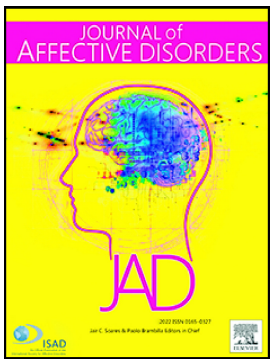

Validating the depression anxiety stress scales (DASS-21) across Germany, Ghana, India, and New Zealand using Rasch methodology

Peter Adu, Tosin Popoola, Naved Iqbal, Oleg N. Medvedev, Colin R. Simpson

PII: S0165-0327(25)00674-3

DOI: https://doi.org/10.1016/j.jad.2025.04.099

Reference: JAD 19268

To appear in:

Received date: 22 September 2024

Revised date: 25 March 2025

Accepted date: 19 April 2025

Please cite this article as: P. Adu, T. Popoola, N. Iqbal, et al., Validating the depression anxiety stress scales (DASS-21) across Germany, Ghana, India, and New Zealand using Rasch methodology, (2025), https://doi.org/10.1016/j.jad.2025.04.099

Validating the Depression Anxiety Stress Scales (DASS-21) across Germany, Ghana, India, and New Zealand using Rasch Methodology

Peter Adu[1*], Tosin Popoola[2], Naved Iqbal[3], Oleg N. Medvedev[5], and Colin R. Simpson[1 6]

[1] *School of Health, Wellington Faculty of Health, Victoria University of Wellington, New Zealand*

[2] *School of Nursing and Midwifery, College of Health, Medicine and Wellbeing, The University of Newcastle, Australia*

[4]*Jamia Millia Islamia, New Delhi, India*

[5]*School of Psychology, University of Waikato, New Zealand*

[6]*Usher Institute, The University of Edinburgh, United Kingdom*

*Corresponding Author, Peter Adu; ORCID: 0000-0002-1931-147X; Victoria University of Wellington, P. O. Box 600, Wellington 6140, New Zealand, Email: peter.adu@vuw.ac.nz

Tosin Popoola, Email: Tosin.Popoola@newcastle.edu.au (ORCID: 0000-0003-2761-7783)

Naved Iqbal, Email: niqbal@jmi.ac.in (ORCID: 0000-0001-5558-8938)

Oleg N. Medvedev, Email: oleg.medvedev@waikato.ac.nz (ORCID: 0000-0002-0529-9764)

Colin R. Simpson, Email: colin.simpson@vuw.ac.nz (ORCID: 0000-0002-5194-8083)

Abstract

Psychological distress is widespread globally, affects both mental and physical health, and has been significantly exacerbated by the COVID-19 pandemic. Reliable and valid instruments to assess psychological distress are needed so that interventions to address this condition can be directed and evaluated. We evaluated the psychometric statistics of the Depression Anxiety Stress Scales (DASS-21) using the Rasch method and a population survey sample dataset from four nations. To meet the requirements of the Rasch model, we randomly selected 500 participants (at least 110 from each country) from a total sample of 1,822 individuals from Germany (n=475), Ghana (n=523), India (n=411), and New Zealand (n=413). Minor testlet modifications were necessary to achieve a satisfactory model fit for the depression ($\chi^2(30)=40.32$, $p=0.099$), anxiety ($\chi^2(32)=31.22$, $p=0.510$), and stress ($\chi^2(20)=29.64$, $p=0.076$) subscales, and the full-scale ($\chi^2(18)=23.93$, $p=0.16$). The scales

were strictly unidimensional with acceptable reliability reflected by Person Separation Index (PSI) values of 0.86 for depression, 0.74 for anxiety, 0.82 for stress, and 0.87 for the total scale. The scales' items showed invariance across the countries and sociodemographic factors. The scales also demonstrated convergent, structural, and external validities. The study supported the reliability and validity of the DASS-21 for assessing psychological distress in four countries. Rasch model algorithms developed in this study enable conversion of the ordinal DASS-21 scale and its subscale scores to interval-level scores, which increases the precision of measurement. Future studies should apply Rasch methodology to the DASS-21 in various cultural contexts, incorporating subgroup analyses.

Keywords: Depression, Anxiety, Stress, Rasch analysis, Validity, Reliability.

Introduction

Depression, anxiety, and stress are common in most populations around the world, and their prevalence has been significantly increased by the COVID-19 pandemic (Bueno-Notivol et al., 2021). For instance, the prevalence of depression symptoms among US adults was over three times higher during the COVID-19 pandemic compared to before (Ettman et al., 2020). These mental health conditions often overlap, and account for a substantial portion of global Disability-Adjusted Life Years (DALYs) and economic burden. In fact, the age-standardised DALYs for depression and anxiety have increased by 16.4% and 16.7%, respectively, between 2010 and 2021 (Mettananda, & Mettananda, 2024). According the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision (DSM-5-TR; American Psychiatric Association [APA], 2022), depression is characterised by persistent negative feelings, loss of interest in activities, and physical symptoms such as changes in appetite and sleep disturbances. However, anxiety depicts persistent worry about future occurrences, while stress reflects continual state of heightened arousal and a tendency to become easily upset (APA, 2022). Notably, the classification of disorders by the DSM-5-TR predominately reflects the Western

perspective of nosology. Such classifications of diseases have been criticised by cultural psychiatrists and psychologists for being overly simplified, especially when considering the complexities of cultural influences on diseases. The presentation of psychiatric symptoms can vary significantly across diverse cultures, a concept known as 'idioms of distress' (Kaiser et al., 2015).

Depression, anxiety, and stress are distinct conditions, but chronic stress can contribute to the development of symptoms associated with both depression and anxiety, such as disrupted sleep patterns and increase irritability (Cohen et al., 1997; Medvedev, 2023). The negative impacts of depression, anxiety, and stress are pervasive, affecting various aspects of life including education, business, health, relationships, and overall quality of life (Schneiderman et al., 2005). In particular, during the COVID-19 pandemic, higher levels of depression, anxiety, and stress were significantly linked to lower levels of COVID-19 vaccination acceptance (Adu et al, 2024; Sekizawa et al., 2022). These conditions have also been found to reduce work performance output (Harvey et al., 2011). Additionally, adolescents who reported a decline in their academic performance were more likely to experience depression (Fröjd et al., 2008). Depression, anxiety, and stress are often found to be comorbid with various other health conditions (APA, 2022).

Assessing distress and its components, such as depression, anxiety, and stress, is important for directing and evaluating effectiveness of interventions to alleviate the impact of distress on mental and physical health. One commonly used psychometric instrument for measuring these negative affective states is the 21-item Depression Anxiety Stress Scales (DASS-21; Lovibond & Lovibond, 1995). The DASS-21 is the short version of the initial 42-item version of this scale (Lovibond & Lovibond, 1995). The DASS-21 has several advantages compared to the original version, including ease of administration, a clearer factor structure, and minimal inter-factor correlations (Lovibond & Lovibond, 1995). The scale can be used to assess depression, anxiety, and stress separately, with each containing seven items, or the overall psychological distress by summing all subscales' scores.

Evidence supporting the psychometric statistics of the DASS-21 using the traditional Classical Test Theory (CTT) method, Confirmatory Factor Analysis (CFA), indicated that it is reliable and valid for assessing depression, anxiety, and stress (Lee & Kim, 2022; Medvedev, 2023; Yıldırım et al., 2018). Internal consistency estimates, using Cronbach's alpha coefficients, range from 0.83 to 0.94 for depression, 0.76 to 0.87 for anxiety, and 0.79 to 0.91 for stress across different countries (Medvedev, 2023). Additionally, the DASS-21 has been translated into about 30 languages, such as such as Arabic, Dutch, Malaysian, Russian, Spanish, Swedish, Thai, German, and Turkish (Bibi et al., 2020; Medvedev, 2023). Currently, the DASS-21 was cited about 15000 across

different studies according to Google Scholar. The use of the traditional CFA in the assessment of instruments has been instrumental in the literature, providing valuable insights into the validity and reliability of measurement scales. However, this method has been criticised for using ordinal measures, which are limited in precision. Moreover, while the optimal estimation of parameters in Confirmatory Factor Analysis (CFA) typically requires large sample sizes (VanVoorhis & Morgan, 2007), high internal consistency in such large samples may be misleading. It might simply indicate that many participants are responding to items in a similar way, rather than the items truly measuring the underlying trait consistently across different contexts or populations (Hays et al., 2000). Other statistical parameters may also disproportionately reflect the characteristics of a specific subgroup within a broader population sample.

Nevertheless, Rasch analysis is known to overcome the limitations of the CTT methods regarding instrument assessment. Rasch analysis assumes that the probability of an item response category being endorsed depends on the difference between the person's ability and the item's difficulty (Hays et al., 2000). This method focuses on the fit between the person response patterns and the item difficulty, providing a more nuanced reliability estimate than overall inter-item correlation typically used in CFA (Medvedev & Krägeloh, 2022). Therefore, the Rasch methodology provides less sample-independent assessment, interval-level scaling, and detailed fit statistic of scales, which aligns with basic assumptions of measurement. These assumptions include ensuring equitable treatment for all users of an instrument, maintaining consistent and proportional scale units (Thurstone, 1931). Furthermore, unlike the CFA, Rasch models do not assume that the steps between adjacent response rating categories are equal.

To date, only two studies have applied Rasch methodology (Rasch, 1960) to assess the psychometric properties of the DASS-21, focusing on Australian (Shea et al., 2009) and New Zealand (Medvedev et al., 2020) populations. While these studies reported optimal fit indices after removing some misfitting items, this approach potentially affects the scale's validity (Boateng et al., 2018). Moreover, the lack of cross-cultural validation in these studies limits the generalisability of their findings and the robustness of the interval scores derived from the DASS-21 (Cooper, 2023). Given these limitations, our study aimed to evaluate and improve the measurement precision of the DASS-21 using Rasch analysis across diverse cultures, specifically samples from Ghana, Germany, India, and New Zealand were used in the study. We also sought to assess the scale's reliability, dimensionality, and structural validity in these diverse populations, examining its convergent and divergent validity, and investigated measurement invariance across the different cultural contexts. We hypothesised that the DASS-21 will demonstrate adequate fit to the Rasch model across the diverse samples, potentially requiring

some modifications to optimise its psychometric properties. Furthermore, we expected the scale to show strong convergent validity with related constructs such as negative affect, in line with previous research (Pezirkianidis et al., 2018). Conversely, we anticipated that the DASS-21 will exhibit weak to non-significant correlations with unrelated measures like positive affect, self-compassion, compassion towards others, and dispositional optimism, supporting its divergent validity (Chan et al., 2022; Medvedev et al., 2021). By addressing these objectives and testing these hypotheses, we aimed to strengthen the validity and reliability of the DASS-21 for use in contemporary cross-cultural research.

Methods

Participants

The total sample included 1,822 individuals recruited from Germany (n=475), Ghana (*n*=523), India (*n*=411), and New Zealand (*n*=413). The data collection occurred in mid-2022 and is part of a broader dataset investigating the links between psychological variables and vaccination attitudes. The age of participants for the full data ranged from 18 to 80 years in India (M_{age}= 26.14; *SD*=8.57), 18 to 89 years in New Zealand (M_{age} = 46.35; *SD* = 18.07), 18 to 63 years in Ghana (M_{age} = 29.48; *SD* = 5.69), and 18 to 87 years in Germany (M_{age} = 44.09; *SD* = 5.57). Chi-square test of independence showed that the randomly selected participants for the Rasch analysis differed statistically significant by country in terms of education levels ($\chi 2$ (42) = 1003.72, $p < 0.001$), sex (χ2 (30) = 63.58, $p < 0.001$), age ($\chi 2$ (6) = 365.15, $p < 0.001$), marital status ($\chi 2$ (45) = 112.00, $p < 0.001$), and employment status ($\chi 2$ (42) = 157.25, $p < 0.001$) across the countries.

Power analysis

According to Hagell and Westergren (2016), a sample size of 500 is recommended for Rasch analysis if a Rasch Unidimensional Measurement Model (RUMM2030) is used to address the issue of chi-square sensitivity to large sample sizes. This sample size is also necessary for conducting Differential Item Functioning (DIF) invariance testing in Rasch analysis. We initially randomly selected 1,000 participants (250 from each country) using Microsoft Excel for the Rasch analysis, representing equal sample sizes for each country in the RUMM2030 and taking the advantage of large sample size. However, to ensure the required sample size was met for this analysis and address the chi-square test's tendency to highlight statistical significance with larger samples, potentially overemphasising data misfit with negligible practical implications (Pelton, 2002), this sample underwent continuous adjustments using the RUMM2030 software. Ultimately, we arrived at a satisfactory final sample of randomly selected 500 participants (120 participants from Germany, 140 from

Ghana, 112 from New Zealand, and 119 from India), mirroring the characteristics and diversity of the entire sample (Hagell & Westergren, 2016).

Procedure

The authors' institutional Human Research Ethics Committee approved this study. Data collection in Ghana and India was conducted through various online platforms such as Facebook, WhatsApp, Twitter, Instagram, and Email. We employed convenience sampling via a snowballing technique for the data collection in Ghana and India, using *selectsurvey.net* software. No incentives were offered to participants from these two countries. For data collection in New Zealand and Germany, the process was facilitated by the data collection company, Qualtrics, and participants received small monetary compensation (e.g., NZ$5 or an opportunity to enter a draw). The adoption of online data collection methods was driven by COVID-19 preventive measures, such as social distancing. Nevertheless, online data collection is growing internationally (e.g., Adu, et al., 2024), and it also offers a way of reaching out to a more diverse population for instance that may not be reached through postal or telephone surveys (Lee et al., 2016). Participants were first asked to provide demographic information before completing the main measures, which typically took about 15 minutes. The questionnaires were presented in English for participants in Ghana, India, and New Zealand, while a validated German version was provided for participants in Germany (e.g., Adu et al., 2024).

Measures

The DASS-21 (Lovibond & Lovibond, 1995), is rated on a 4-point Likert-scale response option: from 0= "*Did not apply to me at all*" to 3= "*Applied to me very much*" and was utilised to measure depression, anxiety and stress. Each subscale contained 7 items including: depression ("I couldn't seem to experience any positive feeling at all"), anxiety ("I was aware of dryness in my mouth"), and stress ("I found it hard to wind down"). The reliability of all scales in this study were assessed using the Rasch-based Person Separation Index (PSI), with the coefficients of DASS-21 reported at the results section.

Compassion towards others was assessed with the Santa Clara Brief Compassion Scale (SCBCS; Hwang et al., 2008). This is a 5-item measure scored on a 7-point Likert scale, ranging from 1= "*not at all true of me* to" 7= "*very true of me"*. Example of an item found on the scale is "I tend to feel compassion for people, even though I do not know them". The reliability of the SCBCS is very good (PSI=0.81; *M*=24.10, *SD*=7.00) (Adu et al., 2024 manuscript under review).

The Life Orientation Test Revised (LOT-R; Glaesmer et al., 2008; Scheier et al., 1994) was used to measure dispositional optimism. This 10-item scale is rated on a 5-point Likert scale ranging from 0 = "*strongly disagree*" to 4 = "*strongly agree*". For instance, a positively worded item on this scale is "In uncertain times, I usually expect the best". The scale demonstrated acceptable reliability (PSI=0.73; *M*=16.10, *SD*=3.62).

The 12-item Self-Compassion Scale-Short Form (SCS-SF; Raes et al., 2011) was employed to measure self-compassion. This scale uses a 5-point Likert-scale response format: 1 = "*Almost Neve*r" to 5 = "*Almost Always"*. An example of positively worded item on this scale is "I try to be understanding and patient towards those aspects of my personality I don't like". The reliability of the scale was found to be acceptable (PSI=71; *M*=31.40, *SD*=7.40).

We used the well-known 20-item Positive Affect (PA) and Negative Affect (NA) Scale (PANAS; Watson et al., 1988) to assess PA and NA. Participants rated each adjective on a 5-point Likert scale, ranging from 1= "*very slightly*" to 5= "*extremely*". Examples of adjectives measuring PA include "interested," "strong," and "proud"; while NA include "anger", "fear", and "sadness". While the reliability coefficient for the PA subscale was excellent (PSI=0.91; *M*=31.40, *SD*=7.40), the NA subscale was found to have very good reliability (PSI=81; *M*=23, *SD*=8.70).

Statistical Analyses

Data handling and preliminary analysis

We conducted the Little's (1988) test ($p>0.05$) to assess the pattern of missingness in the data. We found out that missing values were negligible (<1%) and were ignored as the Rasch model is not sensitive to missing data as it estimates parameters for items and persons independently. All analyses, except for the Rasch analysis, were performed using the IBM Statistical Package for the Social Sciences (SPSS) version 28. The analyses encompassed Pearson's correlation, paired samples *t*-test, descriptive statistics, and psychometric evaluations. By employing the Rasch transformed interval scores for all variables, the data was deemed appropriate for parametric test tools in related analyses.

Rasch model evaluation

To assess the overall fit of the data to the Rasch model, we first examined at the chi-square test for item-trait interaction, aiming for a non-significant result (p>0.05) indicating an acceptable fit (e.g., no significant deviations from Rasch model expectations). We also examined item fit residual values to ensure they fell within an acceptable range (-2.50 to +2.50) and checked residual correlations between items to ensure they were below 0.20 (Tennant & Küçükdeveci, 2023; Wilkinson et al., 2023). Unidimensionality was assessed by comparing person estimates on two groups of items, with high and low loadings on the first principal component of residuals using *t*-tests, where unidimensionality is supported if the number of significant *t*-tests is below 5% or lower bound of confidence intervals overlaps 5% (Hagell, 2015; Smith, 2002). Threshold ordering was evaluated by examining Item Characteristic Curves (ICCs), with the intention to achieve monotonous patterns of items (Andrich, 1978). Good sample targeting was achieved if the difference between the average person measure and the average item measure of the dataset is below 1.00 logit (Finger et al., 2012). Ceiling and floor effects were significant only when more than 20% of participants fell outside the range covered by the items representing the latent variables (Murugappan et al., 2022). Any item local dependence, indicated by residual correlation matrices above 0.20 was addressed using testlets creation methodology (Tennant & Küçükdeveci, 2023; Wilkinson et al., 2023). Reliability of the scale was estimated using PSI. Measurement invariance was evaluated across different sociodemographic factors, including biological sex, age, country of origin, health workers vs. general population, marital status, employment status, and educational levels using the DIF (Hagquist & Andrich, 2017; Pratscher et al., 2019). Age group invariance was assessed by categorizing age into groups (i.e., 18-35 years, 36-50, years, and 61-89) based on percentiles. If an acceptable model was achieved, an algorithm was developed to transform ordinal data into interval-level data (Medvedev et al., 2020). We assessed the convergent and divergent validity of the scales using Pearson's correlation coefficients.

Results

Overall model fit, reliability, and unidimensionality

Our initial analysis (Table 1) indicated that the depression (χ^2 (21)= 60.80, p < 0.001), anxiety (χ^2(42) = 186.89, p < 0.001), and stress (χ^2 (21)=49.00, p < 0.001) subscales, and the full-scale (χ^2 (21)=189.96, p < 0.001), deviated from the Rasch model expectations, as evidenced by a statistically significant interaction between the items and the latent variable. However, there was evidence of strict unidimensionality for all subscales and the total scale (Table 1). The reliability at this initial analysis were also acceptable for depression, anxiety, stress, and the full-scale, ranging from 0.71 to 0.92. Table 2 provides an overview of the individual item

fit statistics within the Rasch model, including item locations, fit residuals, and chi-square parameters. Items 5 and 10 in the depression subscale, item 1 in the stress subscale, and items 2, 15, and 20 in the anxiety subscale showed significant misfit to the model reflected by elevated fit residuals (indicated by asterisks in Table 2). Removing these items could impact construct validity as they represent clinically important distress symptoms (e.g., hopelessness). However, Rasch model misfit can be due to individual items misfit or local dependency between items. Therefore, we examined the residual correlation matrix and identified locally dependent items potentially impacting the Rasch model fit.

<Insert Table 1 here>

<Insert Table 2 here>

<Insert Figure 1 here>

We used testlets method to address local dependency and combined locally dependent items as follows: for depression subscale, items 3+13 and 5+21; for anxiety subscale, items 2+20 and 4+15; and for stress subscale, items 1+11, 6+8, and 12+18. At this point, both the depression ($\chi^2(30) = 40.32$, $p = 0.099$), and stress ($\chi^2(20) = 29.64$, $p = 0.076$) subscales were found to fit well the Rasch model, while the anxiety subscale underwent further testlet modification to achieve the optimal overall Rasch model fit (Table 2; $\chi^2(32) = 31.22$, $p = 0.510$). This resulted in evidence of unidimensionality for all the subscales. An examination of local item dependency in the full-scale indicated three general groups of locally dependent items, as expected, corresponding to the three subscales. As a result, the final three subscales were treated as testlets for the full scale, similar to the study of Medvedev et al. (2020), which then fitted the Rasch model very well ($\chi^2(18) = 23.93$, $p = 0.160$). We then achieved strict unidimensionality for the full scale, with the lower bound of significant t-tests within the 5% cut-off point (Table 1). Additionally, the reliability coefficients for all the scales were found to be acceptable, with PSI values of 0.86 for depression subscale, 0.74 for anxiety subscale, 0.82 for stress subscale, and 0.87 for the total scale. Furthermore, the final items, including the testlets worked appropriately in the final models for all the scales, indicated by the monotonous patterns in the ICCs for items (e.g., Figure 1).

DIF, Person-Item Trait, and Ordinal-to-Interval Conversion

We found that items were invariant across demographic characteristics of participants as evidenced by no significant DIF. This evidence was presented by participants' sex (Figure 2), age (Figure 3), country of origin (Figure 3), health workers vs. the general population (Supplementary Figure S1), marital status (Supplementary Figure S2), employment status (Supplementary Figure S3), and educational levels (Supplementary Figure S4). Further, the results indicated that the proportion of participants falling outside the range of the latent variable items for the scales ranged from 0.8% to 16.8%. Since these percentages were below the 20% threshold for significance of ceiling and floor effects, the person-item distribution thresholds targeted the sample well (Figure 4). Since we achieved satisfactory Rasch model fit indices for all scales, algorithms were developed based on Rasch model person estimates to convert ordinal scores into interval-level data (Table 3). Moreover, we were able to conduct paired samples *t-tests* to determine if there were statistically significant differences between the ordinal and interval level scores for the scales using the same scale range. The t-test used the total sample from the four countries by converting participants' scores into interval-level scores and then comparing these with their ordinal scores. The results showed statistically significant differences between the means of the ordinal (M = 6.81, SE= 0.14) and interval level (M = 7.73, SE = 0.13) scores for the depression subscale $t(1821) = 32.46$, $p < 0.001$); the ordinal (M = 6.04, SE = 0.13) and interval level (M = 7.05, SE = 0.12) scores for the anxiety subscale $t(1821) = 29.47$, $p < 0.001$); the ordinal (M = 7.40, SE = 0.14) and interval level (M = 8.57, SE = 0.12) scores for the stress subscale $t(1821) = 40.34$, $p< 0.001$); and the ordinal (M = 20.25, SD = 16.47) and interval level (M = 24.56, SE = 0.28) of the full-scale $t(1821) = -28.07$, $p<0.001$). Effect sizes (d) were large for depression (d = 1.21), anxiety (d = 1.47), stress (d = 1.23) subscales, and the full-scale (d = 2.56). The observed lower standard error in the *t-test* of differences favoured the interval level scores.

<Insert Figure 2 here>

<Insert Figure 3 here>

<Insert Table 3 here>

Convergent and divergent validity

Results from the Pearson's correlation coefficient (r) revealed a moderate positive correlation between the scores of depression ($r = 0.51$, $p < 0.001$), anxiety ($r = 0.51$, $p < 0.001$), stress ($r = 0.51$, $p < 0.001$) subscales, the full-scale ($r = 0.53$, $p < 0.001$) and negative affect. This supported the convergent validity of the

scales. However, there were statistically significant weak negative correlations between self-compassion scores and the scores of depression (r = -0.37, $p < 0.001$), anxiety (r = -0.32, $p < 0.001$), and stress (r = -0.38, $p < 0.001$) subscales, and the full-scale (r = -0.37, $p < 0.001$). Statistically significant negative correlations were also found between dispositional optimism scores and depression (r = -0.47, $p < 0.001$), anxiety (r = -0.35, $p < 0.001$), and stress (r = -0.37, $p < 0.001$) subscales, and the full-scale scores (r = -0.41, p < 0.001). Compassion towards others measure showed weak negative correlations with depression (r = -0.03, $p < 0.001$), anxiety (r = -0.02, p < 0.001), and stress (r = -0.03, $p < 0.001$) subscales, and the full-scale scores (r = -0.02, $p < 0.001$). Lastly, positive affect scores had weak negative correlations with the scores of depression (r = -0.36, $p < 0.001$), anxiety (r = -0.22, $p < 0.001$), and stress (r = -0.27, $p < 0.001$) subscales, and the full-scale (r = -0.30, $p < 0.001$), which shows evidence of external validity.

Discussion

This study aimed to apply Rasch methodology to assess the psychometric statistics of the DASS-21 across Germany, Ghana, India, and New Zealand. Specifically, we evaluated the scales' structural validity, reliability, convergent and divergent validity, and invariance across sociodemographic factors. We discovered that none of the scales initially fit the Rasch model due to the interrelated nature of specific items, which failed to capture distinct aspects of the latent variables and the overall psychological distress. Implementing a testlet creation modification resolved this issue, leading to satisfactory model fit indices for the depression, anxiety, and stress subscales, as well as for the full scale. The scales were also found to be unidimensional and demonstrated acceptable reliability and internal validity. Moreover, all scales were invariant, with items showing no systematic bias across country of origin, marital status, age, sex, healthcare workers vs. general populace, employment status, and educational level. Meeting expectations of the unidimensional Rasch model allowed us to enhance measurement precision of the DASS-21. This was done by developing algorithms to transform ordinal data into interval-level scores using Rasch model estimations. We have also established external validity of all subscales and the total scale, which showed a positive relation with measures of negative affect, while exhibiting negative correlations with measures of self-compassion, dispositional optimism, positive affect and compassion towards others.

The testlet creation method combines items with high shared variance that are not directly associated with the overarching latent variable of interest, which helps to reduce measurement error while meaningfully capturing specific aspects of the underlying latent variable (Lundgren-Nilsson et al., 2013). Testlet creation

benefits only Rasch-transformed interval-level scores, as testlets are incorporated into the conversion algorithm (Tennant & Küçükdeveci, 2023). However, previous studies, including Medvedev et al. (2020) and Shea et al. (2009) removed items from the DASS-21 to improve model fit, which can potentially impact the scales' construct validity (Boateng et al., 2018). Our study preserved content validity by using testlet creation, keeping all original items in this measure. This method preserves the scales' soundness and underscores its value in Rasch analysis for maintaining scale's structure while improving model fit (Tennant & Küçükdeveci, 2023). Despite the advantages associated with this method, its application in the literature is limited, making the current study a valuable contribution to expanding its use and enhancing Rasch method. We also observed that the scales were unidimensional, meaning a single score is appropriate for measuring depression, anxiety, stress, and general psychological distress (Adu et al., 2024). In other words, items in each subscale capture a single underlying latent trait (e.g., depression), while the total scale reflects the overarching latent trait of psychological distress, which aligns with findings from previous DASS-21 Rasch studies (Medvedev et al., 2020). While Shea et al. (2009) removed some items to achieve an acceptable Rasch model fit and unidimensionality, their findings failed to support the full version of the DASS-21 as a measure of general distress.

The current study provides supporting evidence indicating that the DASS-21 can be adequately used with a non-clinical general population sample in the four countries used in this investigation. While the reliability coefficient for the anxiety subscale met the criteria for group-level assessment, the reliability coefficients for all other scales are suitable for both individual and group assessments (Fisher, 1992). Notwithstanding, the study by Medvedev et al., (2020) revealed reliability coefficients supporting group level assessment. In our study, the general distress scale demonstrated strong sample targeting, suggesting that the full-scale effectively captured the range of symptoms within the sample. This was supported by the monotonous patterns of the ICC.

The use of interval-level scores enhances the precision of capturing the range of distress symptoms, enabling the detection of meaningful changes among participants. Table 3 outlines the process for converting ordinal data to interval-level scores. To illustrate this with the full DASS-21 scale, consider two scenarios: in one case, an individual's distress score decreases from 10 to 5, and another whose score drops from 15 to 10. Although both reductions appear equal on the ordinal scale, each showing a decrease of 5 points, converting these to interval scores reveals a different pattern. The first patient's reduction is 4.75 units, while the second patient's reduction is only 3.21 units in the interval scale (Table 3). This demonstrates that using the Rasch

interval scores can offer a more precise measurement, providing greater clarity on actual changes in distress measures among individuals at both group and individual levels. This approach improves upon the traditional CFA by addressing the limitation of merely assuming equal intervals between score categories without explicitly testing this assumption (Magno, 2009). By incorporating interval-level scores, researchers can more accurately assess these constructs and enhance the validity and reliability of their analyses (Medvedev et al., 2021). The lower standard error observed in the *t-test* of differences for the interval scores suggest reduced measurement error, indirectly indicating that transforming ordinal scores into interval scores enhances a scale's measurement accuracy. Furthermore, the evidence of measurement invariance across numerous sociodemographic factors improves the scale's usefulness, trustworthiness, and consistency across diverse populations with different characteristics, enabling dependable comparisons of study outcomes, especially from the countries under study (Welzel et al., 2023).

The significant positive associations between the scores of the subscales and the measure of negative affect provide evidence of convergent validity for all the scales. The above correlations are congruent with well-established evidence (Pezirkianidis et al., 2018). These associations mean that the scales are theoretically related to similar measures. Conversely, all the scales exhibited statistically significant correlations to measures of self-compassion, positive affect, compassion towards others, and optimism, failing to establish divergent validity for the scales. However, these associations have been consistently reported in the literature, supporting the external validity of the scales (Chan et al., 2022; Medvedev et al., 2021).

All the scales in this study effectively captured the full range of trait levels present in the population, demonstrating good targeting (fFnger et al., 2012) . This contrasts with previous research (Medvedev et al., 2020; Shea et al., 2009) which reported poor targeting due to their samples lacking participants with clinically significant distress, resulting in a lower prevalence of severe distress symptoms. These studies noted that psychiatric measures often report fewer severe symptoms in community samples (Shorey et al., 2022). Despite using a relatively homogeneous convenience sample recruited from the general population via social media, our study found good targeting. This discrepancy might be attributed to the elevated distress levels observed during the COVID-19 pandemic (Bueno-Notivol et al., 2021), resulting in a broader spectrum of distress levels in our sample. Despite the limitations of our sampling technique, the current unidimensional scores of the all the scales remain promising for use in non-clinical settings within the current contexts. This is particularly supported by

parameters such as satisfactory model fit indices, unidimensionality, sound reliability, invariance, and its external validity stemming from the use of four countries from four continents.

Moreover, the discrepancies between previous and current results regarding the DASS-21 may be attributed to cultural influences on the manifestation of mental disorders. Cultures vary in their expression of distress (Kirmayer et al., 1998). For instance, Asians often report more somatic symptoms and Western societies exhibiting more psychological symptoms (Arnault & Kim, 2008; Sheikh & Furnham, 2012). Therefore, it is possible that item 5 ("I found it difficult to work up the initiative to do things") which was removed by previous researchers, e.g., Medvedev et al., (2020), may not typically represent underlying symptoms of psychological distress in these cultural contexts but could indicate distress in other cultures (Kaiser et al., 2015). This complicates the universal acceptance of distress measures, and further highlights the complexities of applying the Western diagnostic frameworks, such as those outlined by the International Classification of Diseases, Eleventh Revision (ICD-11; World Health Organization, 2022) and DSM-5-TR (APA, 2022), globally (Young, 2013). For example, "thinking too much" idioms reflect aspects of experiencing distress and are expressed differently between and within cultures, but it is not captured by many psychiatric diagnoses (Kaiser et al., 2015). Notwithstanding, removing item 5, which measures motivation toward life, from the scale could impact its content validity, as lack of motivation is considered a key indicator of distress (APA, 2022). Consequently, maintaining all items on the already developed scale is vital to preserving its validity (Clark & Watson, 2019), as demonstrated in the current study. Therefore, future studies could consider developing context-specific versions of distress measures while integrating local idioms of distress, especially in contexts where the validity and reliability of the DASS-21 are not fully supported (APA, 2022). For example, modifying item 5 to reflect the cultural conceptualisation of personal motivation could be helpful. In some cultures, the focus on group harmony may reduce individual motivation, while cultures emphasising personal initiative and self-driven achievement will increase personal motivation (Hofstede, 2001). This approach would enhance the effective monitoring of distress and the efficacy or effectiveness of treatment outcomes.

Conclusion

This study has applied Rasch analysis to assess the psychometric statistics of the DASS-21 in Germany, Ghana, India, and New Zealand. The results indicated that the depression, anxiety, and stress subscales, as well as the full scale, meets expectations of the Rasch model, supporting reliability and validity of these measures across the four countries. This was achieved by modifying the items using the testlet creation

methodology. The enhanced scales with converted ordinal data to interval-level scores were found to be reliable and valid measures for assessing psychological distress irrespective of individuals' sociodemographic characteristics in the general population.

References

Adu, P., Mbinta, J., Badu Prempeh, A. S., Grigoryev, D., & Jurcik, T. (2024). Help-seeking for Social Anxiety symptoms in Ghana: Looking through the lens of sexist attitudes. *Current Psychology*. https://doi.org/10.1007/s12144-024-06121-2

Adu, P., Popoola, T., Collings, S., Aspin, C., Medvedev, O. N., & Simpson, C. R. (2024). Psychometric properties of the motors of COVID-19 vaccination acceptance scale in New Zealand: Insights from confirmatory factor analysis. *Current Psychology*, 1-11. https://doi.org/10.1007/s12144-024-05877-x

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*, 561-573. https://doi.org/10.1007/BF02293814

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Bertoli-Barsotti, L. (2004). Effects of Ordered Scores in Rasch Analysis. *Atti della XLII Riunione Scientifica SIS*, 743-746.

Bibi, A., Lin, M., Zhang, X. C., & Margraf, J. (2020). Psychometric properties and measurement invariance of Depression, Anxiety and Stress Scales (DASS-21) across cultures. *International Journal of Psychology*, *55*(6), 916-925. **https://doi.org/10.1002/ijop.12671**

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in public health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Bueno-Notivol, J., Olaya, B., Lasheras, I., Gracia-García, P., Santabárbara, J., & López-Antón, R. (2021). *Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies*. https://doi.org/10.1016/j.ijchp.2020.07.007

Chan, K. K. S., Lee, J. C.-K., Yu, E. K. W., Chan, A. W., Leung, A. N. M., Cheung, R. Y., Li, C. W., Kong, R. H.-M., Chen, J., & Wan, S. L. Y. (2022). The impact of compassion from others and self-compassion on psychological distress, flourishing, and meaning in life among university students. *Mindfulness*, *13*(6), 1490-1498. https://doi.org/10.1007/s12671-022-01891-x

Cohen, S., Kessler, R. C., & Gordon, L. U. (1997). *Measuring stress: A guide for health and social scientists*. Oxford University Press, USA.

Cooper, C. (2023). *An Introduction to Psychometrics and Psychological Assessment: Using, Interpreting and Developing Tests*. Routledge.

Dellaert, F. (2002). The expectation maximization algorithm. *College of Computing, Georgia Institute of Technology*.

Ettman, C. K., Abdalla, S. M., Cohen, G. H., Sampson, L., Vivier, P. M., & Galea, S. (2020). Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. *JAMA network open*, *3*(9), e2019686-e2019686.

Finger, R. P., Fenwick, E., Pesudovs, K., Marella, M., Lamoureux, E. L., & Holz, F. (2012). Rasch analysis reveals problems with multiplicative scoring in the Macular Disease Quality of Life questionnaire. Ophthalmology, 119(11), 2351–2357. 10.1016/j.ophtha.2012.05.031.

Fisher Jr, W. (1992). Reliability statistics. Rasch Measurement Transactions. *Chicago, USA: MESA Press*, *6*, 238.

Fröjd, S. A., Nissinen, E. S., Pelkonen, M. U., Marttunen, M. J., Koivisto, A.-M., & Kaltiala-Heino, R. (2008). Depression and school performance in middle adolescent boys and girls. *Journal of adolescence*, *31*(4), 485-498. https://doi.org/10.1016/j.adolescence.2007.08.006

Glaesmer, H., Hoyer, J., Klotsche, J., & Herzberg, P. Y. (2008). Die deutsche version des Life-Orientation-Tests (LOT-R) zum dispositionellen Optimismus und Pessimismus. *Zeitschrift für Gesundheitspsychologie*, *16*(1), 26-31. https://doi.org/10.1026/0943-8149.16.1.26

Hagell, P. (2015). Testing unidimensionality using the PCA/t-test protocol with the Rasch model: a cautionary note. *Rasch Measurement Transactions*, *28*(4), 1487-1489.

Hagell, P., & Westergren, A. (2016). Sample Size and Statistical Conclusions from Tests of Fit to the Rasch Model According to the Rasch Unidimensional Measurement Model (Rumm) Program in Health Outcome Measurement. *Journal of applied measurement*, *17*(4), 416-431.

Hagquist, C., & Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and quality of life outcomes*, *15*(1), 1-8. https://doi.org/10.1186/s12955-017-0755-0

Harvey, S., Glozier, N., Henderson, M., Allaway, S., Litchfield, P., Holland-Elliott, K., & Hotopf, M. (2011). Depression and work performance: an ecological study using web-based screening. *Occupational Medicine*, *61*(3), 209-211. https://doi.org/10.1093/occmed/kqr020

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item Response Theory and Health Outcomes Measurement in the 21st Century.

Hwang, J. Y., Plante, T., & Lackey, K. (2008). The development of the Santa Clara brief compassion scale: An abbreviation of Sprecher and Fehr's compassionate love scale. *Pastoral Psychology*, *56*, 421-428. https://doi.org/10.1007/s11089-008-0117-2

Kaiser, B. N., Haroz, E. E., Kohrt, B. A., Bolton, P. A., Bass, J. K., & Hinton, D. E. (2015). "Thinking too much": A systematic review of a common idiom of distress. *Social science & medicine*, *147*, 170-183. https://doi.org/10.1016/j.socscimed.2015.10.044

Kirmayer, L. J., Dao, T. H. T., & Smith, A. (1998). Somatization and Psychologization: Understanding Cultural Idioms of Distress.

Lee, B., & Kim, Y. E. (2022). Validity of the depression, anxiety, and stress scale (DASS-21) in a sample of Korean university students. *Current Psychology*, *41*(6), 3937-3946. https://doi.org/10.1007/s12144-020-00914-x

Lee, C.-T., Lin, C.-Y., Tsai, M.-C., Strong, C., & Lin, Y.-C. (2016). Psychometric evaluation and wording effects on the Chinese version of the parent-proxy Kid-KINDL. *Health and quality of life outcomes*, *14*, 1-10. https://doi.org/10.1186/s12955-016-0526-3

Lovibond, S. H., & Lovibond, P. F. (1995). Depression anxiety stress scales. *Psychological Assessment*. https://doi.org/10.1037/t01004-000

Lundgren-Nilsson, Å., Jonsdottir, I. H., Ahlborg, G., & Tennant, A. (2013). 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 and quality of life outcomes*, *11*, 1-9. https://doi.org/10.1186/1477-7525-11-2

Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The international Journal of Educational and Psychological assessment*, *1*(1), 1-11. https://ssrn.com/abstract=1426043

Medvedev, O. N. (2023). Depression Anxiety Stress Scales (DASS-21) in International Contexts.

Medvedev, O. N., Cervin, M., Barcaccia, B., Siegert, R. J., Roemer, A., & Krägeloh, C. U. (2021). Network analysis of mindfulness facets, affect, compassion, and distress. *Mindfulness*, *12*, 911-922. https://doi.org/10.1007/s12671-020-01555-8

Medvedev, O. N., & Krägeloh, C. U. (2022). Rasch measurement model. In *Handbook of assessment in mindfulness research* (pp. 1-18). Springer.

Medvedev, O. N., Krägeloh, C. U., Titkova, E. A., & Siegert, R. J. (2020). Rasch analysis and ordinal-to-interval conversion tables for the Depression, Anxiety and Stress Scale. *Journal of Health Psychology*, *25*(10-11), 1374-1383. https://doi.org/10.1177/135910531875526

Mettananda, K. C. D., & Mettananda, S. (2024). Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the global burden of disease Study 2021.

Pezirkianidis, C., Karakasidou, E., Lakioti, A., Stalikas, A., & Galanakis, M. (2018). Psychometric properties of the depression, anxiety, stress scales-21 (DASS-21) in a Greek sample. *Psychology*, *9*(15), 2933-2950. 10.4236/psych.2018.915170

Pratscher, S. D., Wood, P. K., King, L. A., & Bettencourt, B. A. (2019). Interpersonal mindfulness: Scale development and initial construct validation. *Mindfulness*, *10*, 1044-1061. https://doi.org/10.1007/s12671-018-1057-2

Raes, F., Pommier, E., Neff, K. D., & Van Gucht, D. (2011). Construction and factorial validation of a short form of the self-compassion scale. *Clinical psychology & psychotherapy*, *18*(3), 250-255. **https://doi.org/10.1002/cpp.702**

Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.

Saint Arnault, D., & Kim, O. (2008). Is there an Asian idiom of distress? Somatic Symptoms in Female Japanese and Korean Students. *Archives of psychiatric nursing*, *22*(1), 27. https://doi.org/10.1016/j.apnu.2007.10.003

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of personality and social psychology*, *67*(6), 1063. https://sid.ir/paper/580439/en

Schneiderman, N., Ironson, G., & Siegel, S. D. (2005). Stress and health: psychological, behavioral, and biological determinants. *Annu. Rev. Clin. Psychol.*, *1*, 607-628. **https://doi.org/10.1146/annurev.clinpsy.1.102803.144141**

Sekizawa, Y., Hashimoto, S., Denda, K., Ochi, S., & So, M. (2022). Association between COVID-19 vaccine hesitancy and generalized trust, depression, generalized anxiety, and fear of COVID-19. *BMC public health*, *22*(1), 126. https://doi.org/10.1186/s12889-021-12479-w

Shea, T. L., Tennant, A., & Pallant, J. F. (2009). Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). *BMC Psychiatry*, *9*, 1-10. https://doi.org/10.1186/1471-244X-9-21

Sheikh, S., & Furnham, A. (2012). The relationship between somatic expression, psychological distress and GP consultation in two cultural groups. *Counselling Psychology Quarterly*, *25*(4), 389-402. https://doi.org/10.1080/09515070.2012.735860

Shorey, S., Ng, E. D., & Wong, C. H. (2022). Global prevalence of depression and elevated depressive symptoms among adolescents: A systematic review and meta-analysis. *British Journal of Clinical Psychology*, *61*(2), 287-305. **https://doi.org/10.1111/bjc.12333**

Smith Jr, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of applied measurement*, *3*(2), 205-231.

Statistics, N. C. f. H. (2008). International classification of diseases, tenth revision (ICD-10). In: Diunduh.

Tennant, A., & Küçükdeveci, A. A. (2023). Application of the Rasch measurement model in rehabilitation research and practice: early developments, current practice, and future challenges. *Frontiers in Rehabilitation Sciences*, *4*. https://doi.org/10.3389/fresc.2023.1208670

Thurstone, L. L. (1931). The measurement of social attitudes. *The Journal of Abnormal and Social Psychology*, *26*(3), 249-269. https://doi.org/10.1037/h0070363

VanVoorhis, C. W., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in quantitative methods for psychology*, *3*(2), 43-50.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063. https://doi.org/10.1037/0022-3514.54.6.1063

Welzel, C., Brunkert, L., Kruse, S., & Inglehart, R. F. (2023). Non-invariance? An overstated problem with misconceived causes. *Sociological Methods & Research*, *52*(3), 1368-1400. https://doi.org/10.1177/0049124121995

Wilkinson, S., Ribeiro, L., Krägeloh, C. U., Bergomi, C., Parsons, M., Siegling, A., Tschacher, W., Kupper, Z., & Medvedev, O. N. (2023). Validation of the Comprehensive Inventory of Mindfulness Experiences (CHIME) in English Using Rasch Methodology. *Mindfulness*, 1-15. https://doi.org/10.1007/s12671-023-02099-3

Yıldırım, A., Boysan, M., & Kefeli, M. C. (2018). Psychometric properties of the Turkish version of the Depression Anxiety Stress Scale-21 (DASS-21). *British Journal of Guidance & Counselling*, *46*(5), 582-595. https://doi.org/10.1080/03069885.2018.1442558

Young, G. (2013). Breaking bad: DSM-5 description, criticisms, and recommendations. *Psychological Injury and Law*, *6*, 345-348. https://doi.org/10.1007/s12207-013-9181-8

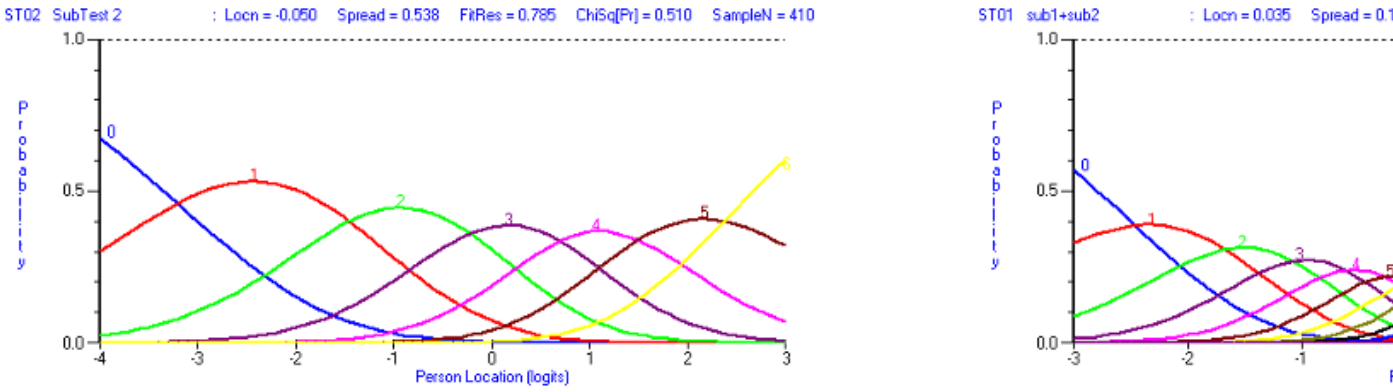

Figure 1. Example of the monotonous pattern Item Characteristic Curves (ICCs) for the final items on the scales for anxiety (bottom left) and depression (bottom right) are ordered.

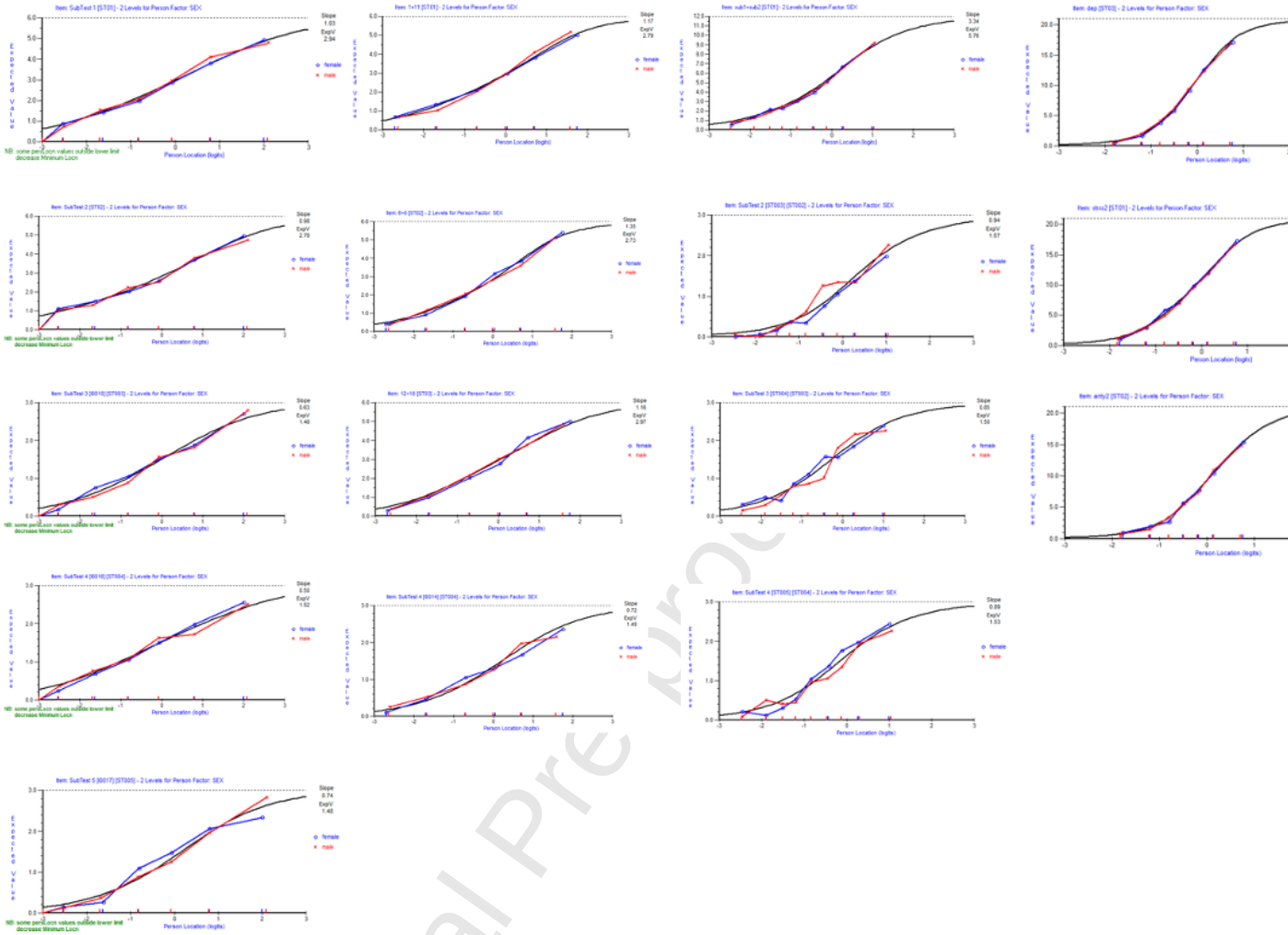

Figure 2. Differential Item Functioning (DIF) curves of the depression (first column from the left), stress (second column from the left), anxiety (third column from the left) subscales, and the full scale (fourth column from the left) for biological sex of the participants.

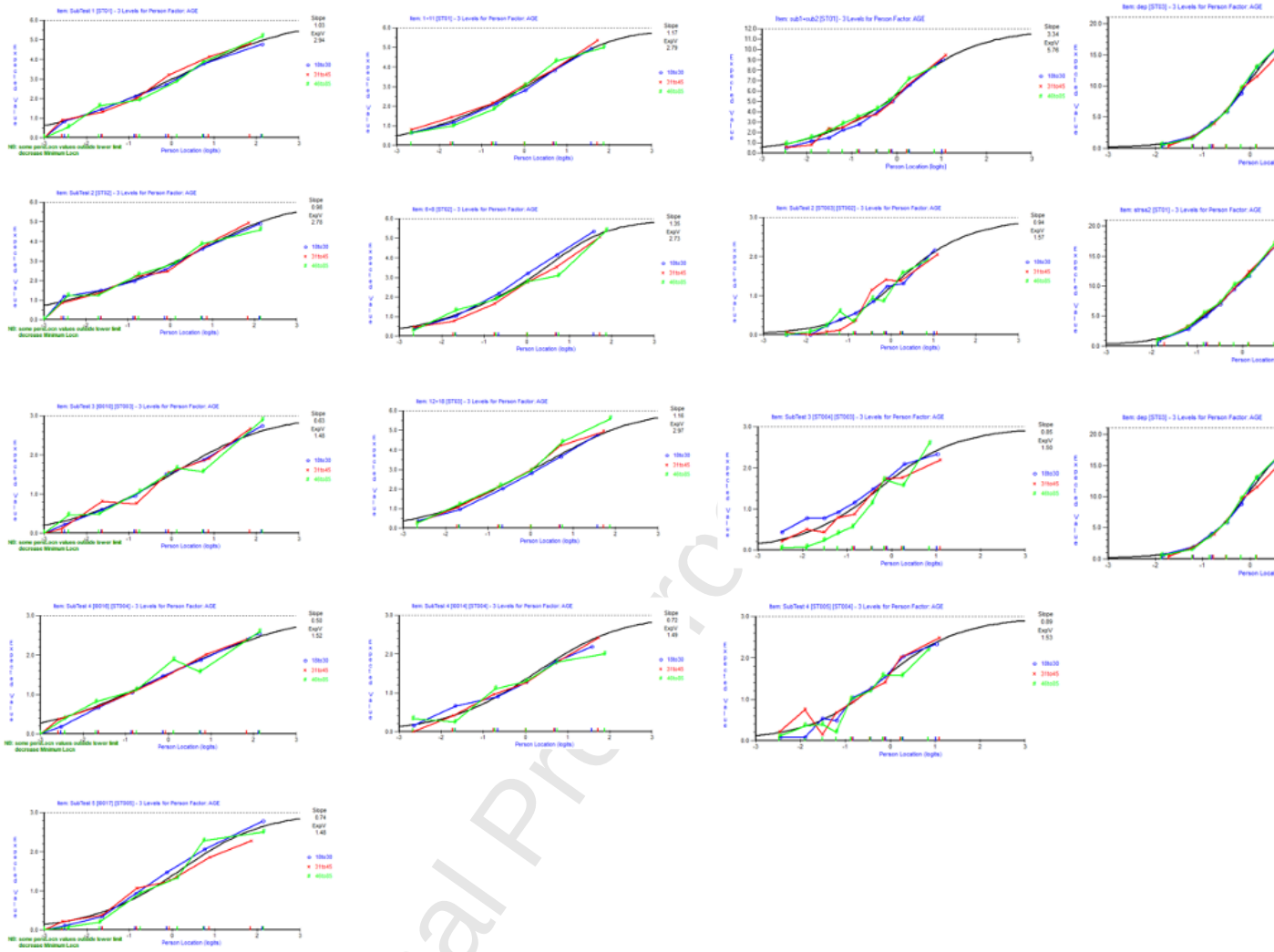

Figure 3. Differential Item Functioning (DIF) curves of the depression (first column from the left), stress (second column from the left), anxiety (third column from the left) subscales, and the full scale (fourth column from the left) for age of the participants.

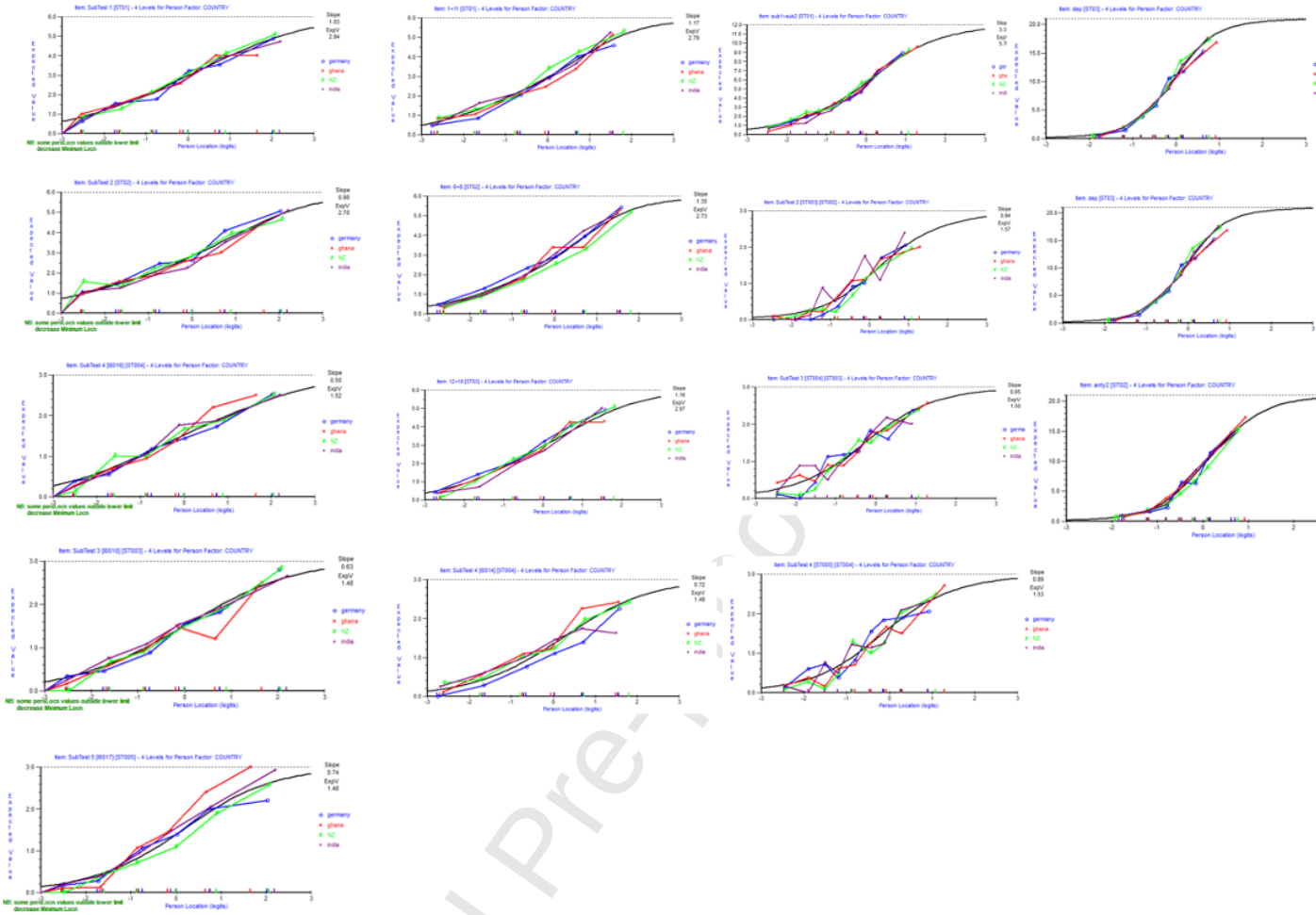

Figure 4. Differential Item Functioning (DIF) curves of the depression (first column from the left), stress (second column from the left), anxiety (third column from the left) subscales, and the full scale (fourth column from the left) for country of origin of the participants.

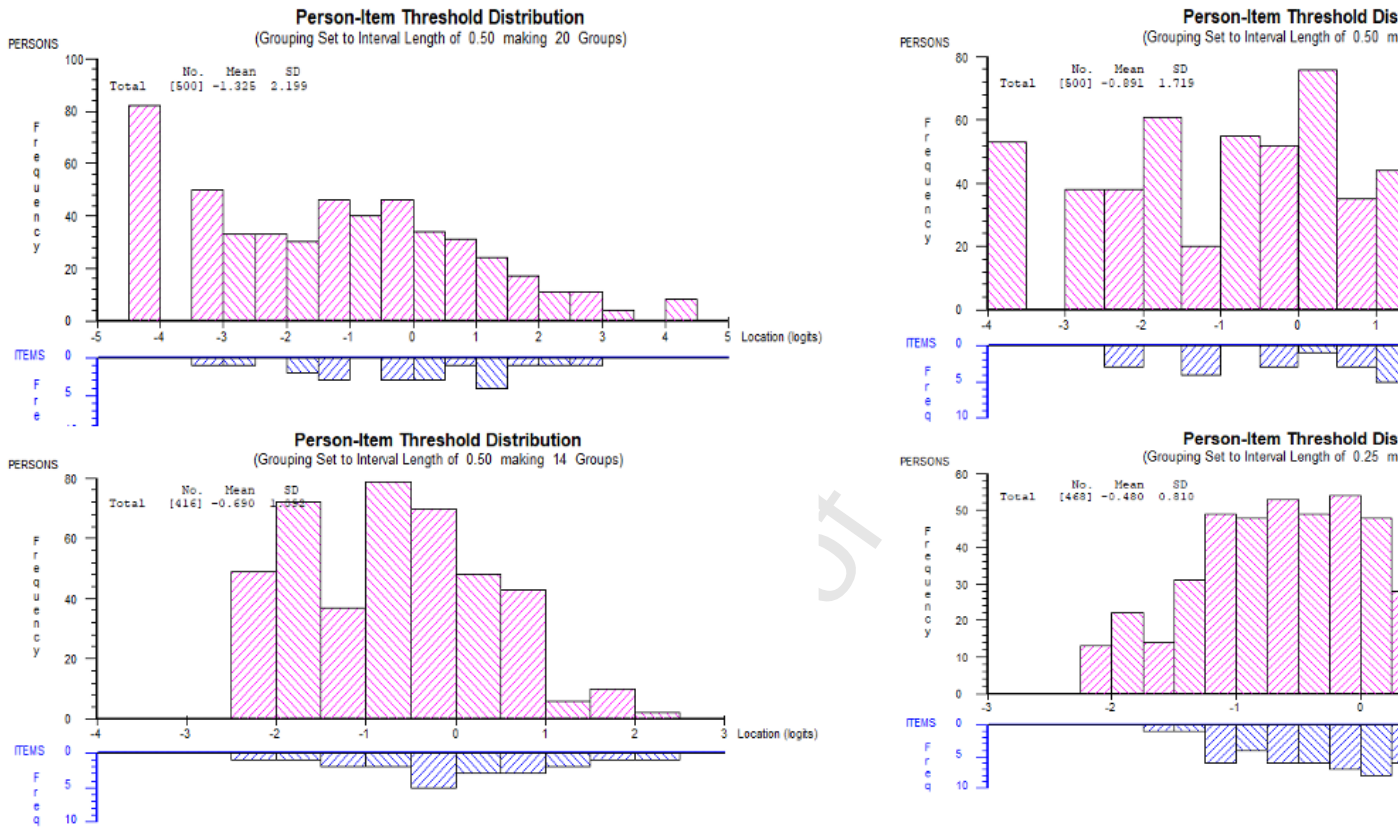

Figure 5. Person-item thresholds distributions for the depression (top left), stress (top right), and anxiety (bottom left) subscales, and the full scale (bottom right)

Table 1. Rasch model fit statistics for the initial and final analysis for all the scales of the DASS-21 (n=500).

Analyses(A)	Item fit residual		Person fit residual		Goodness of fit		PSI	Unidimensionality *t*-test	
	Mean	SD	Mean	SD	χ^2 (df)	*p*		%	Lower bound
Depression subscale									
A1 Initial	-0.15	2.47	-0.42	1.48	60.80(21)	<0.001	0.83	0.0	-1.9 (YES)
A2 Final	-0.08	0.61	-0.46	1.35	40.32(30)	0.099	0.86	1.8	-0.1 (YES)
Stress subscale									
A1 Initial	-0.06	1.61	-0.47	1.64	49.00(21)	<0.001	0.81	0.0	-1.9 (YES)
A2 Final	-0.19	2.60	-0.51	1.34	29.64(20)	0.076	0.82	0.6	-1.3 (YES)
Anxiety subscale									
A1 Initial	-0.11	3.24	-0.38	1.36	186.89(42)	<0.001	0.71	0.0	-1.9 (YES)
A2 Testlets	0.78	2.97	-0.34	1.07	71.60(40)	0.002	0.74	0.2	-1.7 (YES)
A3 Final	-0.26	3.50	-0.40	1.03	31.22(32)	0.51	0.74	0.2	-1.7 (YES)
DASS-21-Full scale									
A1 Initial	0.03	2.87	-0.50	2.35	189.96(21)	<0.001	0.92	2.0	0.1 (YES)
A2 Final	-0.28	0.80	-0.51	1.06	23.93(18)	0.16	0.87	1.6	-0.3 (YES)

Note: PSI=Person Separation Index without extremes.

Table 2. Items fit statistics including the initial and final analyses for all the scales of the DASS-21 (*n*=500).

Item No.	Item content	Location	Fit residual	Chi-square
Depression subscale				
3	I couldn't seem to experience any positive feeling at all	0.09	1.90	4.28
5	I found it difficult to work up the initiative to do things	-0.47	3.49*	11.73
10	I felt that I had nothing to look forward to I felt that I had nothing to look forward to	-0.01	-2.93*	11.47
13	I felt downhearted and blue	-0.15	-0.30	5.23
16	I was unable to become enthusiastic about anything	0.04	-2.17	12.11
17	I felt I wasn't worth much as a person	0.13	-2.39	15.28*
21	I felt that life was meaningless	0.37	1.34	0.69
	Final Analysis: 2 Testlets (T)			
T1	Items: 3=13	-0.03	0.113	3.727
T2	Items: 5+21	-0.05	0.785	5.268
10		-0.042	-0.85	14.107
16		-0.002	-0.059	7.233
17		0.124	-0.401	9.981
Stress subscale				
1	I found it hard to wind down	-0.21	2.59*	8.36
6	I tended to over-react to situations	-0.12	-0.39	5.53
8	I felt that I was using a lot of nervous energy	0.09	0.02	2.90
11	I found myself getting agitated	0.01	-1.91	12.56*
12	I found it difficult to relax	-0.20	-0.95	5.82
14	I was intolerant of anything that kept me from getting on with what I was doing	0.16	1.85	3.94
18	I felt that I was rather touchy	0.28	-0.81	7.88
	Final Analysis: 3 Testlets (T)			

T1	Items: 1+11	-0.14	-0.56	3.59
T2	Items: 6+8	-0.08	-0.96	8.07
T3	Items: 12+18	0.02	-2.71	13.30
14		0.19	3.45	4.69

Anxiety subscale

2	I was aware of dryness of my mouth	-0.45	6.27*	86.16*
4	I experienced breathing difficulty (e.g. excessively rapid breathing, breathlessness in the absence of physical exertion)	0.23	0.26	10.49
7	I experienced trembling (e.g. in the hands)	0.35	-1.04	16.11
9	I was worried about situations in which I might panic and make a fool of myself	-0.28	0.73	4.95
15	I felt I was close to panic	0.15	-3.39*	32.09*
19	I was aware of the action of my heart in the absence of physical exertion (e.g. sense of heart rate increase, heart missing a beat)	-0.14	-0.43	9.12
20	I felt scared without any good reason	0.13	-3.23*	27.97*
	Analysis 2: First 2 Testlets (T)			
T1	Items: 2+20	-0.10	0.40	23.62*
T2	Items: 4+15	0.23	-3.17	21.36*
7		0.35	0.14	13.02
9		-0.31	2.14	4.94
19		-0.16	0.87	8.66
	Final Analysis: 1 Testlet (T)			
T1	Items: T1 +T2	0.04	-5.34	11.20
7		0.37	0.41	10.55
9		-0.27	2.55	2.76
19		-0.13	1.35	6.71

DASS-21-Full scale

	Stress subscale	-0.11	-1.06	10.54
	Anxiety subscale	0.17	0.54	5.48
	Depression subscale	-0.05	-0.31	7.91

Table 3. Ordinal-to-interval level transformation for all the DASS-21scales ordinal total scores in logits and the scale metric (*n*=500)

Depression subscale			Stress subscale			Anxiety Subscale			DASS-21-Full scale					
Ordinal scores	Logit	Interval scores	Ordinal scores	Logit	Interval scores	Ordinal scores	Logit	Interval scores	Ordinal scores	logit	Interval scores	Ordinal scores	Logits	Interval scores
0	-4.45	0.00	0	-3.81	0.00	0	-3.28	0.00	0	-2.77	0.00	32	0.03	31.97
1	-3.49	2.36	1	-2.98	2.46	1	-2.47	2.63	1	-2.20	6.47	33	0.06	32.39
2	-2.79	4.10	2	-2.37	4.23	2	-1.91	4.47	2	-1.85	10.50	34	0.10	32.80
3	-2.27	5.38	3	-1.94	5.52	3	-1.52	5.75	3	-1.63	13.02	35	0.14	33.23
4	-1.85	6.43	4	-1.58	6.56	4	-1.21	6.75	4	-1.47	14.85	36	0.17	33.65
5	-1.48	7.34	5	-1.27	7.48	5	-0.95	7.58	5	-1.34	16.30	37	0.21	34.08

6	-1.15	8.16	6	-0.98	8.33	6	-0.73	8.30	6	-1.24	17.51	38	0.25	34.51
7	-0.85	8.91	7	-0.72	9.12	7	-0.54	8.94	7	-1.15	18.55	39	0.29	34.94
8	-0.57	9.60	8	-0.46	9.86	8	-0.36	9.52	8	-1.07	19.46	40	0.33	35.37
9	-0.30	10.26	9	-0.22	10.56	9	-0.19	10.06	9	-0.99	20.30	41	0.36	35.81
10	-0.04	10.89	10	0.00	11.24	10	-0.03	10.58	10	-0.93	21.05	42	0.40	36.25
11	0.21	11.51	11	0.22	11.87	11	0.13	11.09	11	-0.87	21.76	43	0.44	36.71
12	0.45	12.12	12	0.43	12.48	12	0.28	11.60	12	-0.81	22.41	44	0.48	37.17
13	0.70	12.73	13	0.62	13.06	13	0.44	12.12	13	-0.76	23.03	45	0.52	37.65
14	0.95	13.35	14	0.82	13.64	14	0.61	12.67	14	-0.70	23.61	46	0.57	38.14
15	1.21	14.00	15	1.02	14.23	15	0.79	13.26	15	-0.66	24.17	47	0.61	38.64
16	1.49	14.69	16	1.23	14.86	16	1.00	13.92	16	-0.61	24.71	48	0.66	39.16
17	1.80	15.46	17	1.47	15.55	17	1.23	14.68	17	-0.56	25.24	49	0.70	39.70
18	2.16	16.33	18	1.74	16.35	18	1.51	15.60	18	-0.52	25.74	50	0.75	40.27
19	2.59	17.41	19	2.08	17.36	19	1.88	16.78	19	-0.48	26.23	51	0.81	40.86

											-0.43	26.71		0.86	41.50
20	3.20	18.90	20	2.59	18.84	20	2.41	18.50	20				52		
											-0.39	27.18		0.92	42.19
21	4.04	21.00	21	3.32	21.00	21	3.17	21.00	21				53		
											-0.35	27.64		0.99	42.92
										22			54		
											-0.31	28.09		1.06	43.73
										23			55		
											-0.27	28.54		1.13	44.62
										24			56		
											-0.24	28.97		1.22	45.64
										25			57		
											-0.20	29.41		1.33	46.83
										26			58		
											-0.16	29.84		1.45	48.26
										27			59		
											-0.12	30.28		1.61	50.06
										28			60		
											-0.08	30.70		1.83	52.55
										29			61		
											-0.05	31.12		2.18	56.55
										30			62		
											-0.01	31.55		2.74	63.00
										31			63		

Note: *This conversion table can only be used for complete responses to each of scale's items. To use this table, ordinal raw scores (ordinal scores) should be obtained by adding the observed scores for items on each of the scales. Next, match the ordinal total score (0-21 for subscales; 0-63 for the full scale) to the corresponding interval score in the right column of each scale (0-21 for subscales; 0-63 for the full scale). A final converted score between 0 to 21 for the subscales, and 0 to 63 for the full scale, will be obtained, with higher scores corresponding to higher levels of distress. Alternatively, the current authors can be contacted for guidance on how to use this conversion algorithm.*

Highlights:

- Psychological distress can be reliably measured with the DASS-21
- The DASS-21 can be fitted to the unidimensional Rasch model of measurement.
- The robust measurement invariance enhances the universal acceptability of the interval-level scores of the DASS-21.
- The ordinal-to-interval conversion tables developed improve the precision of the DASS-21.
- The Rasch-converted interval scores of the DASS-21are superior for parametric statistics.