ELSEVIER

Journal of Clinical Epidemiology ■ (2011) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Rasch analysis supports the use of the Depression, Anxiety, and Stress Scales to measure mood in groups but not in individuals with chronic low back pain

Luke Parkitny[a,b,*], James Henry McAuley[a], David Walton[c], Leonardo Oliveira Pena Costa[d,g], Kathryn M. Refshauge[e], Benedict Martin Wand[f], Flavia Di Pietro[a,b], G. Lorimer Moseley[a,h]

[a]*Neuroscience Research Australia, University of New South Wales, PO Box 1165, Randwick, NSW 2031, Australia*
[b]*Prince of Wales Clinical School, University of New South Wales, PO Box 1165, Randwick, NSW 2031, Australia*
[c]*Faculty of Health Sciences, The University of Western Ontario, Ontario N6G 1A1, Canada*
[d]*Universidade Cidade de São Paulo, SP 03071-000, Brazil*
[e]*Faculty of Health Sciences, The University of Sydney, Sydney, NSW 2006, Australia*
[f]*School of Health Sciences, The University of Notre Dame Australia, Fremantle, WA 6959, Australia*
[g]*Musculoskeletal Division, The George Institute for Global Health, Sydney, Australia*
[h]*Sansom Institute for Health Research, The University of South Australia, Adelaide, 5000, Australia*

Accepted 23 May 2011

Abstract

Objective: Chronic pain is a common problem that is associated with mood disorders such as depression. The Depression, Anxiety, and Stress Scales (DASS-21) questionnaire is commonly used to help measure disordered mood. In this study, we used Rasch analysis to analyze the clinimetric properties of the DASS-21 in a chronic low back pain sample.

Study Design and Setting: A Rasch analysis was conducted on data collected as a part of a randomized hospital-based placebo-controlled trial. DASS-21 questionnaires were completed by the 154 enrolled participants.

Results: The DASS-21 subscales fit the Rasch model. No differential item functioning was detected for age, gender, pain severity, or disability. Reliability for individual use was supported for the depression subscale (Person Separation Index [PSI] = 0.86) but group use only for the anxiety (PSI = 0.74) and stress (PSI = 0.82) subscales. A DASS-21 aggregate score of ''negative affect'' lacked fit to the Rasch model ($\chi^2 = 191.48$, $P < 0.001$).

Conclusion: This is the first study that used Rasch analysis to demonstrate that the DASS-21 subscales demonstrate adequate measurement properties for research involving groups with chronic pain. Only the DASS-21 depression subscale demonstrated adequate reliability for use with individuals with chronic pain. The use of a single DASS-21 aggregate score as a measure of ''negative affect'' was not supported.

Keywords: Pain; Mood; Instrument; Questionnaire; Clinimetric; Reliability

1. Introduction

Depression is common in people with chronic pain [1−6] and appears to play a more significant role than physical factors in the transition from acute to chronic pain [7]. In chronic pain, high levels of stress or anxiety are associated with more severe pain and disability, and worse clinical and work-related outcomes [8−11]. Correspondingly, treatment of mood disorders in patients with chronic pain reduces pain and increases quality of life [12−14].

Symptoms of depression are not always recognized in routine clinical practice [15]. Practitioners may fail to screen for depression or might use insensitive diagnostic strategies [16−20]. Consequently, some patients may receive suboptimal management of their chronic pain and related problems. Several instruments have been designed to recognize and measure symptoms of depression, for example, the Beck Depression Inventory [21], the Zung Self-Rating Depression Scale [22], and the Hamilton Rating Scale for Depression [23]. The presence of chronic pain is thought to make interpretation of these scales difficult—patients may endorse certain items for reasons related to

* Corresponding author. Neuroscience Research Australia, University of New South Wales, PO Box 1165, Randwick, NSW 2031, Australia. Tel.: +612-9399-1000; fax: +612-9399-1082.

E-mail address: L.Parkitny@neura.edu.au (L. Parkitny).

0895-4356/$ - see front matter
doi: 10.1016/j.jclinepi.2011.05.010

What is new?

- The clinimetric properties of the Depression, Anxiety, and Stress Scales (DASS-21) were evaluated using Rasch analysis for the first time using a clinical sample. The subscales demonstrated adequate fit to the Rasch model without differential item functioning for the variables age, gender, pain intensity, and disability.
- The properties of the DASS-21 subscales support their use in group research, but the anxiety and stress subscales do not appear to function well for making decisions regarding individual respondents.
- The DASS-21 aggregate score is not supported as a measure of negative affect.
- Analyses found gaps in measurement across the DASS-21 subscales. Any future development of item and category structure should include analyses similar to those used in this study to improve measurement of disordered mood across the spectrum of severity.

their primary presenting problem, which decreases the tool's sensitivity. It has, for example, been recommended that somatic items should be excluded from depression questionnaires [24,25].

The Depression Anxiety and Stress Scales questionnaire (DASS) [26,27] is composed of three 14-item subscales: DASS-depression, DASS-anxiety, and DASS-stress. A shorter 21-item version, composed of three 7-item subscales, has also been developed. The DASS has been evaluated in diverse general and clinical samples, including those with chronic pain [24], using classical test theory approaches. These have generally found good internal consistency [26,28−31] and good construct [26,30], and convergent [26,28,29,32] validity for the 42- and 21-item versions. The Depression, Anxiety, and Stress Scales (DASS-21) has demonstrated comparable clinimetric properties to the 42-item DASS, and has better three-factor structure and discriminant validity [29]. The aggregate DASS score has been suggested to provide a measure of ''negative affectivity'' [33], but the construct validity of this suggestion has not been determined.

Existing analyses of the DASS have been criticized for not meeting contemporary statistical standards [34]; for example, its factor structure may be affected by the respondent's age [35]. Of the more contemporary statistical techniques [36], the Rasch model allows a detailed analysis of predicted scale structure, including item difficulty, category ordering, and examination of item bias [37]. The only published Rasch model analysis of the DASS-21 used a community sample [34], which left unsettled the critical issue of whether pain and disability influence test performance. That DASS-21 data are independent of pain and disability is critical if we are to use it to screen for mood disorders in patients in pain.

1.1. Aims and objectives

The aims of this study were to (1) determine the clinimetric properties of the DASS-21 subscales and the DASS-21 aggregate score in a sample of chronic low back pain patients using Rasch analysis and (2) examine the influence of age, gender, pain, and disability, on responses to the DASS-21 subscales and the DASS-21 aggregate score.

2. Methods

2.1. Sample and data collection

Data were previously collected during a randomized placebo-controlled trial on low back pain (LBP) [38]. Patients were eligible if they were between the ages of 18 and 80 years, able to exercise, spoke English, and were seeking care for chronic nonspecific low back pain (CLBP), defined as LBP that had been present for 12 weeks or longer.

Data were obtained at baseline from the 154 participants included in the trial. Participants provided demographic data, a score of their pain intensity over the previous week, a disability score (Roland-Morris Disability Questionnaire [RMDQ]), and the DASS-21.

The trial was conducted in a large metropolitan hospital in Sydney, Australia and was approved by the University of Sydney Human Research Ethics Committee, Australia.

2.2. Study instruments and scoring

The Pain Intensity Numerical Rating Scale involves the participant rating the intensity of their pain over the last week on an 11-point scale, from ''no pain'' to ''worst pain possible.'' This instrument has been shown to adequately discriminate pain intensity in patients with chronic pain [39,40].

The RMDQ [41] requires participants to score 24 items related to activities of daily living. Higher scores (maximum score = 24) indicate higher disability. Reliability for use in chronic pain [42] and correlations with other measures of physical function [42−45] have been demonstrated for this instrument.

The DASS-21 is a short version of the 42-item questionnaire originally developed by Lovibond and Lovibond [26,27]. It is self-administered and consists of 21 items offering four response categories (0, 1, 2, 3). The questionnaire is separately scored as three 7-item subscales designed to measure the latent constructs of depression, anxiety, and stress. Completion of the scale involves the respondent indicating the extent to which each of 21 statements applied to them over the last week. Each item is assigned a number: 0 ''did not apply to me at all,''

1 ''applied to me to some degree, or some of the time,'' 2 ''applied to me to a considerable degree, or a good part of the time,'' or 3 ''applied to me very much, or most of the time.'' A total score is obtained for each subscale by adding the item scores. Following established practice, we have multiplied all DASS-21 scores by 2 [27]. Higher scores on each DASS subscale indicate a higher severity of each respective construct.

2.3. Analysis plan

2.3.1. Rasch analysis

Until recently [34], the DASS had only been investigated using classical test theory approaches, sometimes referred to as ''weak'' for establishing clinimetrics [46,47], because of their simplicity and ease with which data can meet the necessary assumptions. However, the results of classical testing are coupled to the characteristics of the tested sample, and do not precisely describe the properties of the scale at the item level [37].

The Rasch model [48] evaluates the ability of a scale and its component items to discriminate between people who fall at different points along the continuum of an underlying, or latent, construct. The approach is distinct from classical test methods for evaluating construct validity. In the classical approach, the results of the DASS might be compared with the results of a specified ''gold standard'' of assessment for each of the latent constructs of depression, anxiety, and stress. In Rasch analysis, the goal is to evaluate the degree to which each of the scale items and, in turn, item categories, fit within a mathematical model. This, in short, suggests that responses to individual items should be predictable from the individual's overall scale score, and vice versa. In so doing, evidence for validity of the scale is provided through fit indicators and response thresholds. These results could be viewed as complementary to classical approaches, both of which contribute to evidence for validity of the scale.

In order for the DASS-21 to fit the Rasch model, respondents with a greater severity of the tested psychological construct (as indicated by overall scale score) should endorse individual DASS subscale items located at the upper extreme of the continuum more frequently than respondents with a lesser severity, who should only endorse items located at the lower parts of the scale. Correspondingly, DASS items that are located at the lower end of the continuum have a greater probability of being endorsed than higher items. Any deviation from this expected response structure denotes misfit to the model and suggests that the scale is not measuring a single construct.

Separate Rasch analyses were performed for each of the DASS-21 subscales and of the aggregate ''negative affect'' DASS-21 score [33] using RUMM2030 software [49]. The steps involved in Rasch analysis have previously been published in detail elsewhere [37,46,50]. We tested the properties of the DASS-21 and whether it met the assumptions of the Rasch model. The appropriate form of the polytomous Rasch model for each scale was determined by conducting the likelihood ratio test provided in RUMM2030 [51−53]. Data for which the likelihood ratio test was significant were analyzed using the partial credit model. Unlike the rating scale model, the partial credit model does not assume that threshold distance is uniform across all scale items. We also examined the data for differential item function (DIF) across subgroups formed by age, gender, pain intensity, and disability.

The final samples for each scale excluded persons scoring at the floor and ceiling levels because their item responses do not vary and their standard errors are infinite [54]. Item- and person-fit statistics were examined for departures from model expectations. First, to test whether scale items are endorsed in a way that reflects the severity of the construct being measured, the relative item locations were examined. Person-item residuals were also examined for any item redundancies. Secondly, to investigate whether respondents are able to reliably discriminate between the different response categories, we examined the thresholds. When disordered thresholds occurred, adjacent categories were collapsed to examine if this improved model fit [37,50]. Thirdly, tests comparing expected and observed responses using the chi-square statistic were examined for significance, which would imply departure from the model. Individual item chi-square statistics were also examined for significance with a Bonferroni adjusted α of 0.05. Person responses were examined for extreme responses through the fit residual. Individual fit residuals for either items or persons were considered normal when 99.5% of values fell in the ± 2.5 range and 99.9% of values fell in the ± 3.0 range [46]. Individual fit residuals outside of these criteria would imply items that do not fit the model well and should be re-examined in an effort to identify the reason for poor fit. In these cases, persons providing extreme residual values were compared with those who did not, using a chi-square test of significance (for gender and ethnicity defined as born in or outside Australia) or a one-way analysis of variance (ANOVA) (for age, pain intensity, and disability) to test for any effect of these variables. Finally, the data were examined for differential item functioning. Four subgroups were created: (1) age (18−39, 40−69 years, 70 years and older), gender (male, female), pain intensity (0−3 mild or no pain, 4−6 moderate pain, 7−10 severe pain), and disability (0−8, 9−16, 17−24). DIF was tested by conducting an ANOVA with a Bonferroni adjusted α of 0.05 for each item across the levels of the specified person factor and levels of the latent trait.

Rasch analysis requires that a single dominant construct is being measured or, at least, that the response to each item is consistently affected by the same variables [55]. No single method of demonstrating unidimensionality is uniformly accepted, although overall item fit is strongly indicative. For the purposes of description and to demonstrate similarity between our findings and others [26,27], exploratory factor analyses using varimax rotation were conducted for each of

the individual subscales. The identification of a single factor (eigen values ≥ 1) in each subscale provided further evidence for unidimensionality [56]. From within the Rasch analysis, testing of unidimensionality was performed using the method proposed by Smith [57]. A principal component analysis of residuals was performed and the resultant person estimates were subjected to a series of independent *t*-tests. The tested scale was considered unidimensional if less than 5% of these *t*-tests were significant.

To quantify internal reliability, a Person Separation Index (PSI) was calculated for each scale and interpreted as reliable if PSI >0.7 for group use and >0.85 for individual use [50].

3. Results

3.1. Participants

The analysis was performed on 154 DASS-21 questionnaires, completed without missing data. Details of participant demographic and clinical characteristics are presented in Table 1.

3.2. Depression subscale

A likelihood ratio test was not significant indicating that the Rating Scale version of the Rasch model can be used. As 53 participants recorded an extreme score, 101 were used for the analysis. The depression subscale met the requirements of unidimensionality and showed good overall fit to the Rasch model ($\chi^2 = 32.53$, $P = 0.25$) (Table 2). There were no disordered thresholds. All individual item fit scores were within normal fit residual values. Question 16 showed misfit to the Rasch model expectations on the F statistic ($P < 0.05$) (Table 3). An analysis of the depression subscale with this item deleted did not significantly alter the results. Person-fit residuals were outside the -2.5 to $+2.5$ range in 15 participants (15%). Analyses of these data revealed no significant association with gender, age, disability, pain, or ethnicity. Individual item fit scores did not depart significantly from model expectations. PSI was 0.86.

The sample was spread from approximately -4.5 to $+4.5$ logits with a mean person location of -1.85 (standard deviation [SD] 2.42) (Fig. 1). No differential item functioning was seen for age, gender, pain intensity, or functional limitation.

3.3. DASS-anxiety subscale

A likelihood ratio test was significant which indicated that the Partial Credit Rating version of the Rasch model was appropriate. As 38 participants recorded an extreme score, 116 were used for the analysis. The DASS-21 anxiety subscale met the requirements of unidimensionality and showed good fit to the Rasch model ($\chi^2 = 38.10$, $P = 0.096$). Disordered thresholds were observed in Q2 "*I was aware of dryness of my mouth*," Q9 "*I was worried about situations in which I might panic and make a fool of myself*," Q15 "*I felt I was close to panic*," and Q20 "*I felt scared without any good reason*." Several attempts to rescore these items by collapsing the four response categories

Table 1. Participant demographic and clinical characteristics

	Males, $n = 61$ (39.6%)	Females, $n = 93$ (60.4%)	Total $N = 154$
Age (yr), mean (SD)	52.79 (11.92)	54.27 (12.39)	53.68 (12.81)
Duration of low back pain (wk), mean (SD)	356.31 (425.91)	315.23 (370.25)	331.50 (392.39)
Marital status, *n* (%total)			
Single	14 (9.1)	14 (9.1)	18.2 (18.2)
Married	36 (23.4)	51 (33.1)	56.5 (56.5)
Divorced	8 (5.2)	15 (9.7)	14.9 (14.9)
Widowed	3 (1.9)	13 (8.4)	10.4 (10.4)
Employed, *n* (%total)	15 (9.7)	21 (13.6)	23.4 (23.4)
Education			
School certificate	19 (12.3)	17 (11.0)	23.4 (23.4)
Higher school certificate	15 (9.7)	22 (14.3)	24.0 (24.0)
Trade certificate	5 (3.2)	9 (5.8)	9.1 (9.1)
Diploma	2 (1.3)	7 (4.5)	5.8 (5.8)
Advanced diploma	0 (0)	1 (0.6)	0.6 (0.6)
Bachelor degree	8 (5.2)	15 (9.7)	14.9 (14.9)
Postgraduate degree	1 (0.6)	3 (1.9)	2.6 (2.6)
Other	11 (7.1)	19 (12.3)	19.5 (19.5)
DASS-depression, mean ± SD (range)	10.9 ± 13.1 (0–42)	11.6 ± 13.2 (0–42)	11.3 ± 13.1 (0–42)
DASS-anxiety, mean ± SD (range)	11.5 ± 11.2 (0–42)	12.1 ± 11.9 (0–42)	11.9 ± 11.6 (0–42)
DASS-stress, mean ± SD (range)	13.6 ± 11.7 (0–42)	14.6 ± 12.4 (0–42)	14.3 ± 12.1 (0–42)
DASS-aggregate, mean ± SD (range)	36.0 ± 34.3 (0–126)	38.2 ± 35.7 (0–126)	37.4 ± 35.1 (0–126)
NRS pain score, mean ± SD (range)	6.54 ± 1.9 (2–10)	6.9 ± 2.1 (2–10)	6.73 ± 2.0 (2–10)
RMDQ disability score, mean ± SD (range)	13.9 ± 5.3 (2–24)	12.8 ± 4.6 (1–22)	13.2 ± 4.9 (1–24)

Abbreviations: SD, standard deviation; DASS, Depression, Anxiety, and Stress Scale; RMDQ, Roland-Morris Disability Questionnaire; NRS, numerical rating scale.

Table 2. Fit statistics for the DASS subscales and DASS aggregate score

	Overall fit to Rasch model	Item fit residual, mean (SD)	Person-fit residual, mean (SD)	PSI	Exploratory factor analysis (eigen values >1 (% of variance))	% Of statistically significant *t*-tests
DASS-depression	$\chi^2 = 32.53$; $P = 0.25$	−0.32 (1.43)	−0.71 (1.76)	0.86	1 (81.2)	2.60
DASS-anxiety	$\chi^2 = 38.10$; $P = 0.096$	0.52 (1.24)	−0.43 (1.56)	0.74	1 (62.4)	2.60
DASS-stress	$\chi^2 = 37.47$; $P = 0.11$	−0.17 (1.49)	−0.55 (1.57)	0.82	1 (65.5)	3.25
DASS-aggregate	$\chi^2 = 191.48$; $P < 0.001$	−0.08 (2.38)	−0.61 (2.49)	0.91	2 (63.4, 5.7)[a]	9.09

Abbreviations: SD, standard deviation; DASS, Depression, Anxiety, and Stress Scale; PSI, Person Separation Index.

[a] A third factor (eigen value 0.99) was identified explaining an additional 5.7%.

into three resulted in a failure of the scale to meet the Rasch model requirements ($\chi^2 = 41.55$, $P < 0.048$). The original scoring structure was therefore retained for the subsequent analysis.

All individual item-fit scores were within normal-fit residual values (Table 4). Person-fit residuals were outside the −2.5 to +2.5 range in 10 participants (9%). Analyses of these data revealed no significant association with gender, age, disability, pain, or ethnicity. PSI was 0.74.

The sample was spread from approximately −3.2 to +3.2 logits with a mean person location of −1.08 (SD 1.48) (Fig. 2). No differential item functioning was seen for age, gender, pain intensity, or functional limitation.

3.4. DASS-stress subscale

A likelihood ratio test was not significant which indicates that the Rating Scale version of the Rasch model was appropriate. As 29 participants recorded an extreme score, 125 were used for the analysis. The DASS-21 stress subscale met the requirements of unidimensionality and showed good fit to the Rasch model ($\chi^2 = 37.47$, $P = 0.11$). No disordered thresholds were observed. The fit residual for Q1 "*I found it hard to wind down*" was above 2.5 but below 3.0 ($\chi^2 = 6.02$, $P = 0.20$). Question 18 showed misfit to the Rasch model expectations on the F statistic ($P < 0.05$) (Table 5). An analysis of the stress subscale with this item deleted did not significantly alter the results. Person-fit residuals were outside the −2.5 to +2.5 range in 21 participants (17%). Analyses of these data revealed no significant association with gender, age, disability, or ethnicity. There was a statistically significant difference in the pain reported by these participants compared with that reported by those with normal residuals (mean difference 1.17 (0.46), $P < 0.05$). A PSI value of 0.82 indicated sufficient person separation reliability for group use only.

The sample was spread from approximately −3.5 to +3.5 logits with a mean person location of −0.89 (SD 1.70) (Fig. 3). No differential item functioning was seen for the whole sample for age, gender, pain intensity, and disability.

3.5. DASS-21

A likelihood ratio test was significant which indicated that the Partial Credit Rating version of the Rasch model was appropriate. Analysis of the DASS-21 aggregate score indicated multidimensionality and showed significant misfit to the Rasch model ($\chi^2 = 191.48$, $P < 0.001$). Disordered thresholds were observed in Question 2 "*I was aware of dryness of my mouth*," Question 9 "*I was worried about situations in which I might panic and make a fool of myself*," Question 16 "*I was unable to become enthusiastic about anything*," Question 17 "*I felt I wasn't worth much as a person*," Question 20 "*I felt scared without any good reason*," and Question 21 "*I felt that life was meaningless*."

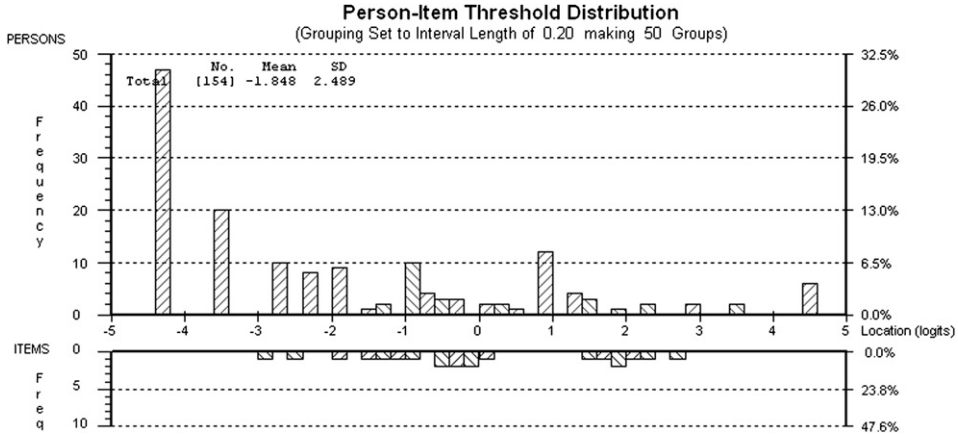

Fig. 1. Person-item threshold distribution for DASS-depression.

Table 3. Item fit statistics for DASS-21 depression subscale

Item	Location (logits)	SE	Fit residual	χ^2	χ^2 probability	*F* statistic	*F* statistic probability
Q13: I felt down-hearted and blue	−0.549	0.159	−0.019	1.953	0.744	1.115	0.354
Q5: I found it difficult to work up the initiative to do things	−0.419	0.160	1.53	2.562	0.634	0.743	0.565
Q3: I could not seem to experience any positive feeling at all	−0.076	0.164	1.114	4.923	0.295	3.450	0.011
Q10: I felt that I had nothing to look forward to	0.092	0.166	−1.749	7.945	0.094	0.570	0.685
Q17: I felt I was not worth much as a person	0.176	0.167	−2.066	8.435	0.077	0.730	0.573
Q16: I was unable to become enthusiastic about anything	0.259	0.168	0.329	2.869	0.580	3.984	**0.005**
Q21: I felt that life was meaningless	0.517	0.172	−1.385	3.846	0.427	1.534	0.199

Abbreviation: DASS, Depression, Anxiety, and Stress Scale.
Q16 is significant at a Bonferroni corrected at $P = 0.05$. Bold signifies misfit to the Rasch model.

The original scoring structure was retained as rescoring of these items had little effect on model fit.

Item-fit scores were outside the ±2.5 range for Questions 1, 2, 4, 10, 17, and 21, with significant χ^2 values for Questions 2, 8, and 4. Differential item functioning was seen in Questions 2, 10, 17, and 21 for age, gender, pain intensity, and functional limitation and in Question 8 for age, pain, and functional limitation.

The sample was spread from approximately −4.5 to +4.5 logits with a mean person location of −1.24 (SD 1.92).

4. Discussion

Our results suggest that the DASS-21 depression subscale is an effective clinical tool for the measurement of depressive symptoms in individuals with chronic low back pain. In contrast, the DASS-anxiety and the DASS-stress subscales are only reliable for use with groups. We also found that responses to the DASS-21 subscales were not affected by age, gender, pain, or disability. Finally, our analysis did not support the use of a DASS-21 aggregate score as a measure of negative affect.

Our analysis showed that the three DASS-21 depression, anxiety, and stress subscales fit to the Rasch model. By performing a factor analysis and examining fit statistics, we concluded that the three subscales predominantly capture their intended unidimensional psychological constructs. However, we found important differences among the reliability indices and item functioning of the subscales, which affect their use as measurement instruments.

The DASS-depression subscale possessed the best overall scale properties. Question 16 (*I was unable to become enthusiastic about anything*) showed possible misfit to the Rasch model on the sensitive *F* statistic. As removal of this item may significantly alter the validity of the subscale and since a re-analysis showed no significant difference to our results, we chose to retain this item. The mean person location suggested that the sample loaded toward lower levels of depression than the average location of the depression subscale items. Furthermore, the significant proportion of the sample reporting zero (47 participants) and maximum (6 participants) responses suggests that the scale possesses some floor and ceiling effects. Scale item locations suggested effective measurement across the measurable spectrum of depression severity. The DASS-depression subscale was found to be composed of items that measure a single latent construct with a PSI of 0.86, supporting its usefulness for making decisions in individuals. On the basis of these and previous published results, provided that the limited measurement ability is considered, the depression subscale appears to be a useful tool for measuring the

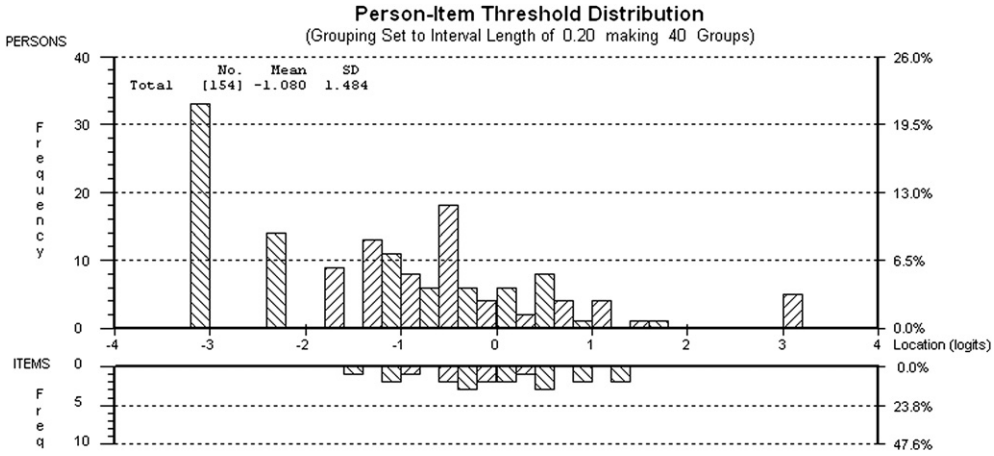

Fig. 2. Person-item threshold distribution for DASS-anxiety.

Table 4. Item fit statistics for DASS-21 anxiety subscale

Item	Location (logits)	SE	Fit residual	χ^2	χ^2 probability	*F* statistic	*F* statistic probability
Q2: I was aware of dryness of my mouth	−0.637	0.109	2.491	7.256	0.123	1.918	0.112
Q9: I was worried about situations in which I might panic and make a fool of myself	−0.064	0.109	0.422	2.627	0.622	0.320	0.864
Q7: I experienced trembling	−0.012	0.122	−0.923	8.224	0.084	2.730	0.033
Q4: I experienced breathing difficulty	0.149	0.120	0.464	1.461	0.833	0.635	0.639
Q20: I felt scared without any good reason	0.167	0.116	−0.473	6.004	0.199	2.492	0.047
Q19: I was aware of the action of my heart in the absence of physical exertion	0.177	0.124	−0.501	5.564	0.234	1.751	0.144
Q15: I felt I was close to panic	0.221	0.119	−1.114	6.968	0.138	1.890	0.117

Abbreviation: DASS, Depression, Anxiety, and Stress Scale; SE, standard error.
None are significant at a Bonferroni corrected at $P = 0.05$.

magnitude of depressive symptoms in a chronic low back pain population.

The DASS-anxiety subscale analyses indicated that the sample loaded toward lower levels of anxiety compared with the average location of the anxiety subscale items. Scale item locations clustered around the −0.1 to 0.2 range creating considerable gaps in measurement across the scale. Similarly to the depression subscale, floor and ceiling effects were observed. The subscale also demonstrated disordered thresholds in four items, which suggests that respondents were unable to reliably distinguish between the levels of the response options. There are several potential explanations. In Question 2 (*I was aware of dryness of my mouth*), it is likely that respondents may find it difficult to discriminate between four categories of mouth dryness thus resulting in misfit to model expectations. Questions 9, 15, and 20 ask patients to quantify feelings of panic and fear; it is possible that these are perceived as extreme reactions. Respondents may thus respond to them in a dichotomous rather than polytomous manner.

Significant gaps were noted across the scale item locations indicating that this subscale is not consistently sensitive to small gradations in anxiety across all the levels. The anxiety subscale was found to measure a single latent construct with a PSI of 0.74. The PSI suggests that this scale should only be used for groups, and is not internally consistent enough for making decisions on an individual level in clinical practice.

The DASS-stress subscale showed good internal consistency, but was also found to be suitable for measuring stress at group-level only (PSI of 0.82). Our analyses indicated that the sample loaded toward lower levels of stress compared with the average location of the anxiety subscale items. Scale item locations cluster around the −0.1 to −0.2 range and around +0.5 creating considerable gaps in measurement across the scale, and floor and ceiling effects, indicating that this subscale is not consistently sensitive to small changes in stress across all levels of the continuum. Question 18 (*I felt that I was rather touchy*) showed possible misfit to the Rasch model on the sensitive *F* statistic. As removal of this item may significantly alter the validity of the subscale and since a re-analysis showed no significant difference to our results, we chose to retain this item.

The use of the DASS-21 aggregate score was not supported in this study. This is in contrast to the findings of Henry and Crawford [33], who found statistical support

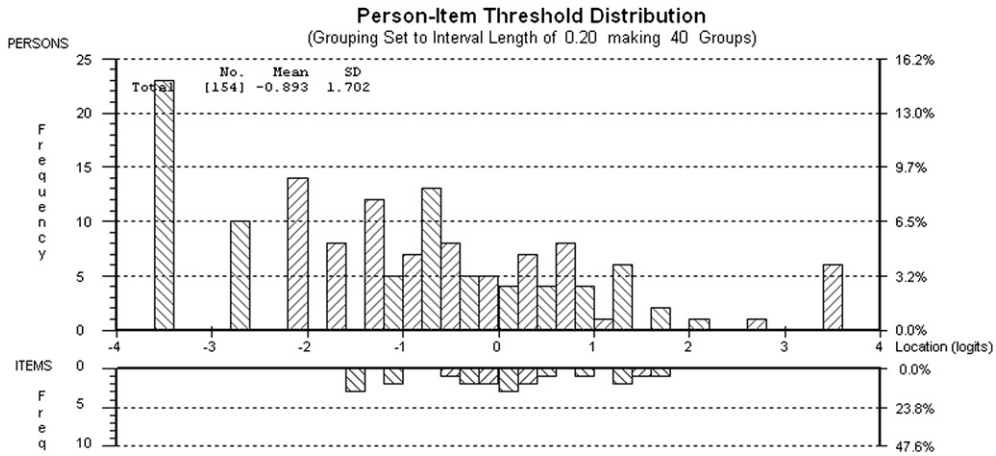

Fig. 3. Person-item threshold distribution for DASS-stress.

Table 5. Item fit statistics for DASS-21 stress subscale

Item	Location (logits)	SE	Fit residual	χ^2	χ^2 probability	*F* statistic	*F* statistic probability
Q1: I found it hard to wind down	−0.487	0.116	2.881	6.018	0.198	1.343	0.258
Q8: I felt that I was using a lot of nervous energy	−0.275	0.117	0.583	4.636	0.327	0.059	0.994
Q6: I tended to overreact to situations	−0.181	0.118	−0.626	0.395	0.983	1.193	0.317
Q12: I found it difficult to relax	−0.115	0.119	−0.660	5.944	0.203	1.014	0.403
Q11: I found myself getting agitated	−0.052	0.119	−0.788	3.559	0.469	1.805	0.132
Q18: I felt that I was rather touchy	0.497	0.127	−1.155	5.900	0.207	4.685	**0.002**
Q14: I was intolerant of anything that kept me from getting on with what I was doing	0.613	0.129	−1.453	11.013	0.026	2.125	0.082

Abbreviation: DASS, Depression, Anxiety, and Stress Scale.
Q18 is significant at a Bonferroni corrected at $P = 0.05$. Bold signifies misfit to the Rasch model.

for the validity of this aggregate score. We found that the DASS-21 aggregate score did not fit the Rasch model, with several items showing significant misfit or disordered response thresholds. Although item rescoring and/or reduction may have helped to create an instrument that fit the model, this was not an aim of the study and we did not wish to attempt the creation of a new scale. Significant restructuring may create a scale that satisfies statistical criteria, but an aggregate of the three DASS subscales would not make clinical sense. It is theoretically possible that the three constructs may change in different directions in a participant, thus we question the construct validity of such a measure.

On the basis of the results of this and other studies [34], we strongly advise against using an aggregate score of the DASS subscales to quantify a negative mood state.

We found that none of the DASS subscales showed differential item functioning for age, gender, pain severity, or disability. These analyses were not performed on the DASS-21 aggregate score as it did not fit the Rasch model. This is a significant new information that supports the use of the DASS-21 subscales to measure depression, anxiety, and stress in people who have different levels of pain and disability. It is of particular importance that we have shown that as pain fluctuates, the DASS-21 can be used to measure symptoms of disordered mood without controlling for pain intensity, which supports the original intent of the scale developers: to create a tool that is free of confounding influences because of pain [26].

The results of the present study using a clinical sample compare favorably with the findings of Shea et al. [34] who performed a Rasch analysis of the DASS-21 using a sample of 420 healthy participants. Similar to their study, we found that the anxiety subscale did not perform and the depression and stress subscales, and that the item pertaining to dryness of the mouth was a consistently problematic item. The PSIs of both studies are also remarkably similar (Shea: D: 0.87, A: 0.76, S: 0.84).

There are some important differences in our results. The mean person-fit residuals were considerably higher in our clinical sample than in the healthy subjects of Shea et al. This is consistent, as we would expect more emotional distress in a sample of people with chronic low back pain. Further, we did not find any clear evidence of differential item functioning, whereas Shea et al. found significant DIF based on gender of the respondents for the depression subscale. Additional research is required to determine why this scale functions differently in healthy men and women but this effect disappears when these people have chronic low back pain.

In conclusion, this study provides significant new information in response to the need for instruments that are valid and reliable for measuring disordered mood in a typical chronic low back pain population. This study supports the continued use of the DASS-21 to measure symptoms of depression in individuals with pain and pain-related disability but only supports the use of the anxiety and stress subscales in research on groups. Although the DASS-21 depression subscale may be used to measure symptoms of depression, the whole DASS-21 does not meet the statistical criteria to support its clinical use. On the basis of this research, the DASS-21 can only be recommended for research use involving groups.

Acknowledgments

This work is supported by GLM grant ID 571090 from NHMRC. This work is also supported by NHMRC grant number 630431.

References

[1] Katon W, Egan K, Miller D. Chronic pain: lifetime psychiatric diagnoses and family history. Am J Psychiatry 1985;142:1156−60.
[2] Wilson KG, Eriksson MY, D'Eon JL, Mikail SF, Emery PC. Major depression and insomnia in chronic pain. Clin J Pain 2002; 18(2):77−83.
[3] Rome HP Jr, Rome JD. Limbically augmented pain syndrome (LAPS): kindling, corticolimbic sensitization, and the convergence of affective and sensory symptoms in chronic pain disorders. Pain Med 2000;1(1):7−23.
[4] Romano JM, Turner JA. Chronic pain and depression: does the evidence support a relationship? Psychol Bull 1985;97(1):18−34.

[5] Baune BT, Caniato RN, Garcia-Alcaraz MA, Berger K. Combined effects of major depression, pain and somatic disorders on general functioning in the general adult population. Pain 2008;138:310−7.
[6] Carroll LJ, Cassidy JD, Cote P. Factors associated with the onset of an episode of depressive symptoms in the general population. J Clin Epidemiol 2003;56(7):651−8.
[7] Pincus T, Burton AK, Vogel S, Field AP. A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. Spine (Phila Pa 1976) 2002;27(5):E109−20.
[8] Bair MJ, Wu J, Damush TM, Sutherland JM, Kroenke K. Association of depression and anxiety alone and in combination with chronic musculoskeletal pain in primary care patients. Psychosom Med 2008;70:890−7.
[9] Munce SE, Stansfeld SA, Blackmore ER, Stewart DE. The role of depression and chronic pain conditions in absenteeism: results from a national epidemiologic survey. J Occup Environ Med 2007;49: 1206−11.
[10] Arnow BA, Hunkeler EM, Blasey CM, Lee J, Constantino MJ, Fireman B, et al. Comorbid depression, chronic pain, and disability in primary care. Psychosom Med 2006;68:262−8.
[11] Kerns RD, Haythornthwaite JA. Depression among chronic pain patients: cognitive-behavioral analysis and effect on rehabilitation outcome. J Consult Clin Psychol 1988;56:870−6.
[12] Lin EH, Katon W, Von Korff M, Tang L, Williams JW Jr, Kroenke K, et al. Effect of improving depression care on pain and functional outcomes among older adults with arthritis: a randomized controlled trial. JAMA 2003;290:2428−9.
[13] Lindsay PG, Wyckoff M. The depression-pain syndrome and its response to antidepressants. Psychosomatics 1981;22(7):571−3, 6−7.
[14] Kroenke K, Bair MJ, Damush TM, Wu J, Hoke S, Sutherland J, et al. Optimized antidepressant therapy and pain self-management in primary care patients with depression and musculoskeletal pain: a randomized controlled trial. JAMA 2009;301:2099−110.
[15] Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. Arch Intern Med 2003;163:2433−45.
[16] Grevitt M, Pande K, O'Dowd J, Webb J. Do first impressions count? A comparison of subjective and psychologic assessment of spinal patients. Eur Spine J 1998;7(3):218−23.
[17] Pignone MP, Gaynes BN, Rushton JL, Burchell CM, Orleans CT, Mulrow CD, et al. Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002;136(10):765−76.
[18] Spitzer RL, Williams JBW, Kroenke K, Linzer M, deGruy FV III, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. JAMA 1994;272(22):1749−56.
[19] Simon GE, VonKorff M. Recognition, management, and outcomes of depression in primary care. Arch Fam Med 1995;4(2):99−105.
[20] Haggman S, Maher CG, Refshauge KM. Screening for symptoms of depression by physical therapists managing low back pain. Phys Ther 2004;84(12):1157−66.
[21] Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561−71.
[22] Zung WW. A Self-Rating Depression Scale. Arch Gen Psychiatry 1965;12:63−70.
[23] Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56−62.
[24] Taylor R, Lovibond PF, Nicholas MK, Cayley C, Wilson PH. The utility of somatic items in the assessment of depression in patients with chronic pain: a comparison of the Zung Self-Rating Depression Scale and the Depression Anxiety Stress Scales in chronic pain and clinical and community samples. Clin J Pain 2005;21:91−100.
[25] Williams AC, Richardson PH. What does the BDI measure in chronic pain? Pain 1993;55:259−66.
[26] Lovibond PF, Lovibond SH. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995;33:335−43.
[27] Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. 2nd ed. Sydney: Psychology Foundation; 1995.
[28] Brown TA, Chorpita BF, Korotitsch W, Barlow DH. Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. Behav Res Ther 1997;35(1):79−89.
[29] Antony MM, Bieling PJ, Cox BJ, Enns MW, Swinson RP. Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. Washington, DC, ETATS-UNIS: American Psychological Association; 1998.
[30] Clara I, Cox B, Enns M. Confirmatory factor analysis of the Depression-Anxiety-Stress Scales in depressed and anxious patients. J Psychopathol Behav Assess 2001;23(1):61−7.
[31] Page AC, Hooke GR, Morrison DL. Psychometric properties of the Depression Anxiety Stress Scales (DASS) in depressed clinical samples. Br J Clin Psychol 2007;46(Pt 3):283−97.
[32] Crawford JR, Henry JD. The Depression Anxiety Stress Scales (DASS): normative data and latent structure in a large non-clinical sample. Br J Clin Psychol 2003;42(Pt 2):111−31.
[33] Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005; 44(Pt 2):227−39.
[34] Shea TL, Tennant A, Pallant JF. Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). BMC Psychiatry 2009;9:21.
[35] Duffy CJ, Cunningham EG, Moore SM. Brief report: the factor structure of mood states in an early adolescent sample. J Adolesc 2005 Oct;28:677−80.
[36] de Vet HC, Terwee CB, Bouter LM. Current challenges in clinimetrics. J Clin Epidemiol 2003;56(12):1137−41.
[37] Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). Br J Clin Psychol 2007;46:1−18.
[38] Costa LO, Maher CG, Latimer J, Hodges PW, Herbert RD, Refshauge KM, et al. Motor control exercise for chronic low back pain: a randomized placebo-controlled trial. Phys Ther 2009; 89:1275−86.
[39] Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? Pain 1994;58:387−92.
[40] Jensen MP, Turner JA, Romano JM, Fisher LD. Comparative reliability and validity of chronic pain intensity measures. Pain 1999 Nov;83(2):157−62.
[41] Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine (Phila Pa 1976) 1983;8(2):141−4.
[42] Jensen MP, Strom SE, Turner JA, Romano JM. Validity of the Sickness Impact Profile Roland scale as a measure of dysfunction in chronic pain patients. Pain 1992;50(2):157−62.
[43] Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis 1986;39(11):897−906.
[44] Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, et al. The Quebec Back Pain Disability Scale: conceptualization and development. J Clin Epidemiol 1996;49(2):151−61.
[45] Leclaire R, Blier F, Fortin L, Proulx R. A cross-sectional study comparing the Oswestry and Roland-Morris Functional Disability scales in two populations of patients with low back pain of different levels of severity. Spine (Phila Pa 1976) 1997;22(1):68−71.
[46] Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Asses 2009;13(12):iii. ix−x, 1−177.
[47] Novick MR. The axioms and principal results of classical test theory. J Math Psychol 1966;3(1):1−18.

[48] Rasch G. Probabilistic models for some intelligence and attainment tests. Copenhagen, Denmark: Danish Institute for Educational Research; 1960.
[49] Andrich D, Sheridan B, Luo G. RUMM2030. Perth, Western Australia: RUMM Laboratory Pty Ltd; 2009.
[50] Tennant A, Conaghan PG. 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 2007;57:1358−62.
[51] Andrich D. A rating formulation for ordered response categories. Psychometrika 1978;43:561−73.
[52] Andrich D. Application of a Psychometric Rating Model to ordered categories which are scored with successive integers. Appl Psychol Meas 1978;2(4):581−94.
[53] Masters G. A Rasch model for partial credit scoring. Psychometrika 1982;47(2):149−74.
[54] Clauser B, Linacre JM. Relating Cronbach and Rasch reliabilities. Rasch Meas Trans 1999;13:696.
[55] Spray JA. Multiple-attempt, single-tem response models. In: Linden WJvd, Hambleton RK, editors. Handbook of modern item response theory. New York: Springer-Verlag; 1997:209−23.
[56] Bjorner JB, Kosinski M, Ware JE Jr. Calibration of an item pool for assessing the burden of headaches: an application of item response theory to the headache impact test (HIT). Qual Life Res 2003;12:913−33.
[57] Smith EV Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002;3:205−31.