REVIEW

Systematic review of the measurement properties of the Depression Anxiety Stress Scales–21 by applying updated COSMIN methodology

Jiyeon Lee[1] · **Eun-Hyun Lee[2]** · **Seung Hei Moon[3]**

Abstract

Purpose The Depression Anxiety Stress Scales (DASS)-21 measures emotional symptoms of depression, anxiety, and stress, is relatively short, and is freely available in the public domain, which has resulted in it being applied to various clinical and non-clinical populations in many countries. The aim of this study was to systematically review the measurement properties of the DASS-21.

Methods The MEDLINE, Embase, and CINAHL databases were searched. The methodological quality of each identified study was assessed using the updated COSMIN Risk of Bias checklist. The quality of the measurement properties of the studies was rated using the updated criteria for good measurement properties. The quality of evidence was rated using a modified version of the GRADE approach.

Results This study included 48 studies in its review. The content validity of the DASS-21 demonstrated sufficient moderate-quality evidence. The instrument exhibited sufficient high-quality evidence for bifactor structural validity and internal consistency. The instrument also showed sufficient high-quality evidence for hypothesis testing of construct validity. Regarding criterion validity, only the DASS-21 Depression subscale demonstrated sufficient high-quality evidence. The measurement invariance across gender demonstrated inconsistent moderate-quality evidence. There was insufficient low-quality evidence for the reliability of each subscale. For responsiveness there was sufficient low-quality evidence for depression and stress subscales, and insufficient very-low-quality evidence for anxiety subscale.

Conclusions The DASS-21 demonstrated sufficient high-quality evidence for bifactor structural validity, internal consistency (bifactor), criterion validity (Depression subscale), and hypothesis testing for construct validity. Further studies are required to assess the other measurement properties of the DASS-21.

Keywords Systematic review · Measurement properties · Outcome measurement instrument · Depression · Anxiety · Stress

✉ Eun-Hyun Lee
  ehlee@ajou.ac.kr

  Jiyeon Lee
  jiyeon.lee@cnu.ac.kr

  Seung Hei Moon
  moonseungh@gmail.com

[1] College of Nursing, Chungnam National University, Daejeon, Republic of Korea

[2] Graduate School of Public Health, Ajou University, 164 World cup-ro, Yeongtong-Gu, Suwon 16499, Republic of Korea

[3] Department of Nursing, Graduate School, Inha University, Incheon, Republic of Korea

Introduction

According to recent global estimates, 615 million people are suffering from depression and/or anxiety, which imposes a high burden on both the affected individuals (e.g., poor function at work or school) and society as a whole (e.g., medical costs) [1]. Numerous self-reported instruments have been developed for the early screening or assessment of people with common mental health problems, of which the Depression Anxiety Stress Scales (DASS)-21 is widely used, relatively short, and freely available in the public domain [2].

The DASS-21 is a short version of the DASS-42 [3] that was developed with the initial aim of measuring negative emotional symptoms of depression and anxiety. During the development process, a third construct corresponding to irritability, tension, and agitation emerged empirically,

and was labeled as "stress." Therefore, the DASS comprises Depression, Anxiety, and Stress subscales, each of which has 14 items [4]. Antony et al. [3] selected seven items from each subscale of the original DASS, and demonstrated the reliability and validity of the DASS-21.

During the last 2 decades, the measurement properties of the original English version of the DASS-21 have been evaluated in both clinical and non-clinical populations [3, 5–8]. The DASS-21 has also been translated into 44 languages (www2.psy.unsw.edu.au/dass/), with its measurement properties studied in various countries, but concerns have emerged about discordant results. For example, its structural validity has variously been reported as having a three-factor, second-order three-factor, bifactor, two-factor, and one-factor structure [5, 9, 10].

Despite the heterogeneity of these findings, we are not aware of any systematic review of the DASS-21. The aim of this study was therefore to systematically review the measurement properties of the DASS-21, by applying the recently updated COnsensus-based Standards for selection of health Measurement INstruments (COSMIN) methodology [11–13].

Methods

Data sources and literature search strategy

The MEDLINE, Embase, and CINAHL databases were searched from their inception up to January 19, 2018. The search strategy consisted of three groups of search terms: name of instrument, type of instrument, and measurement properties. The search terms utilized to identify the name of the instrument (DASS-21) were [("depression" AND "anxiety" AND "stress") OR "depression anxiety stress scales" OR "DASS"]. The search for the type of measurement instrument utilized a modified Patient-Reported Outcome Measures (PROMs) filter developed by the Patient-Reported Outcomes Measurement Group at the University of Oxford (http://phi.uhce.ox.ac.uk). The search terms for measurement properties utilized a validated high-sensitivity search filter developed by Terwee et al. [14].

Eligibility criteria

Studies of the measurement properties of the DASS-21 and reported on in full-text articles in English were included. DASS-21 studies that involved healthy general patients, patients with chronic disease, or patients with psychiatric disorders were all eligible since the instrument was developed without limiting the population of interest. Studies of the DASS-21 involving populations younger than 14 years were not eligible because there are too few data available to confirm the validity of the scale in this age range [15]. Studies in which the DASS-21 had been used in validation tests of other instruments were excluded. Intervention studies in which the DASS-21 was used as an outcome measure were also excluded because no hypotheses about responsiveness had been evaluated.

Selection of studies

The selection process and the included studies are presented in Fig. 1. Duplicates were removed using EndNote,

Fig. 1 Flow diagram of the systematic review according to PRISMA
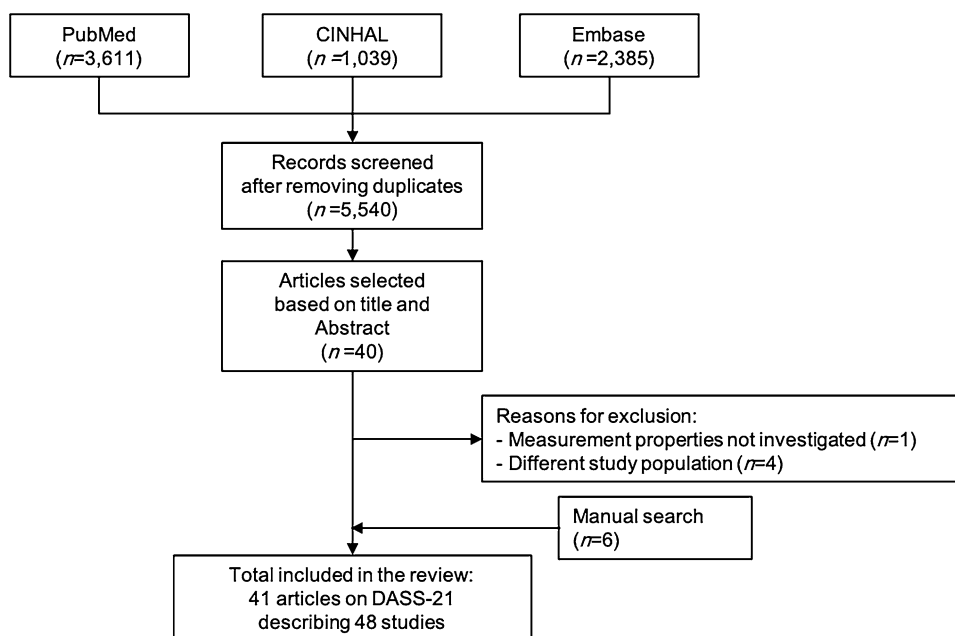
and initial screenings were conducted to remove irrelevant studies based on the title and abstract of the identified studies. The eligibility of the studies was assessed through full-text reviews. The studies were selected by two reviewers (J.L. and S.H.M.) independently. Any disagreements about inclusion were resolved by consensus with a third reviewer (E.-H.L.).

Data extraction

Data were extracted about the population in each study, such as the sample size, age, gender, and target population; on the setting, country, and language where the DASS-21 was administered; and on the results obtained for the measurement properties.

Assessing the risk of bias

The methodological risk of bias in the measurement properties of the included studies was assessed using the newly developed COSMIN Risk of Bias checklist [11, 13]. The changes in the updated COSMIN Risk of Bias checklist include the removal of standards on missing data and handling, sample size, and translation process [11]. The risk of bias in the measurement properties for each study was rated on the same 4-point scale, and determined by taking the lowest rating of any items within each measurement property.

Evaluation of measurement properties for each result

The results for the content validity of each study were rated using five criteria for relevance, one for comprehensiveness, and four for comprehensibility. The results for other measurement properties of each individual study were rated using the updated criteria for good measurement properties as "sufficient (+)", "insufficient (−)", or "indeterminate (?)" [12, 16]. Additional criteria were utilized in the present study because the updated criteria did not include the results of exploratory factor analysis (EFA) for structural validity (+; at least 50% of the variance explained by the factors), or Pearson's correlation coefficients (+; $r \geq 0.80$) for reliability.

For the rating of hypothesis testing for construct validity (convergent validity and known-groups validity), the reviewers decided a priori to apply the well-known Beck Anxiety Inventory (BAI) [17], Back Depression Inventory (BDI) [18], Hospital Anxiety and Depression Scale (HADS) [19], and Positive and Negative Affect Schedule (PANAS) [20] as comparator instruments for convergent validity. For convergent validity, r was expected to be >0.50 for the correlations with the comparator instrument if it measured a similar construct to the DASS-21. Construct validity was rated as sufficient (+) if at least 75% of the results were in accordance with the

hypotheses, insufficient (−) if at least 75% of the results were not, and indeterminate (?) if no hypotheses were defined.

Summary of the evidence and grading of the quality of evidence

For content validity, all results were qualitatively summarized into the following overall ratings for the relevance, comprehensiveness, and comprehensibility of the DASS-21: "sufficient (+)," "insufficient (−)," or "inconsistent (±)" [13]. The results of all studies for each measurement property (except content validity) were qualitatively summarized or quantitatively pooled and summarized as "sufficient (+)," "insufficient (−)," "inconsistent (±)," or "indeterminate (?) [12]. Explanations for inconsistent results were explored using conducting subgroup analyses. For the qualitative summary, the results of studies for measurement properties were summarized, such as by providing the range of values or the percentage of supported hypotheses for construct validity [11]. Quantitative pooling was conducted to perform a meta-analysis for estimating the convergent validity (Pearson correlation coefficients) for hypothesis testing. The R statistical analysis program (version 3.4.3) and the metafor package were utilized [21]. The estimated coefficient values, 95% confidence intervals, and Higgin's I^2 were calculated. The random-effects model was selected considering the heterogeneity of the studies in terms of the diversity of patient samples and various language versions.

The quality of evidence for each measurement property was graded as high, moderate, low, or very low using a modified version of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach [12] while taking into account the risk of bias (methodological quality of the studies), inconsistency of results across studies, imprecision (total sample size of the included studies), and indirectness (evidence from different populations). Indirectness was not applicable to the present study because the DASS-21 was developed without a specific target population or context of use.

If there existed a single study for each measurement property of the DASS-21, the summary and overall rating were not assessed in order to avoid overweighting by that single study. Two authors (E.-H.L. and J.L.) independently performed the above processes from data extraction to grading the quality of evidence, and all three authors convened to produce the final consensus.

Results

DASS-21 studies identified

The database search identified 7085 articles. After removing duplicates, 5540 articles were screened based on their titles and abstracts to remove irrelevant articles. Forty articles then remained, of which five were excluded after full-text screening while six additional articles were identified, resulting in 41 articles [3, 5–10, 22–55] on the measurement properties of the DASS-21. Seven articles each described two studies that examined different structures of the DASS-21. Each study of measurement properties was considered as a separate study. This systematic review included 41 articles that contained reports on 48 studies (Fig. 1).

General characteristics of the articles

Table 1 presents the characteristics of the included articles. The original English version of the DASS-21 was evaluated in 23 articles, with 13 articles from Australia [6, 7, 10, 26, 29, 33, 38, 41, 47, 49, 50, 52, 53] and eight articles from the USA [8, 22, 23, 27, 30, 42, 46, 51]. The most frequently evaluated non-English versions were Malaysian [24, 28, 34, 36, 44, 55] and Portuguese [9, 32, 37, 45]. Most of the studies (n = 18) included a healthy general population. Data were collected in non-clinic settings (n = 20), clinic settings (n = 15), or both clinic and non-clinic settings (n = 6).

Synthesized evidence

The overall ratings of the evidence for each measurement property of the DASS-21 and the quality of evidence for this scale are described below and presented in Table 2. Note that none of the included articles reported on measurement error, and so this was excluded.

Content validity

The most frequently evaluated component related to content validity was comprehensibility as evaluated by patients [22, 24, 25, 35, 37, 43, 45]. Two studies asked professionals about the comprehensiveness of the DASS-21 [39, 43], while none of studies asked either patient or professionals about the relevance of the DASS-21. There was sufficient high-quality evidence for the comprehensiveness of the DASS-21, sufficient moderate-quality evidence for comprehensibility, while there was sufficient but very-low-quality evidence for its relevance. Overall there was sufficient moderate-quality

evidence for the content validity of the DASS-21 [2, 22, 24, 25, 35, 37, 39, 43, 45].

Structural validity

In total, 45 studies from 37 articles assessed the structural validity of the DASS-21 and found several types of factor structures: three factors, as for the original DASS-21 study [3], bifactor, and one factor. Other types of factor structures such as second-order three-factor [22], two-factor [45], and four-factor [50] structures were demonstrated in single studies.

A three-factor structure of the DASS-21 was reported for 29 studies. Twenty studies (68.9%) had at least an "adequate" COSMIN methodological quality rating. Ratings lower than "adequate" were due to small samples [22], methodological flaws (orthogonal rotation [31, 37, 52, 55] or unclear estimation method [44]), reporting the structural validity of a modified DASS-21(18 items [34] or 12 items/9 items [44]), or demonstrating different item loadings compared to the original DASS-21 [35].

The structural validity of the studies that supported a three-factor structure with the same seven items for each subscale was summarized with COSMIN ratings of at least "adequate" quality [3, 5, 7–10, 23, 25, 27, 29, 38, 40, 41, 47, 48]. Among the studies that supported a three-factor structure with at least an "adequate" quality, those having issues of different item loadings [40, 49, 51, 53] or modified structures [7] were excluded from the qualitative summary. Three-factor structural validity was evaluated utilizing EFA (n = 1), confirmatory factor analysis (CFA) (n = 13), or the Rasch model (n = 1). Twelve of 15 studies (80%) exhibited a "sufficient" rating, which is above the criterion value of 75% [11], and so the overall rating for the summarized result was rated as sufficient (+); however, the quality of evidence was rated as moderate because of inconsistencies in the result ratings.

Eight studies evaluated bifactor structural validity [5, 30, 32, 39, 41–43, 54]. All of these studies had a very good methodological quality, with the quality rated as sufficient with a high quality of evidence.

Three studies (described in two articles) found one-factor structural validity [31, 46]. The results of two of the studies were methodologically of low quality, being inadequate and doubtful, and so their results were not summarized, and no grade was given to the associated evidence.

Internal consistency

Internal consistency of the DASS-21 was well supported. In the studies involving a three-factor structure, the subscale values of Cronbach's alpha/uncorrelated error [27] and the Pearson separation index [7] were overall > 0.70 except for

Table 1 Characteristics of the included articles

Authors	Population characteristics				Characteristics of DASS-21 administration		
	N	Age, years	Gender, female (%)	Target population	Setting	Country	Language
Antony et al. [3]	307	28.4–44.9	43–78	Including patients with psychiatric disorders	Clinic and non-clinic	Canada	English
Henry & Crawford [5]	1794	41.0 ± 15.9	54.57	Healthy general	Non-clinic	UK	English
Ng et al. [6]	388	52.3 ± 17.5	61.6	Including patients with psychiatric disorders	Clinic	Australia	English
Shea et al. [7]	420	Reported as age category and percent-ages	66.2	Healthy general	Non-clinic	Australia	English
Sinclair et al. [8]	503	44.7 ± 16.3	52	Healthy general	Non-clinic	USA	English
Apóstolo et al. [9]	1297	48.57 ± 19.9	66.7	Including patients with chronic disease	Clinic	Portugal	Portuguese
Nanthakumar et al. [10]	570 OSA: 285 Non-OSA: 285	OSA: 58.0 ± 5.2 Non-OSA: 58.0 ± 5.2	38.25 38.25	Including patients with chronic disease	Clinic and non-clinic	Australia	English
Daza et al. [22]	98	40.0 ± 13.3	79.5	Including patients with psychiatric disorders	Clinic	USA	Spanish
Norton [23]	895	20.6 ± 4.1	70.3	Healthy general	Non-clinic	USA	English
Musa et al. [24]	263	18–55	62.0	Healthy general	Clinic	Malaysia	Malay
Asghari et al. [25]	551	27.5 ± 7.1	43.4	Including patients with psychiatric disorders	Non-clinic	Iran	Persian
Ownsworth et al. [26]	77	35.3–51.5	44–74	Including patients with chronic disease	Clinic and non-clinic	Australia	English
Gloster et al. [27]	222	67.5 ± 6.0	73.8	Including patients with psychiatric disorders	Non-clinic	USA	English
Ramli et al. [28]	153	20–60	51	Including patients with chronic disease	Clinic	Malaysia	Malaysian
Wood et al. [29]	2045	Reported as age category and percent-ages	56.5	Including patients with chronic disease	Clinic	Australia	English
Osman et al. [30]	1297 Study 1: 887 Study 2: 410	Study 1: 19.4 ± 2.1 Study 2: 19.6 ± 2.8	59.1	Healthy general	Non-clinic	USA	English
Tran et al. [31]	221	27.7 ± 5.5	100	Healthy general	Non-clinic	Vietnam	Vietnamese
Vasconcelos-Raposo et al. [32]	1020	36.7 ± 11.9	57.4	Healthy general	Non-clinic	Portugal	Portuguese
Dahm et al. [33]	123	37.9 ± 14.9	21.1	Including patients with chronic disease	Clinic	Australia	English

Table 1 (continued)

Authors	Population characteristics			Target population	Characteristics of DASS-21 administration		
	N	Age, years	Gender, female (%)		Setting	Country	Language
Oei et al. [34]	2613	30.46	53	Healthy general	Non-clinic	Malaysia, Indonesia, Singapore, Sri Lanka, Taiwan, Thailand	Malaysia, Indonesian, English, Sinhala, Tamil, Mandarin Chinese, Thai
Tonsing [35]	212	32.4 ± 9.0	51.9	Healthy general	Non-clinic	Hong Kong	Nepali
Musa et al. [36]	246	32 (median)	50	Including patients with chronic disease	Clinic	Malaysia	Malay
Vignola & Tucci [37]	242	39.9 ± 12.4	62	Including patients with chronic disease	Clinic	Brazil	Brazilian Portuguese
Gomez et al. [38]	687	43.7 ± 22.4	67	Healthy general	Non-clinic	Australia	English
Bottesi et al. [39]	616	36.3 ± 13.7	48–78.2	Including patients with psychiatric disorders	Clinic and non-clinic	Italy	Italian
Wang et al. [40]	2071 Study 1: Students = 1815 Study 2: Schizophrenia = 166 Healthy controls = 90	Study 1: 18.7 ± 0.8 Study 2: Schizophrenia = 42.7 ± 9.6 Healthy controls = 42.7 ± 9.7	60.74	Including patients with psychiatric disorders	Clinic and non-clinic	China	Chinese
Randall et al. [41]	504	33.1 ± 14.0	19.6	Including patients with chronic disease	Clinic	Australia	English
Moore et al. [42]	2454	Adolescents	47.6	Healthy general	Non-clinic	USA	English
Alfonsson et al. [43]	624 Study 1: Students = 276 Study 2: Primary care = 192 Study 3: Psychotherapy = 156	Study 1: 24.5 ± 6.3 Study 2: 45.3 ± 17.7 Study 3: 35.4 ± 12.2	70.8	Including patients with chronic and psychiatric disorders	Clinic and non-clinic	Sweden	Swedish
Yusoff [44]	706	18–21	60.9	Healthy general	Non-clinic	Malaysia	Malay
Apóstolo et al. [45]	101	45.4 ± 12.5	63.37	Including patients with psychiatric disorders (attending psychiatric consultations)	Clinic	Portugal	Portuguese
Camacho et al. [46]	505	23.9 ± 6.2	73.7	Healthy general	Non-clinic	USA	English
Gomez et al. [47]	269	66.4 ± 7.0	81.78	Healthy general	Non-clinic	Australia	English
Jafari et al. [48]	477	College students	50.10	Healthy general	Non-clinic	Iran	Persian
Johnson et al. [49]	251	66.1 ± 9.7	32.27	Including patients with chronic disease	Non-clinic	Australia	English

Table 1 (continued)

Authors	Population characteristics				Characteristics of DASS-21 administration		
	N	Age, years	Gender, female (%)	Target population	Setting	Country	Language
Johnson et al. [50]	364	18 or older	56.9	Including patients with chronic disease	Clinic	Australia	English
Mahmoud et al. [51]	457	20.0 ± 1.6	66	Healthy general	Non-clinic	USA	English
Parkitny et al. [52]	154	53.6 ± 12.8	60.40	Including patients with chronic disease	Clinic	Australia	English
Wong et al. [53]	123	37.9 ± 14.9	22.8	Including patients with chronic disease	Clinic	Australia	English
Le et al. [54]	1616	16.5 ± 1.0	46.30	Healthy general	Non-clinic	Vietnam	Vietnamese
Rusli et al. [55]	402	18–83	0	Including patients with chronic disease	Clinic	Malaysia	Malay

DASS-21 Depression Anxiety Stress Scale, *OSA* obstructive sleep apnea

the DASS-21 Anxiety subscale [10, 27, 29]. Under the bifactor structure, Cronbach's alpha [5, 32, 39, 41, 43, 54] and coefficient omega [30, 42] for the DASS-21 subscales and the total scale were all > 0.70.

The Cronbach's alpha values for the three-factor structure with at least adequate methodological quality [3, 5, 8–10, 22, 23, 25, 29, 40, 43, 48] were qualitatively summarized. Two studies [7, 27] were excluded from the summary because it evaluated uncorrelated errors (rho), or Pearson Separation Index (PSI) as a statistical value. The qualitatively summarized coefficient alpha values for the three-factor DASS-21 Depression, Anxiety, and Stress structure were 0.83–0.94, 0.66–0.87, and 0.79–0.91, respectively; and the overall rating had sufficient moderate-quality evidence. Cronbach's alpha values for the bifactor structure were qualitatively summarized [5, 32, 39, 41, 43, 54], and two studies [30, 42] were summarized separately because coefficient omega values were used.

The qualitatively summarized coefficient alpha values for the three-factor DASS-21 Depression, Anxiety, and Stress structure were 0.83–0.94, 0.66–0.87, and 0.79–0.91, respectively; and the overall rating had sufficient moderate-quality evidence. Cronbach's alpha values for the bifactor DASS-21 structure were 0.90–0.95 (total scale), 0.82–0.92 (Depression), 0.74–0.88 (Anxiety), and 0.76–0.90 (Stress); the corresponding qualitatively summarized coefficient omega values (two studies) were 0.89–0.97, 0.86–0.99, 0.82–0.99, and 0.85–0.99, respectively. The overall rating was sufficient and of high quality for the internal consistency under the bifactor structure.

Cross-cultural validity/measurement invariance

Six studies assessed cross-cultural validity/measurement invariance [10, 23, 34, 38, 48, 54]. Five of these studies had assessed the cross-cultural validity/measurement invariance based on a three-factor structure, and the sixth study [54] demonstrated a bifactor structure. The quality ratings of the five studies were inconsistent and no explanation was found, and so subgroups by gender, race, country (language), and disease status were explored in an attempt to understand the inconsistency. Subgroup analysis by gender [38, 48] yielded inconsistent moderate-quality evidence regarding measurement invariance. The other three subgroups included only a single study: race [23], country (language) [34], and disease status [10].

Reliability

Reliability was reported for five studies [25, 26, 39, 40, 47]. Only one study [25] evaluated the intraclass correlation coefficient (ICC) for reliability, while the results were insufficient for the remaining studies. The insufficient results

Table 2 Summary of findings

Measurement property	Summary or pooled results	Overall rating	Quality of evidence
Content validity	Content validity [2, 22, 24, 25, 35, 37, 39, 43, 45]	Sufficient (+)	Moderate
	• Relevance	Sufficient (+)	Very low
	• Comprehensiveness	Sufficient (+)	High
	• Comprehensibility	Sufficient (+)	Moderate
Structural validity	• Three-factor structure [3, 5, 7–10, 23, 25, 27, 29, 38, 40, 41, 47, 48]	Sufficient (+)	Moderate: multiple very good studies, inconsistent results
	Qualitative summary: 80% supported [3, 5, 8–10, 23, 27, 29, 38, 40, 41, 47]		
	Total sample size: 10,588		
	• Bifactor structure [5, 30, 32, 39, 41–43, 54]	Sufficient (+)	High: multiple very good studies
	Qualitative summary: all studies supported (100%)		
	Total sample size: 8672		
Internal consistency	• Three-factor structure	Sufficient (+)	Moderate: multiple very good studies, inconsistent results
	Cronbach's alpha values [3, 5, 8–10, 22, 23, 25, 29, 40, 48]		
	DASS-21 depression: 0.83–0.94		
	DASS-21 anxiety: 0.66–0.87 (≥ 0.70 in 80%)		
	DASS-21 stress: 0.79–0.91		
	Total sample size: 9097		
	• Bifactor structure	Sufficient (+)	High: multiple very good studies
	Coefficient omega values [30, 42]		
	DASS-21 total: 0.89–0.97		
	DASS-21 depression: 0.86–0.99		
	DASS-21 anxiety: 0.82–0.99		
	DASS-21 stress: 0.85–0.99		
	Total sample size: 3751		
	Cronbach's alpha values [5, 32, 39, 41, 43, 54]		
	DASS-21 total: 0.90–0.95		
	DASS-21 depression: 0.82–0.92		
	DASS-21 anxiety 0.74–0.88		
	DASS-21 stress 0.76–0.90		
	Total sample size: 5670		
Cross-cultural validity/measurement invariance	• Three-factor structure	Inconsistent (±)	Moderate: multiple very good studies, inconsistency results
	Qualitative summary of studies across gender [38, 48]: 1 of 2 hypotheses supported		
	Total sample size: 1164		
	• Bifactor structure	Not summarized	Not graded
	Qualitative summary of a single study across gender [54]		
	Total sample size: 1616		

Table 2 (continued)

Measurement property	Summary or pooled results	Overall rating	Quality of evidence
Reliability	• Subgroup explanation of studies with time interval of around 2 weeks		Low for all DASS-21 depression, anxiety, and stress: serious risk of bias (only one study of adequate quality), serious inconsistency
	Pearson's correlation coefficients of studies [26, 39]		
	DASS-21 depression: 0.75–0.78	DASS-21 depression: insufficient (−)	
	DASS-21 anxiety: 0.64–0.73	DASS-21 anxiety: insufficient (−)	
	DASS-21 stress: 0.64–0.65	DASS-21 stress: insufficient (−)	
	Total sample size: 206		
Measurement error	No information available		
Criterion validity	• SCID as the gold-standard criterion [27, 31, 33]		
	DASS-21 depression and SCID Depression: AUC = 0.77–0.91	DASS-21 depression: sufficient (+)	High: multiple very good studies
	DASS-21 anxiety and SCID anxiety: AUC = 0.60–0.83	DASS-21 anxiety: insufficient (−)	Moderate: multiple very good studies, inconsistent results
	Total sample size: 566		
Hypothesis testing for construct validity	• Quantitative pooling for convergent validity (pooled correlation coefficients):	DASS-21 depression: sufficient (+)	High: multiple very good studies
	- DASS-21 depression and BDI [3, 22, 25, 27, 39]: 0.73		
	- DASS-21 depression and HADS depression [26, 33, 45]: 0.69		
	- DASS-21 depression and PANAS negative affect [5, 9, 27]: 0.56		
	Subtotal sample size: 5209		
	- DASS-21 anxiety and BAI [3, 22, 27, 39]: 0.75	DASS-21 anxiety: sufficient (+)	High: multiple very good studies
	- DASS-21 anxiety and HADS anxiety [26, 33, 45]: 0.66		
	- DASS-21 anxiety and PANAS negative affect [5, 9, 27]: 0.55		
	Subtotal sample size: 4658		
	- DASS-21 stress and PANAS negative affect [5, 9, 27]: 0.66	DASS-21 stress: sufficient (+)	High: multiple very good studies
	Subtotal sample size: 3313		
	• Qualitative summary for known-groups validity [3, 6, 25, 27, 40]:		High for all DASS-21 depression, anxiety, and stress: multiple very good studies
	- DASS-21 depression: 5 of 5 supported (100%)	DASS-21 depression: sufficient (+)	
	- DASS-21 anxiety: 5 of 5 supported (100%)	DASS-21 anxiety: sufficient (+)	
	- DASS-21 stress: 4 of 5 supported (80%)	DASS-21 stress: sufficient (+)	
	Subtotal sample size: 1314 or higher (not clearly reported)		

Table 2 (continued)

Measurement property	Summary or pooled results	Overall rating	Quality of evidence
Responsiveness	- DASS-21 depression [6, 26]: 2 of 2 hypotheses supported	Depression: sufficient (+)	Low: multiple studies of inadequate quality
	- DASS-21 anxiety [6, 26]: 1 of 2 hypotheses supported	Anxiety: inconsistent (±)	Very low: multiple studies of inadequate quality, with inconsistency
	- DASS-21 stress [6, 26]: 2 of 2 hypotheses supported	Stress: sufficient (+)	Low: multiple studies of inadequate quality
	Total sample: 360 (depression)/351 (anxiety)/370 (Stress)		

AUC area under the receiver operating characteristic curve, *BAI* Beck Anxiety Inventory, *BDI* Beck Depression Inventory, *HADS* Hospital Anxiety and Depression Scale, *PANAS* Positive and Negative Affect Schedule, *SCID* Structured Clinical Interview for DSM-IV Axis 1 Diagnoses

might have been due to a problem with the research method involving long time intervals between the first and second administrations of the DASS-21. Therefore, three studies [25, 26, 39] were qualitatively summarized after eliminating two studies with intervals of 3–6 months [40, 47]. Pearson's correlation coefficients for the two studies were 0.75–0.78 (Depression), 0.64–0.73 (Anxiety), and 0.64–0.65 (Stress). The overall ratings for the Depression, Anxiety, and Stress subscales were insufficient low quality because of a serious risk of bias and serious inconsistency.

Criterion validity

Criterion validity was reported for three studies [27, 31, 33]. The psychiatrist-administered Structured Clinical Interview for DSM-IV Axis 1 Diagnoses (SCID) for depression and anxiety was utilized as the gold-standard criterion for the DASS-21. The DASS-21 Depression and Anxiety subscales demonstrated areas under the receiver operating characteristic curves (AUCs) of 0.77–0.91 for SCID Depression and 0.60–0.83 for SCID Anxiety. Therefore, high-quality evidence of sufficient criterion validity was exhibited for the DASS-21 Depression subscale, and moderate-quality evidence of insufficient criterion validity was exhibited for the DASS-21 Anxiety subscale.

Hypotheses testing for construct validity

Quantitative pooling was applied to the correlations of the DASS-21 Depression subscale with the BDI [3, 22, 25, 27, 39], the HADS Depression subscale [26, 33, 45], and the PANAS Negative Affect subscale [5, 9, 27]; of the DASS-21 Anxiety subscale with the BAI [3, 22, 27, 39], the HADS Anxiety subscale [26, 33, 45], and PANAS Negative Affect subscale [5, 9, 27]; and of the DASS-21 Stress subscale with the PANAS Negative Affect subscale [5, 9, 27] (Table 3; Supplement 1 contains forest plots). Construct validity was supported by high pooled coefficients for the correlations of the DASS-21 Depression with the BDI ($r = 0.73$), the HADS Depression subscale ($r = 0.69$), and the PANAS Negative Affect subscale ($r = 0.56$). The DASS-21 Anxiety subscale demonstrated high pooled coefficients for the correlations with the BAI ($r = 0.75$), the HADS Anxiety subscale ($r = 0.66$), and PANAS Negative Affect subscale ($r = 0.55$). DASS-21 Stress was also strongly correlated with the PANAS Negative Affect subscale ($r = 0.66$). Based on these findings, the overall construct validity was rated as sufficient and the quality of evidence as high in the hypotheses testing.

Five studies [3, 6, 25, 27, 40] evaluated known-groups validity. All known-groups comparisons were conducted

Table 3 Pooled correlation coefficients for construct validity (convergent validity)

DASS-21 subscale and comparators	Estimated correlation coefficient	95% CI		Higgin's I^2 (%)
DASS-21 Depression and BDI [3, 22, 25, 27, 39]	0.73	0.62	0.84	95.80
DASS-21 Depression and HADS Depression [26, 33, 45]	0.69	0.60	0.78	55.77
DASS-21 Depression and PANAS Negative Affect [5, 9, 27]	0.56	0.50	0.63	80.90
DASS-21 Anxiety and BAI [3, 22, 27, 39]	0.75	0.64	0.86	94.31
DASS-21 Anxiety and HADS Anxiety [26, 33, 45]	0.66	0.50	0.83	86.17
DASS-21 Anxiety and PANAS Negative Affect [5, 9, 27]	0.55	0.51	0.60	65.66
DASS-21 Stress and PANAS Negative Affect [5, 9, 27]	0.66	0.58	0.73	91.44

CI confidence interval

while including patients with a psychiatric diagnosis. All results (five out of five) regarding DASS-21 Depression and Anxiety, and 80% of the results (four out of five) regarding DASS-21 Stress were in accordance with the hypotheses supporting known-groups validity. The overall ratings of known-groups validity for the DASS-21 were sufficient high quality.

Responsiveness

Two studies [6, 26] analyzed responsiveness, comparing the DASS-21 scores of patients at admission/predischarge and at discharge. Both demonstrated significant changes in the DASS-21 Depression and Stress scores at discharge, and the results were in accordance with the hypotheses for the Depression and Stress subscales (sufficient rating of low quality because there is a serious risk of bias when using paired t-tests to analyze responsiveness, making this an inappropriate method for evaluating responsiveness). The direction of the Depression and Stress change in two studies [6, 26] was opposite: decreased in psychiatric patients [6] whereas increased among patients with traumatic brain injury [26]. The DASS-21 Anxiety subscale exhibited an inconsistent rating of very low quality because of multiple inadequate studies with inconsistent results which had utilized the paired t-test.

Discussion

This systematic review evaluated 48 studies of the measurement properties of the DASS-21 as reported in 41 articles. Content validity refers to whether the content of an instrument appropriately reflects the construct to be measured, which is the most important measurement property of an instrument [13]. For example, Ailliet et al. [56] noted that the content validity of the Neck Disability Index is poor due to it missing important content, and so they advocated developing a new instrument. With regards to the content validity, the DASS-21 demonstrated sufficient evidence for relevance, comprehensiveness, and comprehensibility. The quality of evidence was high for comprehensiveness, moderate for comprehensibility, and very low for relevance. The presence of sufficient high-quality evidence for comprehensiveness suggests that the DASS-21 includes key concepts. Comprehensibility refers to whether the PROM instructions, items, and response options were understood by the population of interest as intended and also to the wording of the items and whether the response options matched the questions. The lack of qualitative methods for assessing the comprehensibility of the DASS-21 resulted in sufficient moderate-quality evidence. Relevance refers to the relevance of items for the construct, target population, and context of

use of interest, response options, and recall period, and these aspects were not evaluated either by experts or patients in any of the content validity studies of the DASS-21. Further studies are therefore strongly recommended to evaluate the content validity of the DASS-21, especially its relevance.

Most debate regarding the psychometric properties of the DASS-21 has revolved around its underlying structure. The DASS-21 was originally demonstrated with the three factors of its Depression, Anxiety, and Stress subscales; however, alternative structures have been explored due to substantial interfactor correlations ranging from moderate to strong [41]. When interfactor correlations are $r > 0.4$, a bifactor model in which items load on both a general (unidimensional) factor and group factors (potential subscales) may be viable [57]. The existence of a common factor was also supported in the DASS developmental process [2]. The second-order CFA identified a common factor that accounted for 83, 75, and 84% of the variance in the Depression, Anxiety, and Stress subscales. Consistent with this, the best structure derived in the present systematic review was a bifactor structure that exhibited a sufficient high quality of evidence. That is, the DASS-21 items load on a general factor named as a Negative Emotional state (accounting for the common variance among all 21 items) as well as orthogonal group factors named as Depression, Anxiety, and Stress subscales (explaining the item covariance that is independent of the covariance due to the general factor). Osman et al. [30] reported that the item variance of the DASS-21 was explained more by the general factor (62%) than by any of the group factors. These findings have the practical implications that both the total and subscale scores should be calculated separately and considered independently with weightings relative to the total score. The DASS-21 has the merit of providing general information about the negative emotional status of patients as well as each emotional symptom of depression, anxiety, and stress. Establishing cut-points would improve practicality of using the DASS-21.

According to the COSMIN Risk of Bias checklist [11], evidence for structural validity is a prerequisite for the internal consistency and cross-cultural validity/measurement invariance, and these measurement properties focus on relationships between the items constituting an instrument. The present study found that the bifactor structure was optimal for the DASS-21 since this was associated with sufficient high-quality evidence for internal consistency. However, evidence for the bifactor structure measurement invariance could not be assessed due to the availability of only a single study [54]. It is therefore recommended that future studies evaluate the bifactor structure invariance according to gender or language.

Evidence on the reliability of the DASS-21 has been summarized based on studies that tested its reliability using Pearson's correlation coefficients, because only one study

utilized the ICC when analyzing its reliability. Studies utilizing Pearson correlation analysis have produced inconsistent results, even those involving subgroups separated by a time interval of around 2 weeks. According to the COSMIN manual, the measurement quality of the reliability should be rated as doubtful when evaluated by the correlation between two measurements without evidence that no systematic change has occurred or with evidence that a systematic change has occurred. The DASS-21 measures states fluctuating over time or situations rather than traits, and so reliability might not be an important property. The authors decided not to downgrade the measurement quality of each DASS-21 study in relation to the evidence regarding systematic changes between measurements. Downgrading the methodological quality depends on the context of the measurements, and exceptional occasions need to be considered because emotion is a relatively versatile context that can result in systematic changes even in the absence of an apparent cause.

Criterion validity has been defined as "the degree to which the scores of a patient-reported outcome measure are an adequate reflection of a gold standard" [58]. Even though the original version of a shortened instrument is considered as gold standard for a self-reported instrument [59], others have insisted that an expert clinical opinion can be used as a gold standard [60]. A psychiatrist-administered SCID for depression and anxiety was considered as the gold standard in the present study.

Quantitative pooling was conducted for evaluating hypothesis testing (convergent validity). High heterogeneity existed even with a random-effects model. Because correlation coefficients > 0.50 are considered to indicate moderate correlations, wide ranges of the coefficient values might have contributed to the high heterogeneity.

Two studies that evaluated the responsiveness of the DASS-21 used paired t-tests as the statistical analysis technique. According to de Vet et al. [59], the paired t-test is related to the statistical significance of changes in scores rather than their validity. The paired t-test is not recommended as a responsiveness parameter. The context of the response also needs to be considered in a qualitative summary of results. For example, two studies included in the current review measured the DASS-21 scores of patients at admission and discharge; that is, after treatment relative to at admission to the hospital. At discharge, patients with psychiatric disorders exhibited improvements in negative emotional status, whereas patients with brain injuries faced new challenges of returning home with some disability. Researchers need to be careful about the direction of changes in order to avoid results categorized as "inconsistent."

Psychometrically, the DASS-21 exhibited sufficient high-quality evidence for bifactorial structural validity, internal consistency under the bifactor structure,

criterion validity (especially for the depression subscale), and hypothesis testing for construct validity. The synthesized evidence of psychometric properties of the DASS-21 is comparable to that of well-known measures of emotional symptoms such as the CES-D, CESD-R, HADS, and PHQ-9 (which demonstrated strong positive evidence in the set of psychometric properties) when also evaluated with the original COSMIN methodology [61, 62]. Because the current review was based on updated COSMIN methodology, sufficient high-quality evidence (the highest rating) was compared to strong positive evidence (the highest rating) in the previous COSMIN methodology. The CES-D demonstrated strong positive evidence for structural, internal, and construct validity when applied to patients with diabetes [61]. The HADS demonstrated strong positive evidence for structural and internal validity, and moderate positive evidence for construct validity among patients with diabetes. There was conflicting evidence for the structural validity of the PHQ-9, which affects the results regarding internal consistency among patients with diabetes. The CESD-R demonstrated strong positive evidence for structural and internal validity and moderate positive evidence for construct validity among the general public [62].

The wide applicability of the DASS-21 is one of its strengths. The DASS-21 has been validated in healthy general populations as well as patient populations (both psychiatric disease and chronic disease patients). The DASS-21 has been applied to a wide range of populations in terms of age (for subjects older than 14 years). The DASS-21 provides helpful information regarding the negative emotional status of subjects. Unlike the HADS that has established cut-offs for suggesting the presence of clinically meaningful anxiety and/or depression, cut-offs have not yet been established for the DASS-21. Further studies of DASS-21 cut-offs would therefore strengthen the usability of this instrument as a screening tool. One limitation would be using the DASS-21 as an outcome measure because further validation studies regarding its responsiveness are required. Applying the DASS-21 to people younger than 14 years also requires further validation studies.

This study applied the recently updated COSMIN methodology to perform a systematic review of the DASS-21. Having structural validity as an anchor for evaluating internal validity and measurement invariance enabled meaningful evaluation of the structure related to psychometric properties. The updated COSMIN methodology requires authors performing reviews to be knowledgeable about the context of PROMs and related valid measurement instruments. The authors are required to set hypotheses to be tested of different types and magnitudes. Discretion is required regarding each measurement property because some studies provide results of psychometric evaluations performed

using different properties (e.g., criterion validity rather than hypothesis testing).

Conclusions

The DASS-21 exhibited sufficient high-quality evidence for bifactor structural validity, internal consistency under the bifactor structure, criterion validity, and construct validity. The psychometric quality of the DASS-21 is comparable to that of other well-known related measures evaluated using the original COSMIN methodology. The psychometric robustness and wide applicability of the DASS-21 suggest that this scale can be used to understand negative status emotions including depression, anxiety, and stress in both healthy general populations and patient populations. Establishing cut-off points would improve the practicality of applying the DASS-21. The use of the DASS-21 as an outcome measure requires further validation studies regarding responsiveness. The DASS-21 subscales as well as its total score need to be scored and interpreted as individual emotional symptoms of depression, anxiety, and stress as well as overall negative emotions. Further studies are required into its measurement invariance reflecting a bifactor structure, reliability, measurement error, and responsiveness. The updated COSMIN manual provides detailed guidelines for facilitating systematic reviews of PROMs."

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interests.

Informed consent For this type of study informed consent is not required.

Research involving human and animal participants This article does not contain any studies involving human participants or animals performed by any of the authors.

References

1. World Health Organization. (2016). *Investing in treatment for depression and anxiety leads to fourfold return*. Retrieved from October 22, 2018 http://www.who.int/mediacentre/news/releases/2016/depression-anxiety-treatment/en/.

2. Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales* (2nd ed.). Sydney: Psychology Foundation.

3. Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M., & 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.

4. Lovibond, S. H., & Lovibond, P. F. (1995). The structure of negative emotional states: Comparison of the depression anxiety stress scales (DASS) with the Beck depression and anxiety inventories. *Behaviour Research and Therapy, 33*(3), 335–343.

5. 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 nonclinical sample. *British Journal of Clinical Psychology, 44*(2), 227–239.

6. Ng, F., Trauer, T., Dodd, S., Callaly, T., Campbell, S., & Berk, M. (2007). The validity of the 21-item version of the Depression Anxiety Stress Scales as a routine clinical outcome measure. *Acta Neuropsychiatrica, 19*(5), 304–310.

7. Shea, T. L., Tennant, A., & Pallant, J. F. (2009). Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). *BMC Psychiatry, 9,* 21.

8. Sinclair, S. J., Siefert, C. J., Slavin-Mulford, J. M., Stein, M. B., Renna, M., & Blais, M. A. (2012). Psychometric evaluation and normative data for the Depression, Anxiety, and Stress Scales-21 (DASS-21) in a nonclinical sample of U.S. adults. *Evaluation & Health Professions, 35*(3), 259–279.

9. Apóstolo, J. L. S., Tanner, B. A., & Arfken, C. L. (2012). Confirmatory factor analysis of the Portuguese Depression Anxiety Stress Scales − 21. *Revista Latino-Americana de Enfermagem, 20*(3), 590–596.

10. Nanthakumar, S., Bucks, R. S., Skinner, T. C., Starkstein, S., Hillman, D., James, A., et al. (2017). Assessment of the Depression Anxiety and Stress Scale (DASS-21) in untreated obstructive sleep apnea (OSA). *Psychological Assessment, 29*(10), 1201–1209.

11. Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., et al. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome measures. *Quality of Life Research, 27*(5), 1171–1179.

12. Prinsen, C. A. C., Mokkink, L. B., Bouter, N. M., Alonso, J., Patric, D. L., de Vet, H. C. W., et al. (2018). COSMIN guideline for systematic reviews of Patient-Reported Outcome measures. *Quality of Life Research, 27*(5), 1147–1157.

13. Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., et al. (2018). COSMIN methodology for evaluating the content validity of Patient-Reported Outcome Measures: a Delphi study. *Quality of Life Research, 27*(5), 1159–1170.

14. Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research, 18*(8), 1115–1123.

15. Psychology Foundation of Australia. (2017). DASS FAQ (Frequently Asked Questions) 10. Can the DASS be used with children/adolescents? Retrieved from October 22, 2018 http://www2.psy.unsw.edu.au/dass/DASSFAQ.htm#_10.__Can_the_DASS_be_used_with_chil.

16. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42.

17. Beck, A. T., & Steer, R. A. (1990). *Manual for the beck anxiety inventory*. San Antonio, TX: Psychological Cooperation.

18. Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory manual*. San Antonio, TX: Psychological Cooperation.

19. Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica, 67*(6), 361–370.

20. Watson, D., Clark, L. A., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology, 97*(3), 346–353.

21. R Development Core Team. (2010). *R: A language and environment for statistical computing. Vienna, Austria: R foundation for Statistical Computing*. Retrieved from October 22, 2018 http://www.R-project.org.

22. Daza, P., Novy, D. M., Stanley, M. A., & Averill, P. (2002). The depression anxiety stress scale-21: Spanish translation and validation with a Hispanic sample. *Journal of Psychopathology and Behavioral Assessment, 24*(3), 195–205.

23. Norton, P. J. (2007). Depression Anxiety and Stress Scales (DASS-21): Psychometric analysis across four racial groups. *Anxiety Stress and Coping, 20*(3), 253–265.

24. Musa, R., Fadzil, M. A., & Zain, Z. (2007). Translation, validation and psychometric properties of Bahasa Malaysia version of the Depression Anxiety and Stress Scales (DASS). *ASEAN Journal of Psychiatry, 8*(2), 82–89.

25. Asghari, A., Saed, F., & Dibajnia, P. (2008). Psychometric properties of the Depression Anxiety Stress Scales-21 (DASS-21) in a non-clinical Iranian sample. *International Journal of Psychology, 2*(2), 82–102.

26. 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.

27. Gloster, A. T., Rhoades, H. M., Novy, D., Klotsche, J., Senior, A., Kunik, M., et al. (2008). Psychometric properties of the Depression Anxiety and Stress Scale-21 in older primary care patients. *Journal of Affective Disorders, 110*(3), 248–259.

28. Ramli, M., Salmiah, M. A., & Nurul Ain, M. (2009). Validation and psychometric properties of Bahasa Malaysia version of the Depression Anxiety and Stress Scales (DASS) among diabetic patients. *Malaysian Journal of Psychiatry, 18*(2), 1–7.

29. Wood, B. M., Nicholas, M. K., Blyth, F., Asghari, A., & Gibson, S. (2010). The utility of the short version of the Depression Anxiety Stress Scales (DASS-21) in elderly patients with persistent pain: Does age make a difference? *Pain Medicine, 11*(12), 1780–1790.

30. 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 examination of dimensions, scale reliability, and correlates. *Journal of Clinical Psychology, 68*(12), 1322–1338.

31. 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*(1), 24.

32. Vasconcelos-Raposo, J., Fernandes, H. M., & Teixeira, C. M. (2013). Factor structure and reliability of the depression, anxiety and stress scales in a large Portuguese community sample. *The Spanish Journal of Psychology, 16*, E10.

33. Dahm, J., Wong, D., & Ponsford, J. (2013). Validity of the Depression Anxiety Stress Scales in assessing depression and anxiety following traumatic brain injury. *Journal of Affective Disorders, 151*(1), 392–396.

34. Oei, T. P., Sawang, S., Goh, Y. W., & Mukhtar, F. (2013). Using the depression anxiety stress scale 21 (DASS-21) across cultures. *International Journal of Psychology, 48*(6), 1018–1029.

35. Tonsing, K. N. (2014). Psychometric properties and validation of Nepali version of the Depression Anxiety Stress Scales (DASS-21). *Asian Journal of Psychiatry, 8,* 63–66.

36. Musa, R., Ramli, R., Abdullah, K., & Sarkarsi, R. (2011). Concurrent validity of the depression and anxiety components in the Bahasa Malaysia version of the depression and anxiety and stress scale (DASS). *ASEAN Journal of Psychiatry, 12*(1), 66–70.

37. Vignola, R. C. B., & Tucci, A. M. (2014). Adaptation and validation of the depression, anxiety and stress scale (DASS) to Brazilian Portuguese. *Journal of Affective Disorders, 155,* 104–109.

38. Gomez, R., Summers, M., Summers, A., Wolf, A., & Summers, J. (2014). Depression Anxiety Stress Scales-21: Measurement and structural invariance across ratings of men and women. *Assessment, 21*(4), 418–426.

39. Bottesi, G., Ghisi, M., Altoè, G., Conforti, E., Melli, G., & Sica, C. (2015). The Italian version of the Depression Anxiety Stress Scales-21: Factor structure and psychometric properties on community and clinical samples. *Comprehensive Psychiatry, 60,* 170–181.

40. Wang, K., Shi, H. S., Geng, F. L., Zou, L. Q., Tan, S. P., Wang, Y., et al. (2016). Cross-cultural validation of the Depression Anxiety Stress Scale–21 in China. *Psychological Assessment, 28*(5), e88.

41. Randall, D., Thomas, M., Whiting, D., & McGrath, A. (2017). Depression Anxiety Stress Scales (DASS-21): factor structure in traumatic brain injury rehabilitation. *The Journal of Head Trauma Rehabilitation, 32*(2), 134–144.

42. Moore, S. A., Dowdy, E., & Furlong, M. J. (2016). Using the Depression, Anxiety, Stress Scales–21 With US Adolescents: An Alternate Models Analysis. *Journal of Psychoeducational Assessment, 35*(6), 581–598.

43. Alfonsson, S., Wallin, E., & Maathz, P. (2017). Factor structure and validity of the Depression, Anxiety and Stress Scale-21 in Swedish translation. *Journal of Psychiatric and Mental Health Nursing, 24*(2–3), 154–162.

44. Yusoff, M. S. B. (2013). Psychometric properties of the depression anxiety stress scale in a sample of medical degree applicants. *International Medical Journal, 20*(3), 295–300.

45. Apóstolo, J. L. A., Mendes, A. C., & Azeredo, Z. A. (2006). Adaptation to Portuguese of the depression, anxiety and stress scales (DASS). *Revista Latino-Americana de Enfermagem, 14*(6), 863–871.

46. Camacho, A., Cordero, E. D., & Perkins, T. (2016). Psychometric Properties of the DASS-21 Among Latina/o College Students by the US-Mexico Border. *Journal of Immigrant and Minority Health, 18*(5), 1017–1023.

47. Gomez, R., Summers, M., Summers, A., Wolf, A., & Summers, J. J. (2014). Depression Anxiety Stress Scales-21: factor structure and test-retest invariance, and temporal stability and uniqueness of latent factors in older adults. *Journal of Psychopathology and Behavioral Assessment, 36*(2), 308–317.

48. Jafari, P., Nozari, F., Ahrari, F., & Bagheri, Z. (2017). Measurement invariance of the Depression Anxiety Stress Scales-21 across medical student genders. *International Journal of Medical Education, 8,* 116–122.

49. Johnson, A. R., Lawrence, B. J., Corti, E. J., Booth, L., Gasson, N., Thomas, M. G., et al. (2016). Suitability of the Depression, Anxiety, and Stress Scale in Parkinson's disease. *Journal of Parkinson's Disease, 6*(3), 609–616.

50. Johnson, C. E., Bennett, K. S., Newton, J., McTigue, J., Taylor, S., Musiello, T., et al. (2018). A pilot study to assess the validity of the DASS-21 subscales in an outpatient oncology population. *Psycho-oncology, 27*(2), 695–699.

51. Mahmoud, J. S. R., Hall, L. A., & Staten, R. (2010). The psychometric properties of the 21-item Depression, Anxiety, and Stress Scale (DASS-21) among a sample of young adults. *Southern Online Journal of Nursing Research, 10*(4), 21–34.

52. Parkitny, L., McAuley, J. H., Walton, D., Costa, L. O. P., Refshauge, K. M., Wand, B. M., et al. (2012). Rasch analysis supports the use of the depression, anxiety, and stress scales to measure mood in groups but not in individuals with chronic low back pain. *Journal of Clinical Epidemiology, 65*(2), 189–198.

53. Wong, D., Dahm, J., & Ponsford, J. (2013). Factor structure of the depression anxiety stress scales in individuals with traumatic brain injury. *Brain Injury, 27*(12), 1377–1382.

54. Le, M. T. H., Tran, T. D., Holton, S., Nguyen, H. T., Wolfe, R., & Fisher, J. (2017). Reliability, convergent validity and factor structure of the DASS-21 in a sample of Vietnamese adolescents. *PLoS ONE, 12*(7), e0180557.

55. Rusli, B. N., Amrina, K., Trived, S., Loh, K. P., & Shashi, M. (2017). Construct validity and internal consistency reliability of the Malay version of the 21-item depression anxiety stress scale (Malay-DASS-21) among male outpatient clinic attendees in Johor. *The Medical Journal of Malaysia, 72*(5), 264–270.

56. Ailliet, L., Knol, D. L., Rubinstein, S. M., de Vet, H. C. W., van Tulder, M. W., & Terwee, C. B. (2013). Definition of the construct to be measured is a prerequisite for the assessment of validity. The Neck Disability Index as an example. *Journal of Clinical Epidemiology, 66*(7), 775–782.

57. 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*(Suppl 1), 19–31.

58. Mokkink, L. B., Terwee, C. B., Patrick, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delpi study. *Quality of Life Researech, 19*(4), 539–549.

59. de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine: A practical guide (practical guides to biostatistics and epidemiology).* London: Cambridge University Press.

60. Ploit, D. F., & Yang, F. M. (2016). *Measurement and the measurement of change.* Philadelphia: Wolters Kluwer.

61. Cassidy, S. A., Bradley, L., Bowen, E., Wigham, S., & Rodgers, J. (2018). Measurement properties of tools used to assess depression in adults with and without autism spectrum conditions: A systematic review. *Autism Research, 11*(5), 738–754.

62. van Dijk, S. E. M., Adriaanse, M. C., van der Zwaan, L., Bosmans, J. E., van Marwijk, H. W. J., van Tulder, M. W., et al. (2018). Measurement properties of depression questionnaires in patients with diabetes: a systematic review. *Quality of Life Research, 27*(6), 1415–1430.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.