating formulation for ordered response categories, Psychometrika 43 (1978) 561-573.
26. G.N. Masters, A Rasch model for partial credit scoring, Psychometrika 47 (1982) 149-174.
27. E.V. Smith, Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals, J. Appl. Meas. 3 (2002) 205-231.

28. K.B. Christensen, G. Makransky, M. Horton, Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations, Appl. Psychol. Meas. 41 (2017) 178-194.
29. B.D. Wright, Local dependency, correlations and principal components, Rasch Meas. Trans. 10 (1996) 509-511.
30. Å. Lundgren-Nilsson, I.H. Jonsdottir, G. Ahlborg, A. Tennant, Construct validity of the psychological general well-being index (PGWBI) in a sample of patients undergoing treatment for stress-related exhaustion: A Rasch analysis, Health Qual. Life Outcomes 11 (2013) 2.
31. I. Marais, D. Andrich, Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model, J. Appl. Meas. 9 (2008) 105-124.
32. H.S. Pedersen, K.S. Christensen, A. Prior, K.B. Christensen, The dimensionality of the Perceived Stress Scale: The presence of opposing items is a source of measurement error, J. Affect. Disord. 344 (2024) 485-494.
33. C. Chen, H.T. Yip, K.H. Leong, W.C. Yao, C.L. Hung, C.H. Su, et al., Presence of depression and anxiety with distinct patterns of pharmacological treatments before the diagnosis of chronic fatigue syndrome: a population-based study in Taiwan, J. Transl. Med. 21 (2023) 98.
34. J. Nijs, S.Z. George, D.J. Clauw, C. Fernández-de-Las-Peñas, E. Kosek, K. Ickmans, et al., Central sensitisation in chronic pain conditions: Latest discoveries and their potential for precision medicine, Lancet Rheumatol. 3 (2021) e383-e392.

Figure 1. *Person-item threshold distribution for the HADS final analysis.*

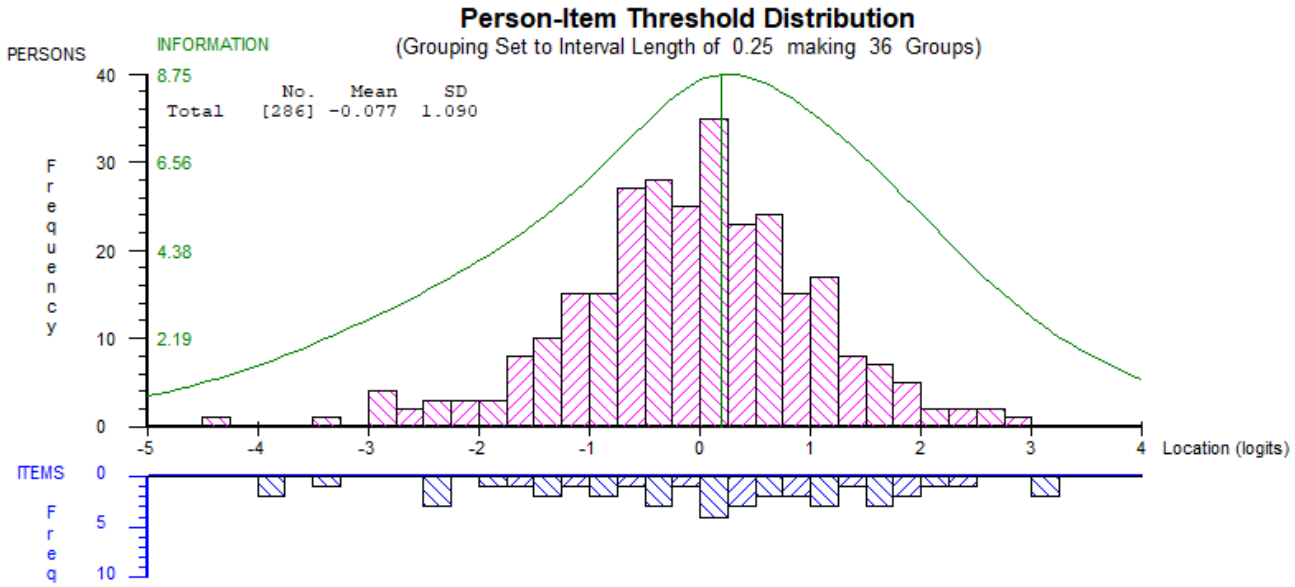

Figure 2. *Item characteristic curves for subtest 2 in the final analysis.*

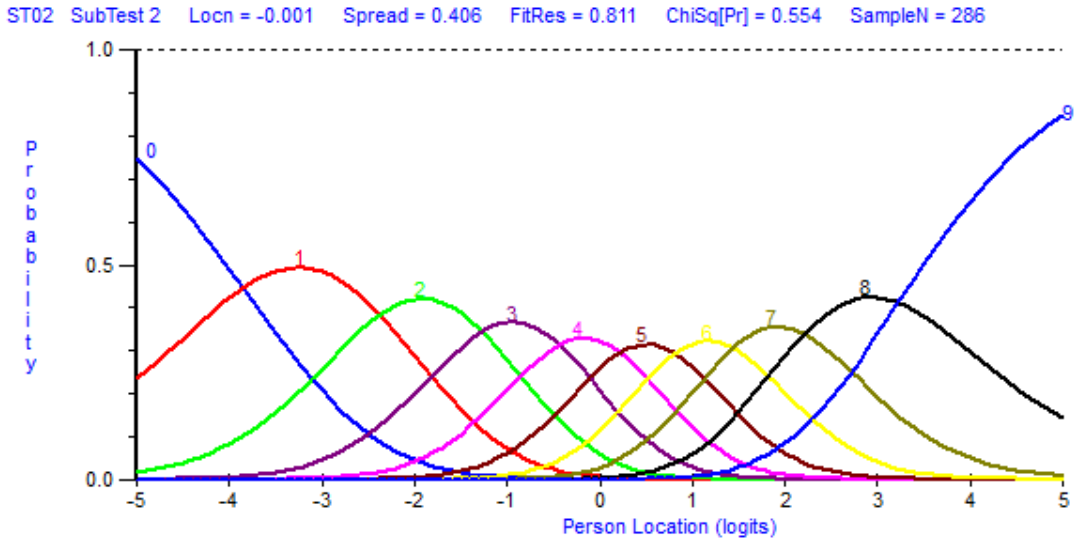

Figure 3. *Item characteristic curves for items 2 (top) and 14 (bottom) in the final analysis.*

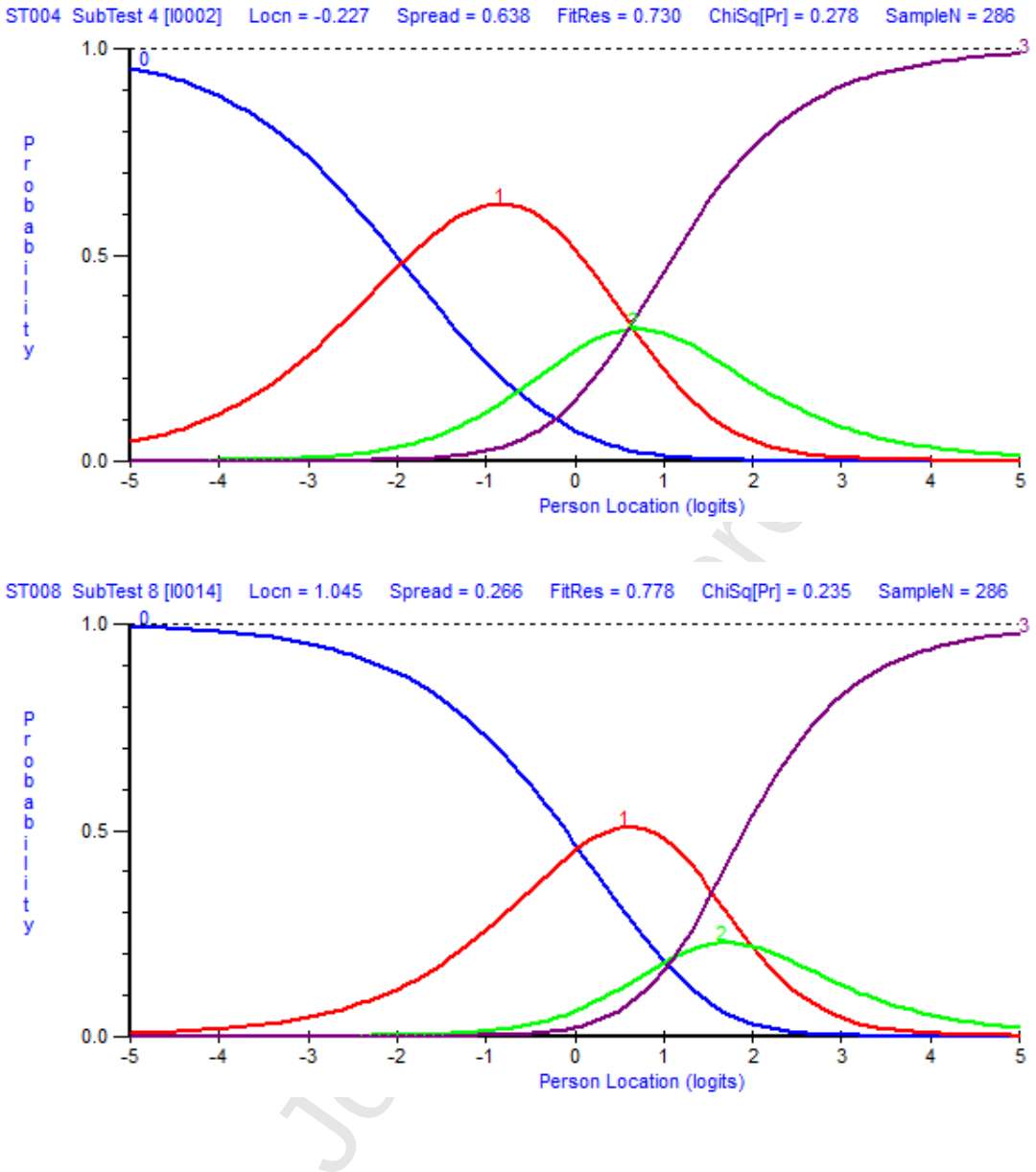

Table 1. *Summary of the Rasch model fit statistics for the initial and final analysis of the Hospital Anxiety and Depression scale (n=286).*

	Item fit residual		Person fit residual		Goodness of fit		PSI	Unidimensionality (Strict)
Analyses	Mean	SD	Mean	SD	χ^2 (*df*)	*p*		Sig. *t*-test in %
Initial	.30	1.95	-.12	1.07	125.47 (42)	<.01	.87	5.2(NO)
Final	.44	1.18	-.08	1.09	58.58 (48)	.146	.87	<1% (YES)

Table 2. *Initial items fit statistics of Hospital Anxiety and Depression Questionnaire (n=350).*

No.	Item content	Location	Fit Residual	Chi Square
	Initial Analysis			
1	I feel tense or 'wound up'.	-1.09	-0.46	0.13
2	I get a sort of frightened feeling as if something awful is about to happen.	-0.29	-0.12	0.06
3	Worrying thoughts go through my mind.	-0.12	0.55	0.97
4	(R) I can sit at ease and feel relaxed.	1.39	-1.52	0.01
5	I get a sort of frightened feeling like 'butterflies' in the stomach.	-0.53	-0.35	0.76
6	I feel restless as I have to be on the move.	0.7	-1.7	0.03
7	I get sudden feelings of panic.	-0.35	-1.17	0.01
8	(R) I still enjoy the things I used to enjoy.	-1.53	3.23*	0.01*
9	(R) I can laugh and see the funny side of things.	0.29	0.72	0.12
10	(R) I feel cheerful.	0.31	2.11	0.13
11	I feel as if I am slowed down.	0.07	4.94	0.01
12	I have lost interest in my appearance.	0.06	-0.68	0.04
13	(R) I look forward with enjoyment to things.	0.11	-1.69	0.04
14	(R) I can enjoy a good book or radio or TV program.	0.97	0.35	0.09
	Final Analysis			
	Subtest 1 (3+5+9+11+12)	-0.02	-0.36	0.91
	Subtest 2 (4+6+8)	-0.01	0.81	0.55
	Item 1	-1.04	0.26	0.32
	Item 2	-0.23	0.73	0.28
	Item 7	-0.31	-0.37	0.08
	Item 10	0.38	2.81	0.27
	Item 13	0.16	-1.14	0.13
	Item 14	1.04	0.78	0.24

Note: Fit Resid= Fit Residual; Chi Sq= Chi Square; *Significant item misfit to the Rasch model $p<0.05$

Table 3. *Ordinal to interval score conversion table for the HADS*

Ordinal	Interval Scores	
Raw Scores	Logits	Scale (0-42)
0	-5.39	0.00
1	-4.46	3.89
2	-3.78	6.70
3	-3.30	8.73
4	-2.91	10.35
5	-2.58	11.73
6	-2.29	12.93
7	-2.03	14.01
8	-1.80	14.99
9	-1.59	15.89
10	-1.39	16.72
11	-1.20	17.49
12	-1.03	18.20
13	-0.87	18.87
14	-0.72	19.49
15	-0.59	20.07
16	-0.45	20.62
17	-0.33	21.14
18	-0.21	21.65
19	-0.09	22.13
20	0.02	22.61
21	0.13	23.07
22	0.24	23.54
23	0.36	24.00
24	0.47	24.47
25	0.58	24.94
26	0.70	25.43
27	0.82	25.93
28	0.94	26.44
29	1.06	26.96
30	1.20	27.51
31	1.33	28.08
32	1.47	28.67
33	1.62	29.30
34	1.78	29.96
35	1.95	30.68
36	2.14	31.47
37	2.36	32.36
38	2.60	33.40
39	2.90	34.65
40	3.29	36.25
41	3.85	38.59
42	4.66	42.00

Note. Conversion table can only be used for respondents with no missing data.

Declarations of interest:

TC is part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust, King's College London. She has grants from Guy's St Thomas Charity Grants, NIHR and UKRI for post COVID syndromes.
She has received travel expenses, accommodation costs, and honorarium for several lectures in Europe and has received travel expenses and accommodation costs for attending American Thoracic Society Conference. She was on the Expert Advisory Panel for Covid-19 Rapid Guidelines.

Highlights

- Rasch analysis validated the HADS in patients with CFS
- Content preserved, subtests resolved local dependence issues without item deletion
- Ordinal-to-interval conversion, 45% reduction in measurement error
- No differential item functioning across gender and age groups detected