Depression, Anxiety, and Stress: How should clinicians interpret the total and subscale scores of the 21-item Depression, Anxiety, and Stress Scales (DASS-21)?

Article *in* Psychological Reports · June 2018

4 authors:

Eu Gene Chin
South East Missouri State University

Erin Buchanan
Missouri State University

Chad Ebesutani
Duksung Women's University

John Young
Manukau Institute of Technology

Abstract

The 21-item Depression, Anxiety, and Stress Scales (DASS-21) is a self-report measure that is easy to administer, quick to score, and is freely available. Widely used in diverse settings and populations, confirmatory factor analytic evidence has accumulated for a bifactor model underlying this multidimensional measure. Studies employing an exploratory bifactor approach to more closely examine its underlying structure and inter-relations of factors, however, have been scarce. This is unfortunate because confirmatory techniques often employ indirect ways of handling model misspecification, whereas exploratory methods enable more direct approaches. Moreover, more precise approaches to modeling an exploratory bifactor structure have not been examined with the DASS-21. Based on several large samples of undergraduate students in the United States, the first two parts of the paper (Studies 1 and 2) utilized both exploratory (M = 19.7 years of age) and confirmatory factor analytic methods (M = 19.7 years of age) following those presented by contemporary multidimensional modeling theorists. Building upon these results, the third part of the paper (Study 3; M = 20.0 years of age) examined sensitivity-/specificity-related indices to provide cut-off score recommendations for a revised DASS-21 instrument based on a newly-identified and supported bifactor structure. Implications of these results are discussed in terms of taxonomy, challenges inherent in multidimensional modeling, and potential use of the revised DASS-21 measure as a component of an actuarial decision-making strategy to inform clinical referrals.

Keywords: anxiety, depression, stress, bifactor, multidimensionality

Short Title: Interpreting the total and subscale scores of the 21-item Depression, Anxiety, and Stress Scales (DASS-21)

Depression, Anxiety, and Stress: How should clinicians interpret the total and subscale scores of the 21-item Depression, Anxiety, and Stress Scales (DASS-21)?

Mental health and behavioral disorders, when combined, function as the third leading cause of disability in the United States (US Burden of Disease Collaborators, 2013). Within this category, anxiety and major depressive disorders make up two of the top three contributors of disabilities (US Burden of Disease Collaborators, 2013). Unfortunately, the vast majority of individuals who have these disorders do not receive adequate treatment, with 78.3% of those with an anxiety disorder and 67.1% of those with a major depressive episode in the past year not receiving any mental health treatment at all (Wang et al., 2005). A myriad of barriers stands in the way of access to proper treatments such as lack of mental health personnel, time-constraints, and limited fiscal resources (Collins, Westra, Dozois, & Burns, 2004), particularly for broad mental health screening. These systemic challenges undermine efforts to regularly detect individuals in need because many clinical instruments require extensive training, incur significant costs to use, or demand substantial time to administer, score, and interpret.

The 21-item Depression, Anxiety, and Stress Scales (DASS-21) is a self-report measure that is easy to administer, quick to score, and freely available (Lovibond & Lovibond, 1995). It is precisely due to these characteristics that it makes a popular screening tool (Tran, Tran, & Fisher, 2013), treatment outcome measure (Ronk, Korman, Hooke, & Page, 2013), and a useful supplement to more comprehensive assessments with clinical populations such as patients with cancer (Fox, Lilis, Gerhart, Hoerger, & Duberstein, 2017) or obstructive sleep apnea (Nanthakumar et al., 2017). The scale also has contemporary support for its application in numerous international settings, including diverse cultures such as Germany, Russia, and China

(Brailovskaia et al., 2017); the United States, Poland, Russia, and the United Kingdom (Scholten, Velten, Bieda, Chi Zhang, & Margraf, 2017); and Australia, Chile, China, and Malaysia (Mellor et al., 2015).

Factor Analytic Issues with the DASS-21

The DASS-21 was originally conceptualized as a correlated three-factor model (Figure 1; Lovibond & Lovibond, 1995), with anxiety, depression, and stress conceptualized as distinct but correlated factors that, when taken together, captured a full spectrum of negative emotional states. The vast majority of studies done with the DASS-21 when used with clinical and community adult samples thus far have either explicitly modeled this 3-factor (correlated) structure (Antony, Bieling, Cox, Enns, & Swinson, 1998; Gloster et al., 2008; Nanthakumar et al., 2017; Scholten et al., 2017) or implicitly assumed that the measure fits the aforementioned structure without conducting tests related to that assumption (Ownsworth, Little, Turner, Hawkes, & Shum, 2008; Ronk et al., 2013; Brailovskaia et al., 2017). Other factor structures have also been suggested by various research groups (e.g., two-factor structure; Fox, Lillis, Gerhart, Hoerger, & Duberstein, 2017), but the literature has yielded more consistent support for a correlated three-factor model when used with adult populations. These studies, however, did not make direct comparisons of the traditional structure to a bifactor model (Figure 2), thus potentially obviating important opportunities to advance psychometric understanding of this widely used instrument.

To provide context on the potential advantages of bifactor models, these approaches are constructed to denote that the instrument simultaneously measures a common underlying factor common to all potential subfactors (Reise, Moore, & Haviland, 2010) as well as domain-specific measurement. For the DASS-21, this suggests that the total score measures a common underlying factor that is able to provide useful information over and above anxiety, depression, and stress subscales as individual constructs. The official manual of the DASS, for example, recommends that users add or average the three subscales together to produce a composite measure of negative emotional symptoms (Lovibond & Lovibond, 1995). This type of procedure may seem reasonable on face value, but it makes the assumption that the anxiety, depression, and stress remain viable subfactors when general distress (i.e., total score) is taken into account (a position that has not been extensively tested). The propensity for such a suggestion to be adopted in clinical practice is potentially high; however, treatment decisions made on this basis would not be directly informed by the majority of extant psychometric study or scalar models.

For example, in the limited subset of studies that directly compare a correlated three-factor model and a bifactor model, more support has been found for the bifactor model. Henry and Crawford (2005) were the first researchers to make such a comparison using *confirmatory* bifactor analysis. Utilizing a large non-clinical sample of adults in the United Kingdom, they made direct comparisons between seven competing models by inspecting global fit indices (i.e., statistics that give the researcher a sense of how well a proposed factor model accounts for observed patterns in the data) and, when factor models were nested within another more restricted model, chi-squared difference tests (i.e., an index that tells the researcher whether or not there are significant differences in the fit of various models; Henry & Crawford, 2005). Given these results, Henry and Crawford (2005) reported that the practice of combining the scores of the subscales appeared to have some psychometric support. Other subsequent confirmatory factor analyses have found similar support for a bifactor model of the DASS-21

(Szabo, 2010; Gomez, 2013; Shaw, Campbell, Runions, & Zubrick, 2017), although *exploratory* bifactor analyses studies with the DASS-21 have been scarce. It has been suggested that this could be attributed to a lack of statistical software that comes pre-packaged with the ability to model appropriate rotation methods (Reise, 2012). This limitation is unfortunate because researchers often rely on indirect ways of handling model misspecification via post hoc inspection of fit and modification indices through confirmatory factor analyses (e.g., Szabo, 2010; Shaw et al., 2017) as opposed to dealing directly with modeling problems via exploratory factor analyses, which offers selective advantages to model revision (Reise, 2012).

To date, only one study has examined the bifactor structure of the DASS-21 via an exploratory bifactor analysis (Osman et al., 2012). With a large sample of non-clinical undergraduate students in the United States, Osman et al. (2012) conducted an exploratory bifactor analyses through a Schimd-Leiman method, which is a reparameterization (orthogonalization) of a second-order exploratory solution (Reise, 2012).[1] Osman et al. (2012) reported that the general distress factor associated with this model accounted for the largest proportion of common variance of the DASS-21 items (61.9%), and that depression, anxiety, and stress factors accounted for smaller amounts of unique variance in the context of a bifactor solution (14.7%, 12.3%, and 11.1% respectively). Osman et al. (2012) also reported that the majority of the items were more strongly associated with the general distress factor than to their own hypothesized specific subfactors, which suggested that interpretation of the scale as a whole

[1] A second-order model for the DASS-21 is a re-expression of the correlations among subfactors (e.g., anxiety, depression, and stress) in a three factor, correlated traits model. In contrast with a bifactor model, a second-order model does not allow researchers to explore the extent to which item variance is split between general distress and subfactors because the relationship between the "higher order" factor (e.g., general distress) and an item of interest is necessarily mediated by a subfactor (e.g., anxiety).

could be more viable and clinically informative than the traditional method. Results, however, were limited to a mono-method bias (i.e., all results were based on self-report measures) and modeling assumptions inherent within a Schmid-Leiman method of conducting an exploratory bifactor analysis.

Although Osman et al. (2012) advanced psychometric understanding of the DASS-21, the results were not without limitations. Reise (2012) and Jennrich and Bentler (2011) expressed concerns that the Schmid-Leiman method may have problems modeling a bifactor structure because it (1) assumes that all items have zero cross-loadings and (2) imposes proportionality constraints when calculating the item loadings for the subfactors and general factor. In other words, the ratio of item to subfactor and subfactor to general factor loadings needs to be equal within each subgroup cluster (e.g., all items within the anxiety subgroup) in order for the Schmid-Leiman method to generate precise estimates of factor loadings in the bifactor model (Reise, 2012). This forced proportional pattern of loadings is not likely to be true in real datasets; consequently, alternative methods that obviate having to make such unlikely assumptions should be explored (Reise, 2012), such as the Jennrich-Bentler (2011) bifactor rotation method.

Current Study

Following these recently proposed recommendations for enhanced bifactor modeling, the current paper addresses the highlighted limitations in understanding the DASS-21 from an exploratory bifactor perspective through a series of studies. In the first study, we evaluated an exploratory bifactor model of the DASS-21 via the Jennrich and Bentler's (2011) bifactor rotation method and directly compared the resultant solution to that resulting from a Schmid-Leiman bifactor solution using the same data. Second, we examined three confirmatory bifactor models, based on (1) the Jennrich-Bentler method, (2) the Schmid-Leiman method, and (3) the original bifactor model proposed by Henry and Crawford (2005). In the third study, we addressed the mono-method bias inherent in Osman et al.'s (2012) study by examining how the DASS-21 scales performs relative to results obtained from a diagnostic semi-structured interview.

Study 1

Study 1 compared two different scaling methods for the DASS-21 derived from the exploratory bifactor approaches outlined above (i.e., Jennrich-Bentler and Schmid-Leiman methods). We paid particular attention to contextualizing item factor loadings as a function of their relationship to the general distress factor and specific subfactors based on the following guidelines: (1) loads strongly (i.e. > .30; McDonald, 1999) on the hypothesized subscale factor (2) loads strongly (i.e. > .30) on the general distress factor, and (3) displays no strong loadings (i.e., < .30) on other factors.

Method

Participants and Procedure

A diverse sample of college students (*N* = 876) from a large, Southeastern University was recruited to complete a battery of measures online that included the DASS-21. The DASS-21 is a Likert-type scale where respondents indicate their level of agreement/disagreement (ranging from 0 = did not apply to me at all to 3 = applied to me very much, or most of the time) to a target statement (e.g., I felt I was close to panic). The study was announced in classes, and students were offered extra credit for their participation. This sample was randomly split into an EFA subsample (Study 1; *n* = 445) and a CFA subsample (Study 2; *n* = 431). Little's Test for

MCAR data was analyzed on these subsamples for the DASS-21, and both tests were significant, $\chi^2_{EFA}(297) = 445.85$, $p < .001$ and $\chi^2_{CFA}(317) = 418.17$, $p < .001$. Data visualization of missing data patterns with the *mice* package in *R*, however, revealed no apparent pattern as each item was missing less than 1% of data, ranging from zero to three missing data points for each DASS-21 item. Because Little's MCAR test (1) does not identify specific variables that violate MCAR, (2) is only useful for testing an omnibus hypothesis that is unlikely to hold in the first place, and (3) assumes missing data patterns share a common covariance matrix (Enders, 2010), we conducted a series of *t*-tests (one for each DASS-21 item) with missing data indicators as grouping variables (i.e., 0 = no missing items on the DASS-21; 1 = at least one missing item on the DASS-21) to obtain a more detailed analyses of missing data patterns (Enders, 2010). We used corrected *p*-values based on Levene's test for equality of variance and controlled for Type I error rates via the Bonferonni-Holms procedure (Holm, 1979). None of the DASS-21 items were significant in relation with missing data indicators, suggesting that missing data pattern was MCAR in both subsamples. Listwise deletion with a small amount of missing data (< 10%) with a MCAR missing data pattern should not significantly bias parameter estimates (Tabachnick & Fidell, 2012). Nineteen participants (4.27%) were excluded due to missing data from the EFA subsample, producing a final EFA subsample of 426 participants (see Table 1 for demographic information).

Comparing the Jennrich-Bentler and Schmid-Leiman Methods

We first employed a bifactor EFA via a bifactor rotation criterion (Jennrich & Bentler, 2011) among the 21 DASS items on our final bifactor EFA subsample, positing a "general distress" factor common to all items and three specific factors corresponding to the three hypothesized subscales (Figure 2). This analysis is a standard exploratory factor analysis with a bi-factor rotation criterion. Factor rotation is used in factor analysis to maximize the interpretability of factor loadings because for any good-fitting solution, there exists an infinite number of factor loading patterns (Brown, 2015). A bifactor rotation criterion is a type of rotation that minimizes the departure of the factor loading matrix from a bifactor structure in order to produce a rotated loading matrix with an approximate bifactor structure (Jennrich & Bentler, 2011). Since the response choices were ordinal, we used weighted least squares estimation (WLS) to model the data. WLS and bifactor rotation criterion are both available in the *psych* package in *R* (Revelle, 2009). Based on the bifactor EFA results, we identified items based on factor loading guidelines previously delineated at the beginning of this study.

For comparison purposes, we employed a Schmid-Leiman bifactor EFA on the same data. We conducted this analysis in *R* using oblique (promax) rotation. The Schmid-Leiman method is composed of two general steps (Schmid & Leiman, 1957). First, the procedure extracts a specified number of group factors from the correlations among items and performs an oblique rotation on the solution to produce an intermediary correlated traits model. Second, the procedure extracts a second-order factor from the correlations among the group factors in this intermediary correlated traits model and performs an orthogonal rotation to produce a final model that has uncorrelated second-order and group factors with corresponding factor loadings. Since the response choices are ordinal, we used polychoric correlation matrices and the minimum residual solution estimation method, which are also available in the *psych* package.

Results

Jennrich-Bentler and Schmid-Leiman Analytic Techniques

Results of the bifactor EFA appear in Table 2. As can be seen, the results of our Jennrich-Bentler bifactor EFA suggested that all items loaded onto the general distress factor (>.30). Six items also loaded adequately and cleanly on the depression specific factor. Very few items, however, loaded onto their theoretically-consistent factors for anxiety and stress (i.e., two items loaded onto anxiety and two items loaded onto stress). These results have interesting and potentially important implications. For example, they suggest that the anxiety and stress subscales may not be very useful subscales above and beyond the general distress factor, given that their interpretation was limited due to comprising only 2 subscale-specific items. The resulting representation of this model is seen in Figure 3. Although an unfamiliar model for the well-known DASS-21, what this model suggests is that general distress among individuals includes the common experience of anxiety and stress, whereas depression, on the other hand, is a uniquely distressing phenomenon associated with unique item variation relative to the other DASS items. Not only does this speak to the psychometric properties of the DASS-21, but also potentially to the nature and structure of psychopathology related to these inter-related concepts of stress, anxiety, and depression. In comparison, the Schmid-Leiman bifactor EFA results (Table 2) suggested that all three subscales may contribute meaningful information above and beyond the general distress factor. All items for this Schmid-Leiman bifactor model loaded onto the general factor, six loaded on the depression subscale, three loaded on the anxiety subscale, and three loaded onto the stress subscale.

Based on these results, as noted above, a new way of conceptualizing and modeling the inter-relations of stress, anxiety, and depression appears warranted. We therefore conducted a follow-up study to more closely examine the psychometric properties from this new structural perspective of the DASS-21.

Study 2

As a follow-up to study 1, we compared the fit indices and factors loadings of the following DASS-21 confirmatory bifactor models:

(1) a bifactor model with a revised depression subfactor (as suggested by the results of the Jennrich-Bentler [2011] method in Study 1);

(2) a bifactor model with revised anxiety, depression, and stress subfactors (as suggested by the results of the Schmid-Leiman method in Study 1); and

(3) a bifactor model with the original three subfactors (Henry & Crawford, 2005).

These confirmatory bifactor models will henceforth be called the bifactor models 1, 2, and 3.

Participants and Procedure

As mentioned in Study 1, a CFA subsample (n = 431) was randomly selected from a larger sample of diverse college students (N = 876) who completed a battery of measures online. Twenty participants (4.64%) were excluded due to having missing data, producing a final sample of 411 participants (see Table 1 for demographic information).

Confirmatory Factor Analyses

Compared with exploratory factor analytic methods, confirmatory factor analytic methods require the researchers to specify all aspects of the factor model such as the number of factors and pattern of item-factor loadings (Brown, 2015). Consequently, a confirmatory approach is generally utilized in later phases of psychometric investigations, that is, after the underlying structure has been analyzed using an exploratory approach and there is reasonable theoretical support for the proposed factor structure (Brown, 2015).

All confirmatory analyses were conducted with the seventh version of Mplus (Muthén & Muthén, 1998-2012). The following fit indices were used to evaluate model fit results for all confirmatory analytic models that converged: Comparative Fit Index (CFI; Bentler, 1990), Tucker-Lewis Index (TLI; Tucker & Lewis, 1973), Root Mean Square Error of Approximation (RMSEA; Steiger, 1990), and Weighted Root Mean Square Residual (WRMR; Yu, 2002). Following published standards for interpretation, high CFI, TLI values (>.95; Hu & Bentler, 1999) and low RMSEA (<.06; Browne & Cudeck, 1993) and high WRMR (>.90; Yu & Muthén, 2002) values were interpreted as indicative of strong model fit. Statistical significance of factor loadings (p <.05) was also examined as a conservative, confirmatory approach to evaluating factor loadings (Hair, Black, Babin, & Anderson, 2010).

Results

Bifactor models 2 and 3 produced residual covariance matrices that were not positive definite (Heywood cases; Dillon, Kumar, & Mulani, 1987). This implied that the input and/or the model-implied matrices for each bifactor model had a determinant of zero (Brown, 2015).[1] Although a number of different reasons could account for Heywood cases (e.g., empirical under-identification, minor data entry problems, small sample size), the most common cause is a misspecified model (i.e., when the theoretical model is significantly different from the model

[1] When a matrix has a determinant of zero, it is singular matrix. A singular matrix, unfortunately, has no inverse. Given the central role of inverse matrices in calculating other multivariate statistics related to model identification (e.g., fitting function), a measurement model that produces such a matrix should not be considered an acceptable CFA solution (Brown, 2015).

supported by the data; Brown, 2015). Brown (2015) noted that matrices composed of a total score that is simply an average or summation of subscale scores often encounter this problem because the total score shares too much redundant variance with other variables in the matrix (e.g., stress subscale). Ultimately, CFA solutions that are based on non-positive definite matrices are not deemed acceptable because a number of multivariate statistics cannot be computed correctly with a singular matrix (Brown, 2015). Consultation with Dr. Linda Muthén, director of the Mplus development team, confirmed that the best solution would be to consider other factor analytic models (L. Muthen, personal communication, August 8, 2015).

Consequently, we only evaluated the global fit indices and factor loadings for bifactor model 1, which is the bifactor model supported by the Jennrich-Bentler (2011) method. The results indicated that this model produced an excellent fit for the data, supporting the findings in study 1: CFI = .98, TLI = .97, RMSEA = .049, and WRMR = .96. Additionally, all items had significant loadings on the general distress factor and all depression questions loaded significantly onto the depression subfactor (every p-value was < .001, two-tailed test; see Figure 3). These results, in combination with improper solutions provided by bifactor models 2 and 3, suggest that bifactor model 1 had the most support among all considered models in terms of factor structure validity.

Study 3

Given results from Study 2, we calculated sensitivity/specificity-related indices for the total and the revised depression subscale scores of the DASS-21 (as suggested by bifactor model 1). Specifically, we calculated sensitivity, specificity, positive predictive values (PPV), and negative predictive values (NPV) for various cut-off scores for the total and revised depression subscale scores derived from this model. We also calculated diagnostic odds ratio (DOR) as a global measure of diagnostic accuracy. This index, together with NPV and PPV values, was used to recommend a reasonable cut-off score for applied usage in the field (for clinical and research purposes), which was optimized to reduce false negatives given the potential screening function of the DASS-21.

Participants and Procedure

Data were obtained from an independent group of participants (*N* = 293) who participated in a semi-structured diagnostic interview (the Anxiety Disorders Interview Schedule, Diagnostic Statistical Manual – 4th edition; ADIS-IV; Brown, Di Nardo, & Barlow, 2004) and who also completed a battery of measures via pen and paper. The study was announced in undergraduate classes, and students were offered extra credit for their participation. Similar to Studies 1 and 2, we utilized Little's MCAR test, data visualization of missing data patterns, and a series of *t*-tests (Enders, 2010) to examine missing data patterns in the sensitivity/specificity sample. Although Little's MCAR test was significant $\chi^2(119) = 228.45$, $p < .001$, data visualization of missing data patterns via the *mice* package in *R* revealed no apparent pattern as each item was missing less than 1% of data, ranging from zero to three missing data points for each DASS-21 item. ADIS-IV items had no missing data. We used corrected *p*-values based on Levene's test for equality of variance and controlled for Type I error rates via the Bonferonni-Holms procedure (Holm, 1979). Results suggested that none of the DASS-21 and ADIS-IV items were significant in relation to missing data indicators, suggesting that missing data pattern was MCAR for the sensitivity/specificity sample. Thus, 11 participants (3.7%) were excluded due to missing data, producing a final sample of 282 participants (see Table 1 for demographic information).

The ADIS-IV is a semi-structured diagnostic interview designed to assess current episodes of anxiety disorders and other highly comorbid disorders (e.g., mood and substance use disorders) based on the DSM-IV classification of mental health disorders (Brown et al., 2004). Brown, Di Nardo, Lehman, and Campbell (2001) reported good to excellent reliability for all DSM-IV diagnoses produced based on the ADIS-IV interview (≥ .60 kappa coefficients; Fleiss, Nee, & Landis, 1979). Likewise, Brown, Chorpita, and Barlow (1998) reported evidence that supported the construct validity of this interview, which was a five-factor model consistent with DSM-IV typology. This conceptual model for the interview has been the predominant method of interpretation since that time.

Sensitivity and Specificity Indices

Diagnostic status for Anxiety or Major Depressive Disorders (MDD), as derived from the ADIS-IV interviews, served as the reference criteria for sensitivity and specificity analyses. Anxiety disorders included Panic Disorder, Social Phobia, Generalized Anxiety Disorder, Obsessive-Compulsive Disorder, Specific Phobia, and Posttraumatic-Stress Disorder (all via DSM-IV taxonomy). The presence of any diagnosis in *any* of these domains (including MDD) yielded a discriminant value of “1” (or “positive”) and the absence of diagnosis across *all of these* domains yielded a value of “0” (or “absent”). On the other hand, diagnostic status for MDD served as the sole criterion for the revised depression subscale of the DASS-21.

All sensitivity and specificity indices were calculated via Statistical Package for Social Sciences (SPSS; IBM Corp., 2013). Calculation of PPV, NPV, and DOR values for major depressive disorder and general psychological distress were based on prevalence rates reported in national epidemiological studies: 6.9% for MDD (Substance Abuse and Mental Health Services

Administration, 2013) and 9.4% for non-specific psychological distress (9.4%; Moriarty, Zack, Holt, Chapman, & Safran, 2009).

Sensitivity refers to the probability that a person will score positively on a screener (e.g., above a cut-off score on the revised depression scale) assuming that he/she truly has the disorder(s) of interest. PPV refers to the probability that a person truly has the disorder(s) of interest given that he/she scores positively on the screener, which is generally the metric of most utility in applied contexts. On the other hand, specificity refers to the probability that a person will score negatively on a screener (e.g., below a cut-off score on the revised depression scale) assuming that he/she truly does not have the disorder(s) of interest. NPV refers to the probability that a person does not have the disorder(s) of interest given that he/she scores negatively on a screener.

The DOR of a test is the ratio of the odds of true positives (i.e., being correctly identified as having a disorder) among those with the condition relative to the odds of false positives (i.e., being incorrectly identified as having the disorder) among those without the condition. It is calculated as follows: (true positives/false negative) divided by (false positives/true negatives). DOR values are considered to be more robust global estimates of a test's diagnostic accuracy compared to other conventional indices (e.g., Area Under the Curve [AUC], Youden's Index, diagnostic effectiveness) because they account for varying sensitivity and specificity values and are less affected by disease prevalence estimates (Šimundić, 2009). Given the likelihood of broader familiarity to readers, however, we also report AUC values for comparison purposes (Figure 4). While AUC values can be useful for making general comparisons of two or more diagnostic tests, we encourage readers to keep in mind that AUC values do not say anything about sensitivity, specificity, PPV, or NPV values (Šimundić, 2009). Accordingly, it is possible to have two tests that have identical AUC values but exhibit divergent performance when applied in the context of a screening optimized to reduce false negatives.

Results

Examination of DOR, NPV, and PPV values for the DASS-21 total scale (Table 3) suggested that sixteen points or more would be a reasonable screening cut-off score for screening any type of anxiety disorder or MDD. At this cut-off score, the odds of correctly identifying an individual with any anxiety disorder or MDD were 5.4 times higher than the odds of mistakenly labeling an individual as having any anxiety disorder or MDD. A student who scores sixteen points or less (negative screen) on the DASS-21 total scale has a 93% probability of truly not having an anxiety disorder or MDD. On the other hand, a student who scores more than sixteen points (positive screen) has a 30% probability of truly having an anxiety disorder or MDD.

Examination of DOR, NPV, and PPV values for the DASS-21 revised depression subscale (Table 4) suggested that two points or more would be a reasonable cut-off score for screening for MDD. At this cut-off score, the odds of correctly identifying an individual with MDD were 11.2 times higher than the odds of mistakenly labeling an individual as having MDD. A student who scores two points or less (negative screen) on the revised depression scale has a 98% probability of truly not meeting criteria for MDD. On the other hand, a student who scores more than two points (positive screen) has an 19% probability of truly meeting criteria for MDD.

Discussion

Consistent with previous literature (Henry & Crawford, 2005; Osman et al., 2012; Shaw et al., 2017; Szabo, 2010), the DASS-21 items appear to have a bifactor structure. In contrast with previous studies, this study demonstrated that anxiety and stress subfactors of the DASS-21 do not possess strong factorial validity when examined with more recently recommended factor analytic techniques. The presence of a general distress factor in the DASS-21 should not come as a surprise in light of current seminal theories, such as the tripartite model of anxiety and depression (Clark & Watson, 1991). This model posits that anxiety and depression share a superordinate temperamental trait labeled negative affect, while a lack of positive affect distinguished unipolar depressive symptoms from anxiety symptoms. Given the superordinate nature of these temperamental constructs, it would make sense that the reported symptoms of anxiety and depression would manifest themselves in a similar pattern. Modern transdiagnostic treatments, such as the Unified Protocol (Barlow et al., 2017) were developed based on these theories and directly target the amelioration of negative affect and increase in positive affect.

In the context of a stepped care health delivery system (Bower & Gilbody, 2005), the revised DASS-21 could also be a component of an actuarial decision-making strategy to determine whether or not it is feasible or necessary to refer individuals for additional assessment. Results of using the instrument to screen in such a way could contribute to one or more subsequent strategies, each with known cost, sensitivity, specificity, and clinical utility. For example, one can be at least 90% confident that an individual scoring 16 points or below on the DASS-21 total score will not need further services for depression or anxiety. On the other hand, a student who scores more than 16 points could be asked if they would be willing to complete further assessments given that they have approximately a one in four chance (30% probability) of meeting criteria for an emotional disorder. In combination with models estimating lost productivity, negative outcomes (such as decreased happiness), medical expenditures, and other costs incurred during the course of treatment, one could arrive at a fairly accurate estimation of economic utility for each stage of stepped care and method of screening when applied to a systems-level (e.g., Markov analysis) approach (Naimark, Krahn, Naglie, Redelmeier, & Detsky, 1997). This could streamline clinical decision-making, support better quality of life, lower costs for stakeholders, and ultimately enhance the evidentiary basis for service provision within the system.

Previous studies have also examined other instruments that screen for anxiety or depression in similar samples. Khubchandani, Brey, Kotecki, Kleinfelder, and Anderson (2016), for example, examined the screening performance of a two-item version of the Patient Health Questionnaire (PHQ-2) for detecting a depression diagnosis with a large non-clinical sample of Midwestern undergraduate students (*M* = 18.9 years, *SD* = 1.2 years). Khubchandani et al. (2016) did not report sensitivity, specificity, PPV, NPV, or DOR indices, but did report values for Area Under the Curve (AUC), which is essentially a more general calculation of a measure's global diagnostic accuracy (i.e., missing information on more specific parameters such as sensitivity and specificity values; Šimundić, 2009). The AUC value for the PHQ-2 was .78, which is similar to the revised DASS-21 MDD subscale (AUC = .76). Khubchandani et al. (2016), however, did not use semi-structured interviews to verify the self-reported past 12-month diagnoses of participants, thus further comparisons to current findings on these dimensions were not feasible.

Shean and Baldwin (2008), on the other hand, did employ the use of a semi-structured interview to verify the screening performance of the Beck Depression Inventory-II (BDI-II) and the Center for Epidemiological Studies-Depression Scale (CESD) scales with a large, non-clinical sample of students in the United States (*M* = 20.3 years, *SD* = 2.7 years). The "mild depression" cutoffs of the BDI-II and CESD produced similar sensitivity, specificity, and NPV values (e.g., sensitivity = .73, specificity = .84, NPV = .94 for the BDI-II) compared with our revised DASS-21 depression subscale (sensitivity = .79; specificity = .75, NPV = .98). The PPV values for the BDI-II (.48) and CESD (.42) were noticeably higher than the same calculation for the revised DASS-21 depression subscale (.19); however, the final screening instruments in this earlier study were more than three times longer than the revised DASS-21 depression subscale (six items vs. 20 or 21). To date, no study has examined the screening performance of a measure that measures both anxiety and depression, which limited our basis for comparison using the total score of the DASS-21 from the current study.

More generally, this study illustrates the challenges in the field regarding knowing how and when to interpret subscales over total scores when dealing with multidimensionality. Many contemporary psychometric studies dealing with bifactor models often use indirect ways of handling modeling misspecification such as post hoc inspection of global fit and modification indices (Reise, 2012). If the initial model specified in CFA is markedly different than what is supported by data, however, post hoc revisions to a fitted model often lead to incorrect conclusions (Brown, 2015). Moreover, it is equally important to utilize an appropriate method of modeling an exploratory bifactor analysis. Studies 1 and 2 clearly demonstrate the different conclusions one would arrive at using the Jennrich and Bentler and Schmid-Leiman methods of modeling an exploratory bifactor structure, including computation issues that could result from too much redundant variance between the total and subscale scores. These findings are an actuarial demonstration of recently advanced statistical methodology supported by contemporary theorists in this domain (Reise, Morizot, & Hayes, 2007; Reise et al., 2010; Reise, 2012) and the findings highlight the importance of using the most precise methods available to enable optimal development of instrumentation—and through application of these instruments, more data-informed approaches to clinical practice.

Limitations

The present paper is, however, associated with limitations worth noting. This paper, for instance, used a convenient sample of predominantly Caucasian college students in the Southeastern part of the United States. Because our sample consisted of a normative undergraduate population, bifactor models two and three (examined in Study 2) may fit other populations (e.g., general adult population; Henry & Crawford, 2005). Nonetheless, the PPV and NPV values obtained in our study are still relevant for usage in screening efforts among college students who may not present in clinical settings (e.g., in a multi-staged tiered screening in collaboration with residence life, academic affairs, and student health center; Khubchandani et al., 2016). In addition, this study did not account for other extraneous factors that could influence symptom manifestation (e.g., use of psychotropic medications, socioeconomic status), which could have contributed to systematic variation in item responses (and thus error) due to not being attributed to the targets DASS constructs. Any such effects, however, were unlikely to make a substantial impact on the results due to the large sample sizes utilized across all parts of the study.

Conclusions

Despite the aforementioned limitations, results from our study provide important implications for the DASS-21. There appears to be a reliable general distress factor underlying the DASS-21 and a lack of psychometric precision for the anxiety and stress subscales when the general distress factor is taken into consideration. The total score could be a useful screening instrument for identifying individuals who could benefit from treatments targeting non-specific psychological distress. Within the context of a stepped health care delivery system, the total score could also be a component of an actuarial decision-making strategy to inform clinical referrals and lower costs for stakeholders. More research examining these issues prospectively is needed, in addition to the application of methods contained in this study to other multidimensional scales.

Declaration of conflicting interests

The authors declare that there are no conflicts of interest.

References

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10*(2), 176-181. doi: 10.1037/1040-3590.10.2.176

Barlow, D. H., Farchione, T. J., Bullus, J. R., Gallagher, M.W., Murray-Latin, H., Sauer-Zavala, S., . . . Cassiello-Robbins, C. (2017). The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for Anxiety Disorders: A randomized clinical trial. *JAMA Psychiatry, 74*(9), 875-884. doi: 10.1001/jamapsychiatry.2017.2164

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. doi: 10.1037/0033-2909.107.2.238

Bower, P., & Gilbody, S. (2005). Stepped care in psychological therapies: access, effectiveness, and efficiency. Narrative literature review. *The British Journal of Psychiatry, 186*(1), 11-17. doi: 10.1192/bjp.186.1.11

Brailovskaia, J., Schönfeld, P., Zhang, X. C., Bieda, A., Kochetkov, Y., & Margraf, J. (2017). A cross-cultural study in Germany, Russia, and China: Are resilient and social supported students protected against depression, anxiety, and stress. *Psychological Reports, 121*(2), 265-281. doi:10.1177/0033294117727745

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: The Guildford Press.

Brown, T. A., Chorpita, B. F., & Barlow, D. H. (1998). Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology, 107,* 179-192. doi: 10.1037//0021-843x.107.2.179

Brown, T. A., Di Nardo, P. D., & Barlow, D. H. (2004). *Anxiety Disorders interview schedule adult version: Client interview schedule.* Albany, NY: Graywind Publications.

Brown, T. A., Di Nardo, P. A., Lehman, C. L., & Campbell, L. A. (2001). Reliability of DSM-IV anxiety and mood disorders: Implications for the classification of emotional disorders. *Journal of Abnormal Psychology, 110,* 49-58. doi: 10.1037//0021-843x.110.1.49

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-159). Newbury Park, CA: Sage.

Clark, L. A., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*(3), 316-336. doi: 10.1037/0021-843X.100.3.316

Collins, K. A., Westra, H. A., Dozois, D. J. A., Burns, D. D. (2004). Gaps in assessing treatment for anxiety and depression: Challenges for the delivery of care. *Clinical Psychology Review, 24,* 583-616. doi: 10.1016/j.cpr.2004.06.001

Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. *Psychological Bulletin, 101,* 126-135. doi: 10.1037/0033-2909.101.1.126

Enders, C. K. (2010). *Applied missing data analysis* (1st ed.). New York, NY: The Guildford Press.

Fleiss, J. L., Nee, J. C., & Landis, J. R. (1979). Large sample variance of Kappa in cases of different sets of raters. *Psychological Bulletin, 86*(5), 974-977. doi: 10.1037//0033-2909.86.5.974

Fox, R. S., Lilis, T. A., Gerhart, J., Hoerger, M., Duberstein, P. (2017). Multiple group confirmatory factor analysis of the DASS-21 depression and anxiety scales: How do they perform in a cancer sample. *Psychological Reports 121*(3), 548-565. doi: 10.1177/0033294117727747

Gloster, A. T., Rhoades, H. M., Novy, D., Klotsche, J., Senior, A., Kunik, M., . . . Stanley, M. A. (2008). Psychometric properties of the Depression Anxiety and Stress Scale-21 in older primary care patients. *Journal of Affective Disorders*, *110*(3), 248-259. doi: 10.1016/j.jad.2008.01.023

Gomez, R. (2013). Depression Anxiety Stress Scales: Factor structure and differential item functioning across women and men. *Personality and Individual Differences*, *54*, 687-691. doi: http://dx.doi.org/10.1016/j.paid.2012.11.025

Hair, J., Black, W. C., Babin, B., & Anderson, R. E. (2010). *Multivariate data analysis.* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44*(2), 227-239. doi: 10.1348/014466505X29657

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics. 6,* 65-70.

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

IBM SPSS Statistics For Windows, Version 22.0. [Computer software]. Chicago, IL: SPSS Inc.

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*, *76*(4), 537-549. doi: 10.1007/s11336-011-9218-4

Khubchandani, J., Brey, R., Kotecki, J., Kleinfelder, J., & Anderson, J. (2016). The psychometric properties of PHQ-4 Depression and Anxiety Screening Scale among college students. *Archives of Psychiatric Nursing, 30*, 457 – 462. doi: http://dx.doi.org/10.1016/j.apnu.2016.01.014

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales*, (2nd ed.). Sydney, Australia: Psychology Foundation of Australia.

McDonald, R. P. (1999). *Test theory: A unified approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mellor, A., Vinet, E. V., Xu, X., Mamat, N. H., Richardson, B., & Roman, F. (2015). Factorial invariance of the DASS-21 among adolescents in four countries. *Journal of Psychological Assessment*, *31*(2), 138-142. doi: 10.1027/1015-5759/a000218

Moriarty, D. G., Zack, M. M., Holt, J.B., Chapman, D. P., Safran, M.A. (2009). Geographic patterns of frequent mental distress: U.S. adults, 1993-2001 and 2003-2005. *American Journal of Preventive Medicine, 46,* 497-505.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide: Seventh Edition*. Los Angeles, CA: Muthén & Muthén

Naimark, D., Krahn, M. D., Naglie, G., Redelmeier, D. A., & Detsky, A. S. (1997). Primer on medical decision analysis: Part 5-working with Markov processes. *Medical Decision Making*, *17*(2), 152-159. doi: 10.1177/0272989X9701700205

Nanthakumar, S., Bucks, R. S., Skinner, T. C., Starkstein, S., Hillman, D., James, A., & Hunter, M. (2017). Assessment of the Depression, Anxiety, and Stress Scale (DASS-21) in Untreated Obstructive Sleep Apnea (OSA). *Psychological Assessment, 29*(10), 1201-1209. doi: 10.1037/pas0000401

Osman, A., Wong, J. L., Bagge, C. L., Freedenthal, S., Gutierrez, P. M., & Lozano, G. (2012). The Depression Anxiety Stress Scales-21 (DASS-21): Further examinations of dimensions, scale reliability, and correlates. *Journal of Clinical Psychology, 68*(12), 1322-1338. doi: 10.1002/jclp.21908

Ownsworth, T., Little, T., Turner, B., Hawkes, A., & Shum, D. (2008). Assessing emotional status following acquired brain injury: The clinical potential of the depression, anxiety and stress scales. *Brain Injury, 22*(11), 858-869. doi: 10.1080/02699050802446697

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667-696. doi: 10.1080/00273171.2012.715555

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores, *Journal of Personality Assessment*, *92*(6), 544-559. doi: 10.1080/00223891.2010.496477

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*(1), 19-31. doi: 10.1007/s11136-007-9183-7

Revelle, W. (2009). *Psych: Procedures for psychological, psychometric, and personality research*. Retrieved from http://personality-project.org/r/psych.manual.pdf

Ronk, F. R., Korman, J. R., Hooke, G. R., & Page, A. C. (2013). Assessing clinical significance of treatment outcomes using the DASS-21. *Psychological Assessment, 25*(4), 1103-1110. doi: 10.1037/a0033100

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*, 53-61. doi: 10.1007/BF02289209

Scholten, S., Velten, J., Bieda, A., Zhang, X. C., & Margraf, J. (2017). Testing measurement invariance of the Depression, Anxiety, and Stress Scales (DASS-21) across four countries. *Psychological Assessment*, *29*(11), 1376-1390. doi: 10.1037/pas0000440

Shaw, T. T., Campbell, M. A., Runions, K. C., Zubrick, S. R. (2017). Properties of the DASS-21 in an Australian community adolescent population. *Journal of Clinical Psychology*, *73*(7), 879-892. doi: 10.1002/jclp.22376

Shean, G., & Baldwin, G. (2008). Sensitivity and specificity of depression questionnaires in a college-age sample. *The Journal of Genetic Psychology, 169*(3), 281-288. doi: 10.3200/gntp.169.3.281-292

Šimundić, A. M. (2009). Measures of diagnostic accuracy: Basic definitions. *Electronic Journal of the International Federation of Clinical Chemistry and Laboratory Medicine, 19*(4), 203-211.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173–180. doi: 10.1207/s15327906mbr2502_4

Substance Abuse and Mental Health Services Administration. (2013). *Results from the 2012 National Survey on Drug Use and Health: Mental Health Findings* (NSDUH Series H-47, HHS Publication No. [SMA] 13-4805). Retrieved from https://www.samhsa.gov/data/sites/default/files/NSDUHresults2012/NSDUHresults2012.pdf

Szabo, M. (2010). The short version of the Depression Anxiety Stress Scales (DASS-21): Factor structure in young adolescent sample. *Journal of Adolescence, 33*(1), 1-8. doi: 10.1016/j.adolescence.2009.05.014

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Tran, T. D., Tran, T., & Fisher, J. (2013). Validation of the depression anxiety stress scales (DASS-21) as a screening instrument for depression and anxiety in a rural community-based cohort of northern Vietnamese women. *BMC Psychiatry, 13*. doi: 10.1186/1471-244X-13-24

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10. 10.1007/BF02291170

US Burden of Disease Collaborators. (2013). The state of US Health, 1990-2010: Burden of diseases, injuries, and risk factors. *Journal of the American Medical Association, 310*(6), 591-608. doi: 10.1001/jama.2013.13805

Wang, P. S., Lane, M. L., Olfson, M., Pincus, H. A., Wells, K. B., Kessler, R. C. (2005). Twelve-Month use of mental health services in the United States. *Archives of General Psychiatry*, *62*, 629-640. doi: 10.1001/archpsyc.62.6.629

Yu, C. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes.* (Unpublished Doctoral Dissertation). University of California, Los Angeles. Retrieved from http://www.statmodel.com/download/Yudissertation.pdf

Yu, C., & Muthen, B. (2002, April). *Evaluation of model fit indices for latent variable models with categorical and continuous outcomes*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Table 1

Demographic Information for the EFA, CFA, and Sensitivity/Specificity Samples

Demographic	Study 1: EFA *n* = 426	Study 2: CFA *n* = 411	Study 3: Sensitivity/Specificity *N* = 282
	n (%)[a]	*n (%)*[b]	*n (%)*[c]
Age			
Range	17 - 54	17 - 58	18 - 53
Mode	19.0	19.0	19.0
Mean	19.7	19.7	20.0
Median	19.0	19.0	19.0
Gender			
Female	255 (59.9)	246 (59.9)	79 (28.0)
Male	171 (40.1)	165 (40.1)	202 (71.6)
Missing	0 (0)	0 (0)	1 (0.4)
Race/Ethnicity			
African American	83 (19.5)	83 (20.2)	79 (28.0)
Caucasian	317 (74.4)	296 (72.0)	184 (65.2)
Asian	9 (2.1)	13 (3.2)	6 (2.1)
Hispanic/Latino	9 (2.1)	9 (2.2)	0 (0)
Other	7 (1.6)	10 (2.4)	11 (3.9)
Missing	1 (0.2)	0 (0)	1 (0.4)
Marital Status			
Single	399 (93.7)	395 (96.1)	268 (95.0)
Co-habitating/ Domestic partner	16 (3.8)	10 (2.4)	5 (1.8)
Married	6 (1.4)	4 (1.0)	5 (1.8)
Divorced	3 (0.7)	0 (0)	1 (0.4)
Other	1 (0.2)	1 (0.2)	2 (0.7)
Missing	1 (0.2)	1 (0.2)	1 (0.4)

Note. EFA = Exploratory Factor Analyses; CFA = Confirmatory Factor Analyses.
[a]In the EFA subsample, 19 participants (4.27%) were excluded due to missing data, producing a final sample of 426 participants.
[b]In the CFA subsample, 20 participants (4.64%) were excluded due to having missing data, producing a final sample of 411 participants.
[c]In the sensitivity/specificity sample, 11 participants (3.7%) were excluded due to missing data, producing a final sample of 282 participants.

Table 2

Exploratory Bifactor Analysis Solutions on the 21-item Depression, Anxiety, and Stress Scale

	Exploratory Bifactor Analysis via Jennrich-Bentler				Exploratory Bifactor Analysis via Schmid-Leiman				Abbreviated Item Content
Scale/Item	G	DEP	ANX	STRESS	G	DEP	ANX	STRESS	
Depression									
3	**.57**	**.33**	.03	.15	**.62**	**.40**	.03	.08	couldn't experience positive feelings
5	**.49**	.14	.01	**.31**	**.50**	.24	.06	.18	difficult to do things
10	**.62**	**.42**	-.02	.01	**.65**	**.49**	-.03	.08	nothing to look forward to
13	**.70**	**.34**	-.03	.04	**.71**	**.45**	.03	.10	down-hearted and blue
16	**.65**	**.42**	-.04	-.05	**.69**	**.47**	< .01	.10	unable to be enthusiastic
17	**.60**	**.35**	.03	-.11	**.68**	**.47**	.02	-.02	felt wasn't worth much
21	**.54**	**.43**	.13	-.11	**.68**	**.54**	.00	-.18	life was meaningless
Anxiety									
2	**.35**	-.05	.10	.11	**.39**	.02	.19	.02	dryness in mouth
4	**.47**	.05	**.33**	-.03	**.57**	.08	.27	-.18	breathing difficulty
7	**.52**	-.01	**.49**	-.02	**.67**	.01	**.39**	-.26	trembling
9	**.70**	-.21	.10	.06	**.71**	-.02	**.36**	.16	worried about making a fool of myself
15	**.72**	-.15	.05	-.16	**.75**	.06	**.32**	.14	close to panic
19	**.69**	.05	.24	-.18	**.65**	.16	.26	-.20	aware of the action of the heart without physical exertion
20	**.59**	.14	.28	.01	**.71**	.23	.23	-.13	scared without any good reason
Stress									

1	**.43**	-.09	-.04	**.50**	**.41**	.01	.14	**.40**	hard to wind down
6	**.57**	-.08	-.07	.14	**.55**	.06	.20	.25	tended to over-react
8	**.69**	-.17	.19	.13	**.72**	-.05	.38	.17	using a lot of nervous energy
11	**.68**	.02	**.34**	.16	**.60**	.23	.06	**.53**	found myself getting agitated
12	**.69**	-.03	-.12	**.35**	**.65**	.17	.14	**.45**	difficult to relax
14	**.68**	-.19	-.14	-.12	**.65**	.07	.24	.28	intolerant of barriers
18	**.66**	-.05	-.05	-.10	**.66**	.16	.21	.11	rather touchy

Note. G = general distress; DEP = depression; ANX = anxiety. *Note*. A factor loading is bolded if it meets the criteria described in the procedure of study 1 as having a factor loading of greater than .30.

Table 3

Sensitivity and specificity indices for the total score of the 21-item Depression, Anxiety, and Stress Scales (DASS-21)

Cut-off score	Sensitivity	Specificity	PPV	NPV	DOR
0.5	.91	.13	.10	.94	1.61
1.5	.88	.20	.10	.94	1.89
2.5	.84	.29	.11	.95	2.14
3.5	.77	.37	.11	.94	1.90
4.5	.73	.46	.12	.94	2.28
5.5	.70	.54	.14	.95	2.69
6.5	.65	.62	.15	.95	3.09
7.5	.64	.67	.17	.95	3.60
8.5	.59	.74	.19	.95	3.94
9.5	.52	.78	.19	.94	3.77
10.5	.51	.81	.22	.94	4.54
11.5	.47	.84	.23	.94	4.60
12.5	.40	.85	.21	.93	3.72
13.5	.35	.88	.23	.93	3.92
14.5	.34	.88	.22	.93	3.64
15.5	.32	.90	.25	.93	4.15
16.5	**.29**	**.93**	**.30**	**.93**	**5.40**
17.5	.26	.94	.30	.92	5.26
18.5	.21	.94	.27	.92	4.26
19.5	.19	.95	.28	.92	4.37
20.5	.16	.95	.25	.92	3.62
21.5	.16	.95	.24	.92	3.48
22.5	.14	.96	.26	.92	3.87
23.5	.13	.96	.24	.91	3.36
24.5	.11	.96	.23	.91	3.03
26.0	.10	.96	.20	.91	2.55
27.5	.07	.96	.15	.91	1.79
28.5	.06	.96	.14	.91	1.64
29.5	.05	.98	.20	.91	2.47
31.0	.04	.99	.28	.91	3.84
33.0	.03	.99	.22	.91	2.71

Note. PPV = Positive Predictive Power; NPV = Negative Predictive Power. DOR = Diagnostic Odds Ratio. Cutoff values are the averages of two consecutive ordered observed test values. Suggested screening cut-off score for each subscale is bolded. Diagnostic status for any Anxiety or Major Depressive Disorder was based on the ADIS-IV as described in the procedure of study 3.

Table 4

Sensitivity and specificity indices for the revised depression subscale of the 21-item Depression, Anxiety, and Stress Scales (DASS-21)

Cut-off score	Sensitivity	Specificity	PPV	NPV	DOR
0.5	.79	.53	.11	.97	4.20
1.5	.79	.65	.14	.98	6.86
2.5	**.79**	**.75**	**.19**	**.98**	**11.19**
3.5	.58	.81	.19	.96	6.01
4.5	.47	.86	.20	.96	5.68
5.5	.37	.90	.22	.95	5.31
6.5	.37	.94	.32	.95	9.64
7.5	.32	.95	.32	.95	8.88
8.5	.21	.96	.29	.94	6.75
9.5	.21	.97	.31	.94	7.53
10.5	.11	.97	.23	.94	4.30

Note. PPV = Positive Predictive Power; NPV = Negative Predictive Power. DOR = Diagnostic Odds Ratio. Cutoff values are the averages of two consecutive ordered observed test values. Suggested screening cut-off score for each subscale is bolded. Diagnostic status for any Major Depressive Disorder was based on the ADIS-IV as described in the procedure of study 3.

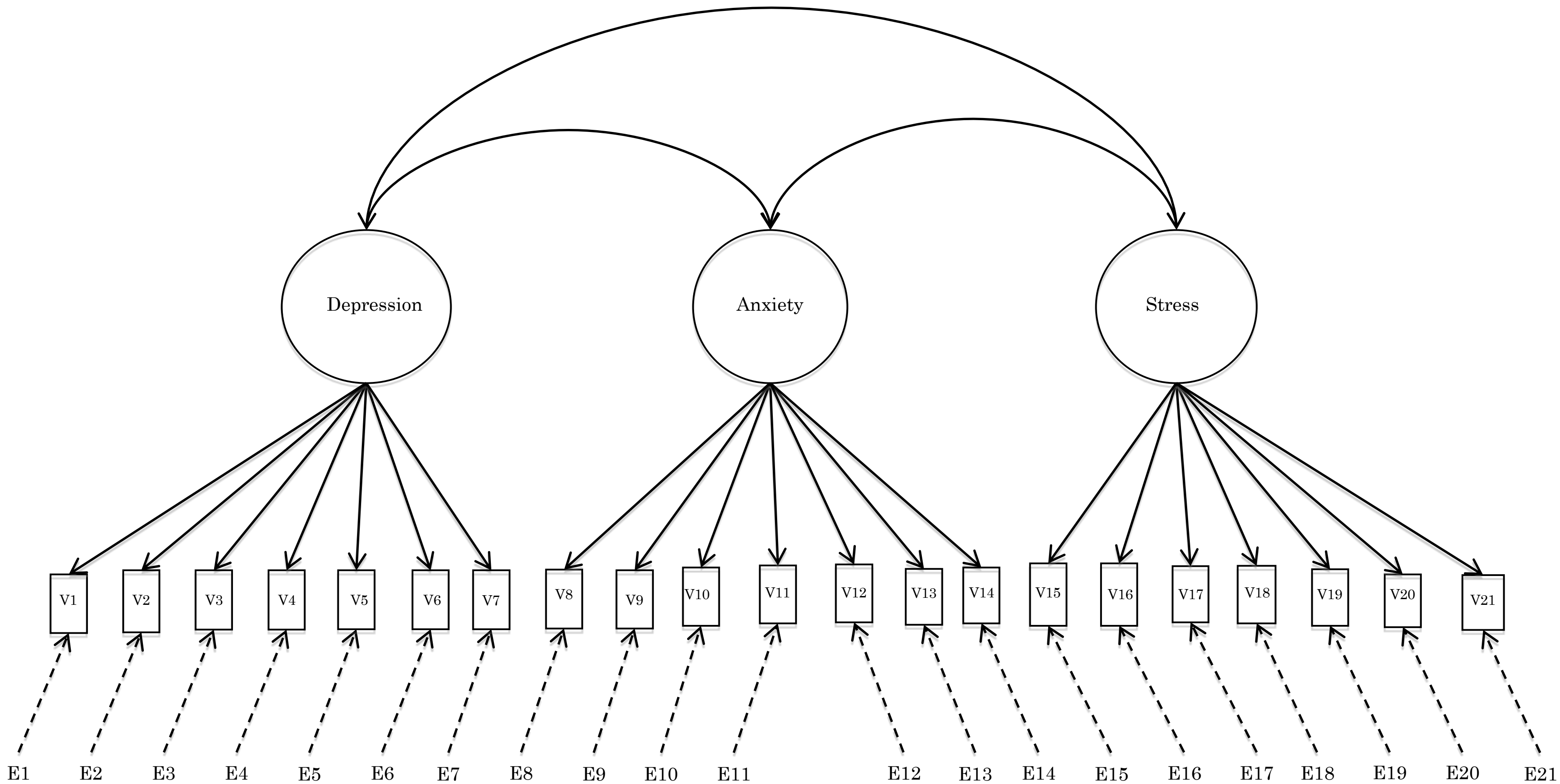


Figure 1. Path diagram for a correlated traits model. Boxes represent items and ovals represent latent constructs. Straight arrows represent correlations and dotted arrows represent error.

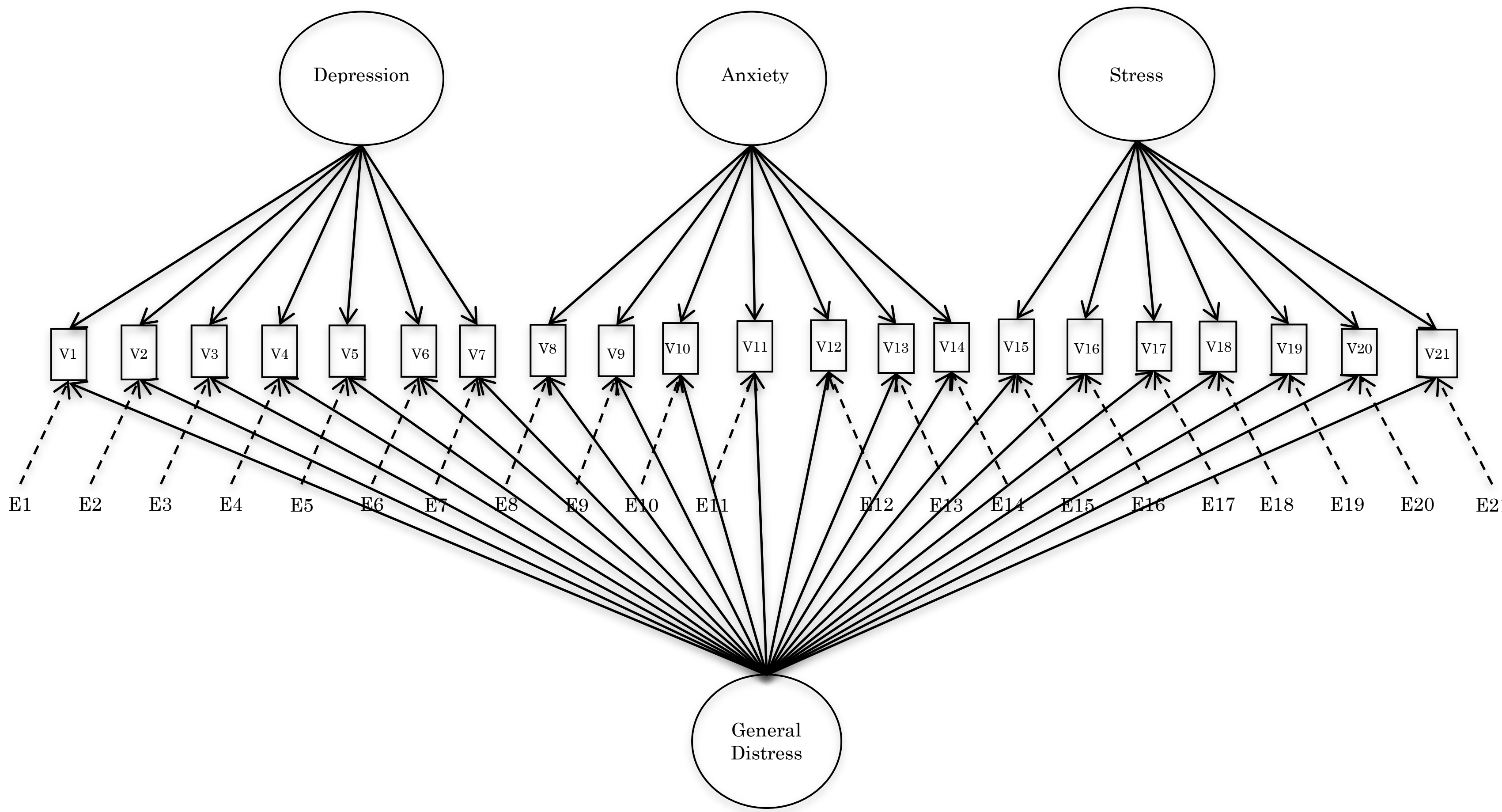


Figure 2. Bifactor model with depression, anxiety, and stress as subfactors. Boxes represent items and ovals represent latent constructs. Straight arrows represent factor loadings and dotted arrows represent error.

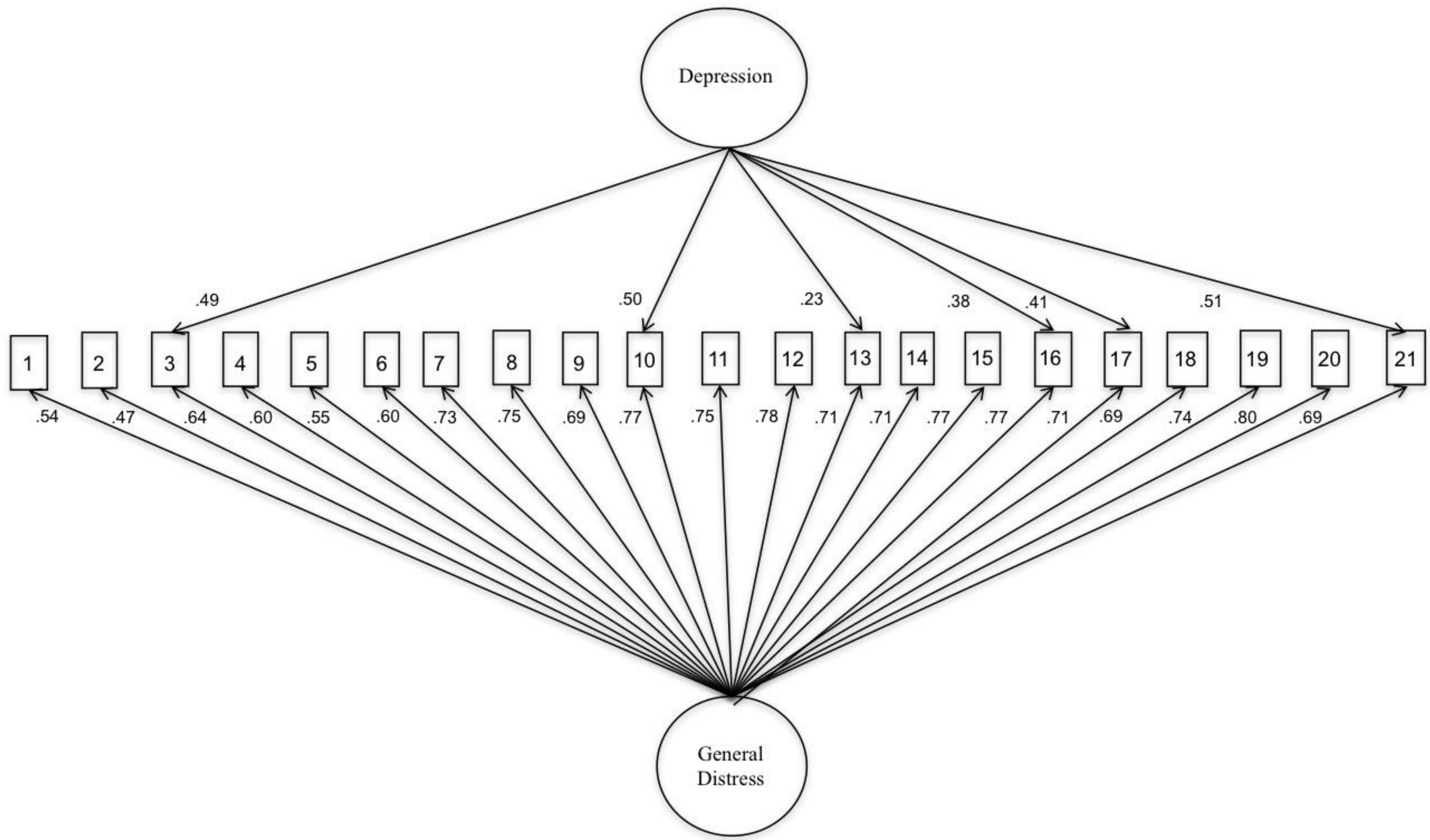
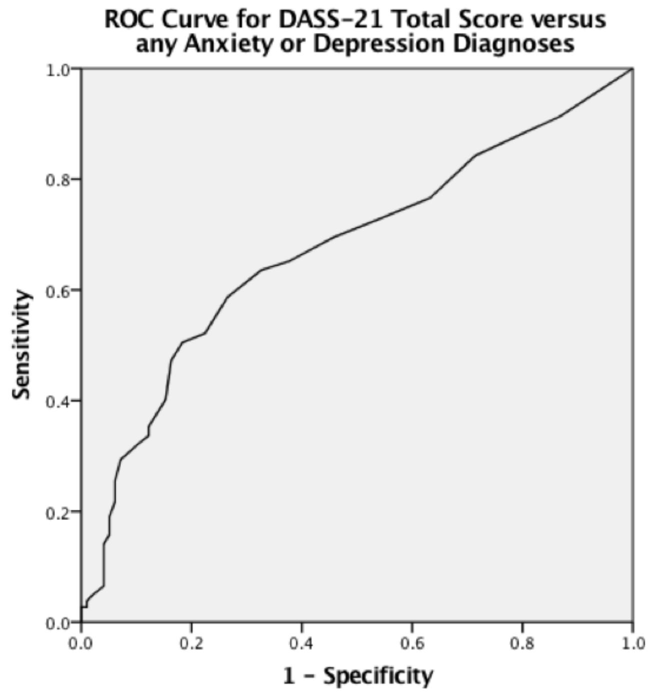
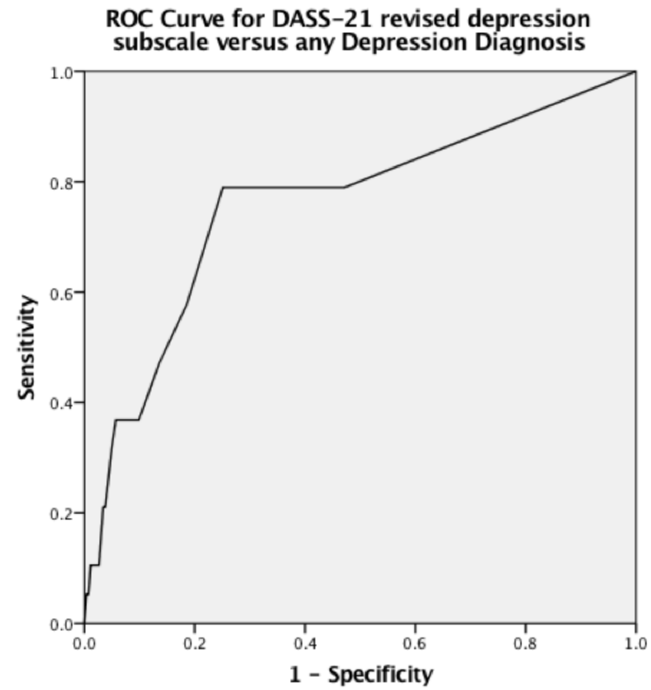


Figure 3. Path diagram for results of the confirmatory bifactor model 1, which is the bifactor model supported by the Jennrich-Bentler (2012) method. Boxes represent items and ovals represent latent constructs. All factor loadings were statistically significant ($p < .001$, two-tailed test). All error terms for indicators were uncorrelated.



Area Under the Curve (AUC)	*SD*	Significance	95% Confidence Interval	
			Lower Bound	Upper Bound
.67	.03	<.001	.61	.74



Area Under the Curve (AUC)	*SD*	Significance	95% Confidence Interval	
			Lower Bound	Upper Bound
.76	.06	<.001	.63	.89

Figure 4. The closer the curve is located to the upper-left corner and the larger the Area under the Curve (AUC), the better the instrument is at discriminating between a person with a disorder of interest and a person without the disorder of interest. While AUC values can be useful for making general comparisons of two or more diagnostic tests, AUC values do not say anything about PPV, NPV, and DOR values.