

Test–retest reliability and sensitivity to change of the dimensional anxiety scales for DSM-5

Susanne Knappe,^{1*} Jens Klotsche,¹ Franziska Heyde,¹ Sarah Hiob,¹ Jens Siegert,¹
Jürgen Hoyer,¹ Anja Strobel,² Richard T. LeBeau,³ Michelle G. Craske,³ Hans-Ulrich
Wittchen,¹ and Katja Beesdo-Baum¹

¹ Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Dresden, Germany

² Department of Psychology, Process-Oriented Assessment, Technische Universität Dresden, Dresden, Germany

³ Department of Psychology, University of California, Los Angeles, California, United States

Objective. This article reports on the test–retest reliability and sensitivity to change of a set of brief dimensional self-rating questionnaires for social anxiety disorder (SAD-D), specific phobia (SP-D), agoraphobia (AG-D), panic disorder (PD-D), and generalized anxiety disorder (GAD-D), as well as a general cross-cutting anxiety scale (Cross-D), which were developed to supplement categorical diagnoses in the *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (DSM-5).

Methods. The German versions of the dimensional anxiety scales were administered to 218 students followed up approximately 2 weeks later (Study 1) and 55 outpatients (23 with anxiety diagnoses) followed-up 1 year later (Study 2). Probable diagnostic status in students was determined by the DIA-X/M-CIDI stem screening-questionnaire (SSQ). In the clinical sample, *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV) diagnoses were assessed at Time 1 using the DIA-X/M-CIDI. At Time 2, the patient-version of the Clinical Global Impression–Improvement scale (CGI-I) was applied to assess change.

Results. Good psychometric properties, including high test–retest reliability, were found for the dimensional scales except for SP-D. In outpatients, improvement at Time 2 was associated with significant decrease in PD-D, GAD-D, and Cross-D scores.

Discussion. Major advantages of the scales include that they are brief, concise, and based on a consistent template to measure the cognitive, physiological, and behavioral symptoms of fear and anxiety. Further replication in larger samples is needed. Given its modest psychometric properties, SP-D needs refinement.

Conclusion. Increasing evidence from diverse samples suggests clinical utility of the dimensional anxiety scales.

Received 16 May 2013; Accepted 20 August 2013; First published online 8 November 2013

Key words: Anxiety, assessment, diagnostic, DSM, psychometric.

Objective

In the 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5), categorical diagnostic criteria for the anxiety disorders are supplemented by dimensional measures to provide information on disorder severity. Previous DSM classifications aimed

to generate almost exclusive and preferably narrow categorical diagnostic groups and diagnostic subtypes, but they neglected however symptom heterogeneity within and co-morbidity across disorder categories.¹ Hence, information on disorder severity is desirable to more precisely map clinical heterogeneity and to improve diagnostic classification,² as well as for allocation to targeted diagnostic procedures, delineation of treatment strategies, and monitoring of the course of symptom over time (ie, across treatment).

Disorder severity ratings, assigned by clinicians, represent a composite of clinician ratings as well as patient (or patient proxy) ratings. For this reason, a set

*Address for correspondence: Susanne Knappe, Technische Universität Dresden, Institute of Clinical Psychology and Psychotherapy, Chemnitz Str. 46, 01187 Dresden, Germany.
(Email: knappe@psychologie.tu-dresden.de)

This article was generated as part of the DSM-5 Work Group activities, ©2012 by the American Psychiatric Association.

of brief (10-item) patient self-rated scales was developed by members of and advisors to the DSM-5 Anxiety, OC Spectrum, Posttraumatic, and Dissociative Disorders work group.* These dimensional anxiety scales are considered to reliably assess the frequency of physiological, cognitive, and behavioral avoidance symptoms for social anxiety disorder, specific phobia, agoraphobia, panic disorder, and generalized anxiety disorder.† All scales follow a common template but differ by introductory statements and reference points for the particular anxiety disorder. In addition to the disorder-specific anxiety scales, a cross-cutting scale (Cross-D) that applies to all anxiety disorders was developed (see LeBeau *et al.*³ for further details). Given their brevity and common template, the dimensional scales are considered to facilitate and supplement diagnostic assessment, in particular for respondents with more than one anxiety disorder.

Previous investigations supported the scales' unidimensionality (internal consistency), convergent and discriminant validity, and sensitivity to clinical severity for both the English and German versions in clinical and non-clinical samples of adults and students,^{3,4} as well as in a Dutch sample of children.⁵ Further, classification performance pointed to accurate discrimination between individuals with versus without a threshold anxiety diagnosis,⁴ and a cut-off score based on additional assessment of impairment and distress was found to assist in the discrimination between subthreshold and threshold anxiety disorders.⁶ Utility of the cross-cutting dimensional anxiety scale was not clearly supported, as the performance of the Cross-D was inferior to the disorder-specific dimensional scales.^{3,4}

Nonetheless, only limited evidence exists with regard to test-retest reliability of the dimensional anxiety scales. On the one hand, anxiety disorders tend to naturally wax and wane,⁷ which may reduce test-retest reliability of assessment instruments. On the other hand, sensitivity to change is crucial for clinical assessment⁸ to allow for both interpersonal comparisons (ie, Is an anxiety disordered patient as anxious as another patient with the same anxiety disorder?) as well as for intrapersonal comparisons (ie, Has the patient's level of anxiety decreased after treatment?). Test-retest reliability has been tentatively established in a small nonclinical sample of undergraduates.³

Thus, this article reports on 2 studies that assessed test-retest reliability in a large convenience sample of students (Study 1) and sensitivity to change in an adult patient sample that underwent psychological treatment for anxiety disorders (Study 2) in Germany. Based on results from previous investigations, we hypothesized high test-retest reliability for all dimensional scales except for specific phobia (SP-D) because psychometric properties for this scale have been consistently lower compared to the other scales, and have shown the lowest test-retest reliability in one previous study.³ In the clinical sample, we expect the scales to prove sensitivity to change as demonstrated by an association between decrease in the scales and (independently measured) mental health improvement.

Methods

Study 1: Participants

Participants of Study 1 were $N = 218$ students (Time 1) who were enrolled in psychology courses during the summer session of 2012 at the Technische Universität Dresden. Students who were age 18 and older were contacted via email and asked to complete an online survey. After granting informed consent, participants provided information on sociodemographics and responded to the DIA-X/M-CIDI stem screening questionnaire (SSQ⁹) to screen for anxiety and other disorder symptoms. Participants were also asked whether they had ever contacted a doctor, psychologist, or a psychiatrist because of any psychological or emotional problems.

Then, the dimensional anxiety scales were presented along with previously validated measures. Approximately 10 days later (Time 2; mean number of days = 10.5, $SD = 2.9$, range: 4 to 23 days after Time 1), $N = 137/218$ (62.8%) students completed the anxiety dimensional scales for a second time (Time 2). Students received course credit or vouchers for participation. The study was approved by the Ethics Committee of the Medical Faculty of the Technische Universität Dresden (No. EK-28022012). Demographic characteristics and the diagnostic profile are presented in Table 1. Dropout analyses revealed differences between participants, who participated at both times compared to those who were not reassessed with regard to age, employment status, and branch of study as well as higher rates of panic disorder and generalized anxiety disorder in those assessed twice. For gender, family status, other diagnostic categories, help-seeking, and scores on dimensional scales, no differences were found (all p -values $> .05$; Suppl. Table S1).

Study 2: Participants

Between November 2010 and February 2011 (Time 1), 102 treatment-seeking adults who consulted the university

* Members and advisors of the Anxiety Disorders Subwork Group of the DSM-5 Anxiety, OC Spectrum, Post-traumatic, and Dissociative Disorder Work Group are Drs. Gavin Andrews, Susan M. Bögels, Michelle G. Craske, Katharine A. Phillips, Murray B. Stein, and Hans-Ulrich Wittchen, and Drs. Lynn Alden, David H. Barlow, Katja Beesdo-Baum, Richard G. Heimberg, Devon Hinton, Stefan G. Hofmann, Donald F. Klein, Ronald M. Rapee, and Richard Zinbarg.

† Dimensional scales were also developed for post-traumatic stress disorder, obsessive compulsive disorder, and separation anxiety disorder, but these, however, are not the focus of this article.

TABLE 1. Sociodemographic characteristics and diagnostic profile of Study 1 and Study 2 participants

	Study 1 (student sample)						Study 2 (clinical sample)					
	time 1 (N = 218)		time 1 only (N = 81)		time 2 (N = 137)		time 1 (N = 102)		time 1 only (N = 47)		time 2 (N = 55)	
	N	%	N	%	N	%	N	%	N	%	N	%
Gender												
female	177	81.2	66	81.5	111	81.0	70	68.6	31	65.96	39	70.9
male	41	18.8	15	18.5	26	18.9	42	31.4	16	34.04	1	29.1
Age in years; mean(Sd), range	24.39 (5.3), 18–47		25.59 (5.98), 19–47		23.67 (4.7), 18–47		32.0 (11.6), 18–66		3 4.59 (11.09) 18–60		33.7 (12.1), 19–67	
18–34 years	209	95.87	75	92.6	134	97.8	70	68.63	33	70.2	37	67.3
35–49 years	9	4.13	6	7.4	3	2.2	21	20.59	9	19.2	12	21.8
50–64 years	-	-	-	-	-	-	10	9.80	5	10.6	5	9.1
65+	-	-	-	-	-	-	1	0.98	0	0.0	1	1.8
Family status												
single	135	61.9	45	55.4	90	65.7	77	75.5	7	14.9	18	32.7
married/living with partner	82	37.6	35	43.2	47	34.3	16	15.7	36	76.6	33	60.0
divorced/widowed	1	0.5	1	1.2	0	0.0	9	8.8	4	8.5	4	7.3
Employment-status												
student/in education	194	89.4	68	84.0	126	92.7	29	28.4	6	12.8	18	32.7
employed	17	7.8	12	14.8	5	3.7	49	48.0	21	44.7	28	50.9
jobless	2	0.9	1	1.2	1	0.7	9	8.8	11	23.4	3	5.5
other (one missing in study 1)	5	1.8	0	0.0	4	2.9	15	14.7	9	19.2	6	10.9
Branch of study (if student; N = 194)												
School of Mathematics and Natural Sciences (incl. Psychology)	143	73.7	52	64.1	102	81.0	-	-	-	-	-	-
School of Engineering Sciences	3	1.6	2	2.5	1	0.8	-	-	-	-	-	-
School of Civil and Environmental Engineering	5	2.6	2	2.5	3	2.4	-	-	-	-	-	-
School of Humanities and Social Sciences	37	19.1	22	27.2	18	14.3	-	-	-	-	-	-
Medical School-Medicine	4	2.1	2	2.5	2	1.6	-	-	-	-	-	-
Others	2	1.0	1	1.2	0	0.0	-	-	-	-	-	-
Diagnostic profile ^a												
any anxiety disorder ^b	100	45.9	47	58.0	53	38.7	46	48.0	23	51.06	23 ^f	41.8
social anxiety disorder ^c	29	13.3	12	14.8	17	12.4	17	16.7	9	17.02	9 ^f	16.4
specific phobia ^c	18	8.3	9	11.1	9	6.6	19	18.6	10	21.28	9 ^f	16.4
agoraphobia ^c	10	4.6	4	4.9	6	4.4	15	14.7	10	21.28	5 ^f	9.1
panic disorder ^d	58	26.6	28	34.6	30	21.9	10	9.8	6	12.77	4 ^f	7.3
generalized anxiety disorder ^d	56	25.7	27	33.3	29	21.2	17	16.7	8	17.02	9 ^f	16.4
Help-seeking ^e												
yes	63	28.9	23	28.4	97	70.8	-	-	-	-	-	-
no	155	71.1	58	71.6	40	29.2	-	-	-	-	-	-

time 1–baseline assessment at time 1; time 1 only – only baseline, but no follow-up assessment; time 2 – assessed at baseline and follow-up

^a Study 1: SSSQ, positive affirmation of the respective screening question; Study 2: DIA-X/M-CIDI diagnoses in the past 4 weeks

^b includes social anxiety disorder; specific phobia, agoraphobia, panic disorder, generalized anxiety disorder

^c in the past 12 months

^d lifetime

^e response to 1 item: 'Did you ever contact a doctor, a psychological because of psychological or emotional problems?'

^f current at Time 1

outpatient clinic for psychotherapy at the Technische Universität Dresden (Germany) for mental health problems completed the dimensional anxiety scales and previously validated measures along with a standardized diagnostic interview to establish *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV) diagnoses of mental disorders (DIA-X/M-CIDI⁹). Previous reports^{4,6} are based on Time 1 sampling. For this study, from January to August 2012 (Time 2; mean number of months = 15.8; SD = 2.7, range: 10–20 months after Time 1), N = 55/102 (53.9%) adults were reassessed with the dimensional and previously validated anxiety scales, supplemented with

general questions about the current status of mental and physical health, irrespective of whether they were allocated to treatment after Time 1 assessment. As shown in the study flow chart (Figure 1), 40/55 (72.7%) participants started cognitive behavioral psychotherapy in the outpatient clinic, and of those, 26 (65.0%) completed therapy (mean number of sessions = 28.4; SD = 14.9; range: 8–60 sessions) at Time 2, 8 (20.0%) were still in treatment (mean number of sessions = 22.0; SD = 8.5; range: 7–34 sessions) at Time 2, and 6 (15.0%) dropped out of treatment (mean number of sessions = 16.8; SD = 12.1; range: 1–31 sessions). All participants provided

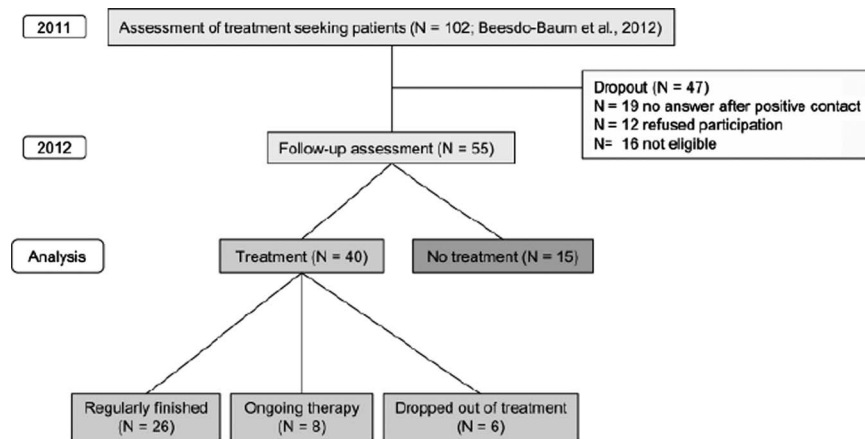


FIGURE 1. Flow chart of participants through both points of time.

written informed consent. Demographic characteristics and the diagnostic profile of Study 2 participants are presented in Table 1. For Time 1, there were no significant differences in demographic characteristics or diagnoses between participants that were reassessed at Time 2 compared to those who were not reassessed (all p -values < .05).

Assessments

DSM-5 dimensional anxiety scales

Development and composition of the dimensional anxiety scales have been described previously.³ Briefly, a 10-item template was created to assess the frequency of cognitive symptoms, physical symptoms, and escape and avoidance behaviors related to fear and anxiety. This general template was then adapted for social anxiety disorder (SAD-D), specific phobia (SP-D), agoraphobia (AG-D), panic disorder (PD-D), and generalized anxiety disorder (GAD-D) through the use of different introductory statements and different reference points throughout the items. Scales were translated into German language by one of the authors (HUW), and quality of the translation was ensured via back-translation by a native English speaker. In the German version of the scales (available upon request; see LeBeau *et al.*³ for the English version of the dimensional scales), each 10-item dimensional scale was supplemented by 2 items to assess disorder-related impairment (eg, for SAD-D: “I couldn’t take care of important issues because of fear of social situations”) and distress (“I felt stressed and burdened because of my problems related to social situations”). All items were anchored to the past 4 weeks and were rated on a 5-point Likert-type scale ranging from 0 (“never”) to 4 (“all of the time”).

A sum score was created (including distress and impairment ratings, possible range of scores 0–48). Unidimensionality and psychometric properties of the 10- and 12-item versions were comparable, and scale reliability of both versions was excellent.⁴ For the German Cross-D scale (10 items), the item “needed help to cope with anxiety” was omitted, and instead, an item assessing distress (“I felt stressed out because of my fear and anxiety.”) was used (sum score range from 0–40).

Respondents who were not reassessed at Time 2 exhibited similar scores on the dimensional anxiety scales as compared to participants who participated at both assessments (Suppl. Table S2).

Previously validated measures

Based on their wide use and the strength of their psychometric properties, the German versions of several established self-report scales for each of the anxiety diagnoses were used to examine convergent and discriminant validity of the dimensional anxiety scales: the Brief Symptom Inventory [BSI¹⁰; a 53-item scale assessing symptoms in nine dimensions including anxiety (panic disorder) and phobia (agoraphobia)]; the Fear Questionnaire [FQ¹¹; assesses how likely an individual is to avoid certain situations (agoraphobia: FQ-A, 5 items; social phobia: FQ-Soc, 5 items; blood/injury-specific phobia subtype: FQ-BI, 5 items)]; the Mini-SPIN¹² (a 3-item screening assessment tool for generalized social anxiety disorder); and the GAD-7¹³ (a 7-item measure assessing symptoms of general anxiety disorder). Psychometric properties of these previously validated measures are recognized to be high.¹⁴ The Patient-Reported Outcome Measurement Information System–Anxiety scale (PROMIS-Anx¹⁵), a 7-item scale assessing the frequency of anxiety symptoms over the

past week, was included as an established measure of general anxiety symptoms with reliability above .89 for the majority of the score distribution.¹⁶

Study 1: specific assessments

Previous psychometric analyses of AG-D and SP-D used only proxy validators to establish validity.^{4,6} Hence, for this study, the Panic and Agoraphobia Scale (PAS¹⁷) was additionally used to assess severity within 13 items and 5 subscales of the main factors of panic disorder and agoraphobia (panic attacks, avoidance, anticipatory anxiety, disability, and worries about health). Similarly, the Fear Survey Schedule (FSS-III^{18,19}) was administered to assess a larger spectrum of specific fears such as fears of animals, injury/illness, classical phobias (eg, being alone), social stimuli, noises, and assorted other stimuli (eg, falling, insects).

Study 2: specific assessments

To assess change in mental health state in Study 2 participants (patients), the Clinical Global Impression–Improvement Scale (CGI-I) was used to rate the severity of symptoms as a result of treatment, ie, the change a patient made after a treatment initiated.²⁰ The validity and the sensitivity to change of the CGI-I have been established.²¹ The outpatient clinic uses this instrument both for therapists and also uses an adapted version for patients to monitor the progress of the treatment. Since objective CGI-I ratings of the therapists were both not actual and not available for all participants of Time 2, the patient-version of CGI-I²² was used to estimate change. Participants were asked to rate to what extent their mental or psychosomatic complaints improved or worsened in comparison to the beginning of treatment on a 7-point Likert-type scale.

Statistical Analyses

Statistical analyses were performed with the statistical software package STATA 12,²³ except for the confirmatory factor analyses, which were conducted using MPLUS 6.1.²⁴ No adjustment for multiple testing was applied, because the individual tests were related to individual hypotheses and adjustment would treat them as reflecting a global hypothesis, which is questionable in substantive terms.²⁵

Total scores were calculated for each of the dimensional anxiety scales (SAD-D, SP-D, GAD-D, AG-D, PD-D; 10- and 12-item versions), for the Cross-D, and also for the validated scales (Mini-SPIN, FQ, GAD-7, PAS, FSS-III, FPRMIS-Anx). The BSI was scored using means.

For Study 1, psychometric properties of the specific dimensional anxiety scales and the Cross-D were

determined in concordance with previous analyses.^{4,6} Briefly, (uni-)dimensionality of the dimensional anxiety scales was examined with confirmatory factor analysis.²⁶ Adequate model fit was evaluated by the fit indices as suggested by Hu and Bentler.²⁷ Internal consistency was determined according to Cronbach's alpha coefficient. Convergent and discriminant validity of the dimensional scales was analyzed with Pearson correlations between the dimensional scales and each of the validated scales. The test for correlated correlation coefficients²⁸ was used for comparing the correlation coefficients for conceptually similar measures (eg, SAD-D and Social Phobia Scale of the FSS) to conceptually different measures (eg, SAD-D and PAS, or SAD-D and GAD-7). For the Cross-D, Pearson correlations between the Cross-D total score and the total scores on each of the dimensional scales were calculated. To establish classification performance, the area under the curve (AUC) was calculated to determine the accuracy at which each dimensional scale differentiates between affected and non-affected participants, and among those with and without help-seeking.

Test-retest reliability (Study 1) was examined using intraclass correlational coefficients (ICCs) between the total score on each of the 5 disorder-specific dimensional scales and the Cross-D at Time 1 with the total score on each scale at Time 2. Being consistent with standards in the field and LeBeau *et al.*,³ test-retest reliability needed to exceed .70.

As Study 2 builds on the sample of Beesdo-Baum *et al.*,⁴ results on psychometric properties of the dimensional scales will not be repeated here. The primary aim of reassessment of the clinical sample was to examine the scales' sensitivity to change (Study 2). Therefore, linear regression analyses were computed with CGI-I as independent variable and the change in scores of the dimensional anxiety scales (Time 2 minus Time 1) as the dependent variable. Since Beesdo-Baum *et al.*⁴ demonstrated that patients with DSM-IV-defined specific diagnoses scored higher on the corresponding disorder-specific scale, the regression analyses for the dimensional scales were computed for the participants with the respective diagnoses (as assessed at Time 1). Due to the non-normal distribution of the dimensional scores (Time 1) and previously validated measures (Time 2; tested with the Shapiro-Wilk test) and the small sample size, the standard errors and 95% confidence intervals for parameter estimates of the linear regression analyses were estimated by bootstrapping for bias accelerated.²⁹

Welch's test for continuously distributed data and chi-square statistics for categorical data were used to analyze differences between participants who were assessed both times and those who dropped out after the first assessment in both Studies 1 and 2.³⁰

TABLE 2. Descriptives for the dimensional anxiety scales (10-items) and Cross-D (9 items)

	Study 1 (student sample)						Study 2 (clinical sample)					
	time 1 (N = 218)		time 1 only (N = 81)		time 2 (N = 137)		time 1 (N = 102)		time 1 only (N = 47)		time 2 (N = 55)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
SAD-D	4.7	4.7	5.0	5.8	4.8	5.6	11.3	9.3	10.3	9.6	5.8	6.9
SP-D	3.2	5.2	4.1	6.4	2.8	5.1	8.2	9.0	9.3	10.9	4.2	7.5
AG-D	2.0	4.5	2.7	5.9	1.6	4.3	6.9	9.4	8.8	10.8	3.7	7.8
PD-D	2.0	4.0	2.3	5.0	2.2	4.1	7.6	9.3	9.3	10.9	4.2	7.6
GAD-D	5.3	5.3	6.1	5.6	4.9	5.8	11.8	9.9	12.5	10.3	7.2	7.8
Cross-D	6.1	5.8	6.8	6.0	5.7	6.3	14.0	9.2	14.2	10.1	8.1	8.5

M mean, SD standard deviation

SAD social anxiety disorder, SP specific phobia, AG agoraphobia, PD panic disorder, GAD generalized anxiety disorder, Cross-D cross-cutting dimensional scale time 1 - baseline assessment at time 1; time 1 only - only baseline, but no follow-up assessment; time 2 - assessed at baseline and follow-up

TABLE 3. Results of Confirmatory Factor Analyses for dimensionality (one) and scale reliability in Study 1 (N = 218 students)

Scale	Parameter	10-item scale	12-item scale
SAD-D	Chi2 (df); p Value	67.7 (34); p < .001	101.5 (51); p < .001
	CFI/TLI/SRMR	0.97/0.96/0.04	0.97/0.95/0.04
	reliability ¹	0.89	0.92
SP-D	Chi2 (df); p Value	127.1 (33); p < .001	206.2 (51); p < .001
	CFI/TLI/SRMR	0.92/0.88/0.05	0.89/0.86/0.06
	reliability ¹	0.90	0.91
AG-D	Chi2 (df); p Value	193.7 (32); p < .001	300.6 (51); p < .001
	CFI/TLI/SRMR	0.89/0.85/0.07	0.87/0.83/0.07
	reliability ¹	0.91	0.93
PD-D	Chi2 (df); p Value	137.2 (31); p < .001	229.9 (49); p < .001
	CFI/TLI/SRMR	0.92/0.88/0.05	0.90/0.87/0.05
	reliability ¹	0.88	0.90
GAD-D	Chi2 (df); p Value	96.19 (33); p < .001	167.8 (52); p < .001
	CFI/TLI/SRMR	0.93/0.90/0.05	0.91 0.88/0.05
	reliability ¹	0.87	0.90
Cross-D	Chi2 (df); p Value	96.45 (34); p < .001	
	CFI/TLI/SRMR	0.94/0.92/0.05	
	reliability ¹	0.90	

SRMR = 0 indicates perfect fit; SRMR = 0.05 good fit and SRMR = 0.08 is considered adequate fit CFI (Comparative Fit Index) and TLI (Tucker-Lewis Index; Relative Noncentrality Index); values close to 1 indicate good fit, values around 0.9 indicate acceptable fit

¹ Cronbach's Alpha

SAD social anxiety disorder, SP specific phobia, AG agoraphobia, PD panic disorder, GAD generalized anxiety disorder, Cross-D cross-cutting dimensional scale

Results

Because of similar results and in order to be consistent with previous publications,^{3,4} only results of the 10-item versions of the dimensional anxiety scales will be reported (results of the 12-item versions were similar and are available upon request). Descriptives for the dimensional anxiety scales (10-items) and Cross-D for Study 1 and Study 2 are presented in Table 2. There were no significant differences in dimensional scale scores between participants of both time points compared to those who were only assessed at Time 1.

Study 1: confirmation of psychometric properties

Unidimensionality was established for each dimensional scale (Table 3), and internal consistency of each of the scales ranged between .87 for GAD-D and .91 for AG-D, and was .90 for Cross-D.

Convergent validity was demonstrated by higher correlations between dimensional anxiety scales and corresponding previously validated anxiety scales as compared to correlations with noncorresponding measures (Table 4). The corresponding correlation coefficients were significantly higher for SAD-D and

TABLE 4. Convergent and discriminant validity of 10-item dimensional anxiety scales in Study 1 (N = 218 students)

Dimensional Anxiety Scales	Previously Validated Scales																					
	Mini-SPIN		FQ-Subscales				BSI-Subscales				PAS		FSS-III-Subscales				GAD-7					
	Corr ¹	P Value ²	Social Phobia		Blood-Injury Phobia		Agoraphobia		Phobia (agoraphobia)		Anxiety (panic)		Social Phobia		Specific Phobia		Agoraphobia		Corr ¹	P Value ²		
			Corr ¹	P Value ²	Corr ¹	P Value ²	Corr ¹	P Value ²	Corr ¹	P Value ²	Corr ¹	P Value ²	Corr ¹	P Value ²	Corr ¹	P Value ²	Corr ¹	P Value ²				
SAD-D	.69 (<.001)	(ref.)	.65 (<.001)	.152	.16 (.016)	<.001	.41 (<.001)	<.001	.56 (<.001)	.050	.59 (<.001)	.023	.41 (<.001)	<.001	.67 (<.001)	.278	.19 (.005)	<.001	1.54 (<.001)	.020	1.60 (<.001)	.030
SAD-D	.69 (<.001)	.152	.65 (<.001)	(ref.)	.16 (.016)	<.001	.41 (<.001)	<.001	.56 (<.001)	.022	.59 (<.001)	.092	.41 (<.001)	<.001	.67 (<.001)	.352	.19 (.005)	<.001	.54 (<.001)	.05	.60 (<.001)	.127
SAD-D	.69 (<.001)	.278	.65 (<.001)	.352	.16 (.016)	<.001	.41 (<.001)	<.001	.56 (<.001)	.012	.59 (<.001)	.037	.41 (<.001)	<.001	.67 (<.001)	(ref.)	.19 (.005)	<.001	.54 (<.001)	.020	.60 (<.001)	.048
SP-D	.31 (<.001)	.392	.33 (<.001)	.476	.33 (<.001)	(ref.)	.22 (<.001)	.067	.37 (<.001)	.291	.44 (<.001)	.058	.37 (<.001)	.289	.40 (<.001)	.168	.33 (<.001)	.484	.48 (<.001)	.013	.36 (<.001)	.352
SP-D	.31 (<.001)	.398	.33 (<.001)	.463	.33 (<.001)	.484	.22 (<.001)	.073	.37 (<.001)	.280	.44 (<.001)	.048	.37 (<.001)	.286	.40 (<.001)	.109	.33 (<.001)	(ref.)	.48 (<.001)	.006	.36 (<.001)	.337
AG-D	.42 (<.001)	.143	.43 (<.001)	.140	.16 (.018)	<.001	.49 (<.001)	(ref.)	.67 (<.001)	<.001	.51 (<.001)	.331	.61 (<.001)	.016	.42 (<.001)	.135	.17 (.012)	<.001	.66 (<.001)	<.001	.44 (<.001)	.224
AG-D	.42 (<.001)	<.001	.43 (<.001)	<.001	.16 (.018)	<.001	.49 (<.001)	<.001	.67 (<.001)	(ref.)	.51 (<.001)	<.001	.61 (<.001)	.103	.42 (<.001)	<.001	.17 (.012)	<.001	.66 (<.001)	.398	.44 (<.001)	<.001
AG-D	.42 (<.001)	.020	.43 (<.001)	.020	.16 (.018)	<.001	.49 (<.001)	.016	.67 (<.001)	.103	.51 (<.001)	.020	.61 (<.001)	(ref.)	.42 (<.001)	.010	.17 (.012)	<.001	.66 (<.001)	.139	.44 (<.001)	.020
AG-D	.42 (<.001)	<.001	.43 (<.001)	<.001	.16 (.018)	<.001	.49 (<.001)	<.001	.67 (<.001)	.398	.51 (<.001)	<.001	.61 (<.001)	.139	.42 (<.001)	<.001	.17 (.012)	<.001	.66 (<.001)	(ref.)	.44 (<.001)	<.001
PD-D	.34 (<.001)	<.001	.38 (<.001)	<.001	.30 (<.001)	<.001	.42 (<.001)	<.001	.50 (<.001)	.003	.63 (<.001)	(ref.)	.64 (<.001)	.418	.42 (<.001)	<.001	.24 (<.001)	<.001	.58 (<.001)	.141	.48 (<.001)	<.001
PD-D	.34 (<.001)	<.001	.38 (<.001)	<.001	.30 (<.001)	<.001	.42 (<.001)	<.001	.50 (<.001)	.040	.63 (<.001)	.418	.64 (<.001)	(ref.)	.42 (<.001)	<.001	.24 (<.001)	<.001	.58 (<.001)	.121	.48 (<.001)	.003
GAD-D	.45 (>.001)	<.001	.43 (<.001)	<.001	.30 (<.001)	<.001	.37 (<.001)	<.001	.51 (<.001)	<.001	.67 (<.001)	.015	.51 (<.001)	<.001	.57 (<.001)	<.001	.32 (<.001)	<.001	.56 (<.001)	<.001	.73 (<.001)	(ref)

ref. = reference correlation for test of correlation coefficient, printed in bold

¹ Correlation (significance level)

² P Value for comparing correlation coefficient between a dependent variable (ref.) and a set of independent variables in the row

p ≤ .05

SAD social anxiety disorder, SP specific phobia, AG agoraphobia, PD panic disorder, GAD generalized anxiety disorder

TABLE 5. Classification performance of the dimensional anxiety and Cross-D scales in Study 1 (N = 218 students)

	Group 1			Group 2			Group 1 vs Group 2		Group 3			Group 4			Group 3 vs Group 4	
	SSQ-item ^{a,b} negatively affirmed			SSQ-item ^{a,b} positively affirmed			Comparison ^c and Classification		SSQ-item ^{a,b} positively affirmed WITHOUT help-seeking			SSQ-item ^{a,b} positively affirmed WITH help-seeking			Comparison ^c and Classification	
	N	M	SD	N	M	SD	P	AUC	N	M	SD	N	M	SD	p	AUC
SAD-D	189	3.4	3.7	29	12.7	7.8	<.001	0.89	13	10.2	5.3	16	14.8	9.1	.095	0.64
SP-D	200	2.7	4.7	18	8.9	7.0	.001	0.79	11	8.1	4.7	7	10.3	9.8	.596	0.53
AG-D	208	1.4	3.4	10	14.1	7.8	<.001	0.86	4	11.8	7.8	6	15.7	8.1	.446	0.73
PD-D	160	0.8	1.7	58	5.2	6.2	<.001	0.77	32	4.2	4.9	26	6.5	7.5	.183	0.59
GAD-D	162	3.9	3.9	56	9.2	6.8	<.001	0.73	26	7.7	6.6	30	10.5	6.8	.129	0.61
Cross-D	118	3.7	3.9	100	8.9	6.4	<.001	0.76	59	7.2	5.5	41	11.3	7.0	.002	0.68

N number, M mean, SD standard deviation

^a based on positive affirmation of the SSQ-screening question regarding the corresponding anxiety disorder

^b for Cross-D: based on positive affirmation of any of the SSQ-screening questions for the anxiety disorders

^c Comparisons are based on univariate regressions, $p < .05$; AUC area under the curve

GAD-D ($p < .001$). The corresponding correlations between PD-D and BSI-Anxiety and between PD-D and PAS were similar to the noncorresponding correlation between PD-D and FSS-Agoraphobia ($p = .121$, $p = .141$). Further, the corresponding correlation between AG-D and FQ-Agoraphobia was mostly similar to their noncorresponding correlations (BSI-Anxiety: $p = .331$, Mini-SPIN: $p = .143$, FQ-Social Phobia: $p = .140$, FSS-Social Phobia: $p = .135$, GAD-7: $p = .224$). For SP-D, where the corresponding measures cover only 1 (FQ-Blood/Injury Phobia: blood/injury) or 2 (FSS-Specific Phobia: animal and blood/injury) disorder subtypes, moderate correlation coefficients between SP-D and both of the corresponding measures ($r = .33$, $r = .33$) were observed. Corresponding correlations for the specific dimensional scales and PROMIS-Anx were moderate to high (r range from .53 for SP-D to .84 for GAD; $r = .78$ for Cross-D; all p -values $< .001$), and similarly, they were moderate to high for the specific dimensional scales and Cross-D (r range from .53 for SP-D to .84 for PD-D; $r = .77$ for PROMIS-Anx; all p -values $< .001$).

As shown in Table 5, participants who positively affirmed the specific SSQ-stem question with and without help-seeking reported higher scores on the specific dimensional anxiety scales and the Cross-D than respondents without an indication for an anxiety disorder ($p < .05$). Classification performance was good to high for the dimensional anxiety scales and the Cross-D with AUC between .76 for Cross-D and .89 for SAD-D. Of note, among positively screened participants, scores were apparently higher for those with help-seeking than for those without, though these differences were only statistically firm for Cross-D ($p < .05$). Here, AUC was low to moderate, ranging between .53 for SP-D to .67 for Cross-D.

Study 1: test-retest reliability

At Time 2, $N = 137/218$ participants were included in the test-retest analyses. ICCs were calculated between the dimensional anxiety scales completed at Time 1 and Time 2: With the exception of SP-D (ICC = .66), strong test-retest reliability for the dimensional anxiety scales (ICCs range from .77 to .81) and the Cross-D (ICC = .85) could be demonstrated. No significant differences emerged for test-retest correlations between participants who completed second time assessment within the allotted time frame and those who did not (Suppl. Table S3).

Study 2: sensitivity to change

Regression analyses based on 55/102 patients who participated at Time 1 and Time 2 assessments revealed that the greater the improvement in CGI-I, the higher the improvement (lower score) in the mean difference score of the dimensional anxiety scales (Time 2 minus Time 1; Figures 2a-2f). Specifically, for participants with any anxiety disorder at Time 1, one point more on CGI-I (improved) resulted in 2.74 points less (mean difference, $p < .05$) on Cross-D at Time 2.

Discussion

A set of brief dimensional self-rating questionnaires for social anxiety disorder, specific phobia, agoraphobia, panic disorder, and generalized anxiety disorder, as well as a general cross-cutting dimensional anxiety scale were developed by the Anxiety Disorders Subgroup of the DSM-5 Anxiety, OC Spectrum, Posttraumatic, and Dissociative Disorder Work Group in order to support the categorical approach of diagnostic classification of

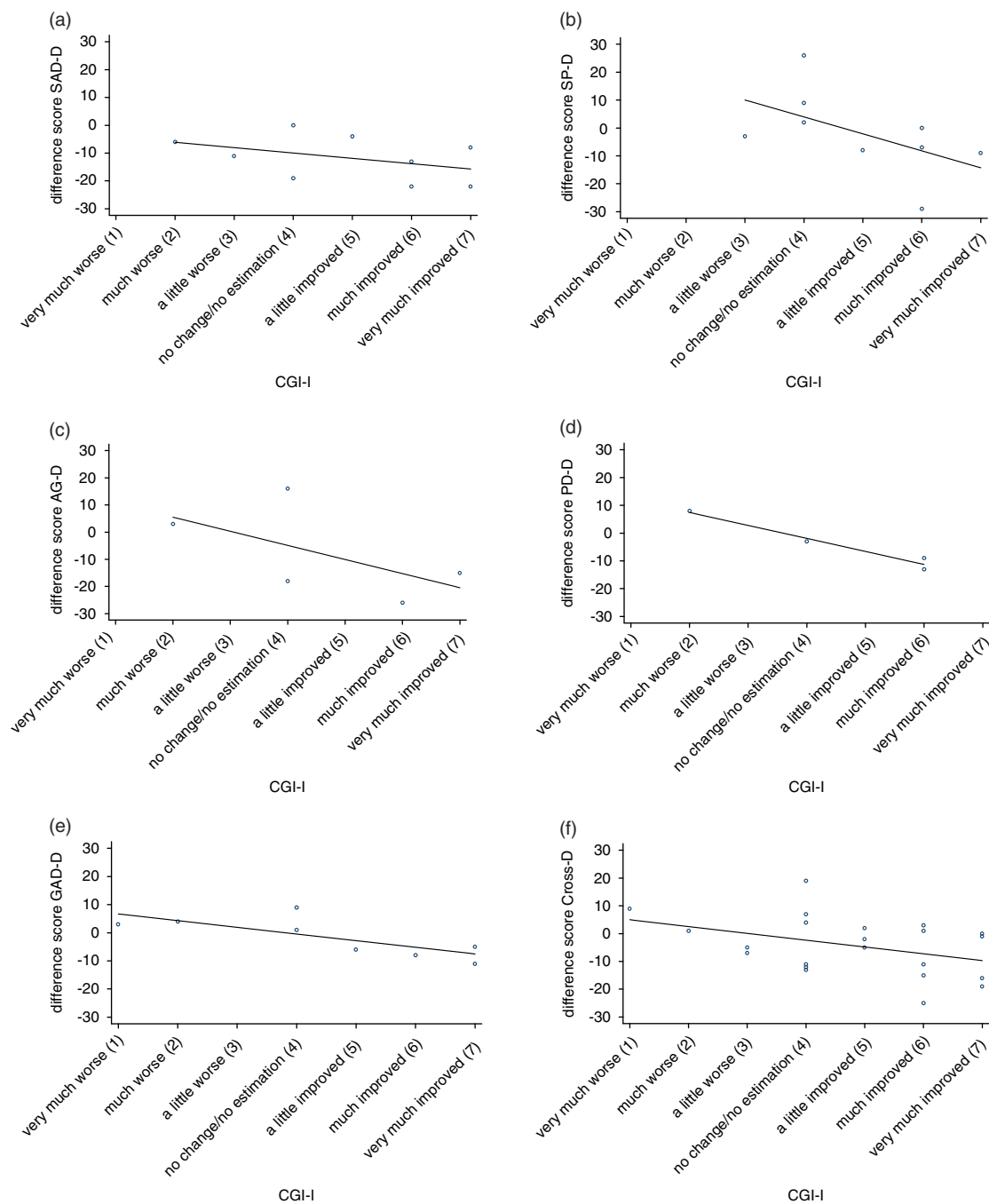


FIGURE 2. Scatterplots of the relation between the difference of scores on the dimensional anxiety scales (Time 2 – Time 1) and the Clinical Global Impression—Improvement Scale (CGI-I) from linear regression analyses. (a) Relation between difference of the SAD-D scores (Time 2 – Time 1) and CGI-I (Beta = -1.9 [-4.6 to 0.8], $p = .166$); (b) relation between differences of the SP-D scores (Time 2 – Time 1) and CGI-I (Beta = -6.1 [13.4 to 1.3], $p = .106$); (c) relation between the difference of AG-D scores (Time 2 – Time 1) and CGI-I (Beta = -5.2 [-13.3 to 2.9], $p = .206$); (d) relation between the difference of PD-D scores (Time 2 – Time 1) and CGI-I (Beta = -4.7 [-5.9 to -3.4], $p < .001$); (e) relation between the difference of GAD-D scores (Time 2 – Time 1) and CGI-I (Beta = -2.4 [-4.3 to -0.5], $p = .014$); (f) relation between the difference of Cross-D scores (Time 2 – Time 1) and CGI-I (Beta = -2.7 [-5.2 to -0.3], $p = .030$).

DSM-5 by a dimensional assessment. Our results from study 1 replicate and expand previous findings^{3,4,6} on the clinical utility and good psychometric properties of the

dimensional scales to a larger convenience sample of young adults. Of note, all scales were administered online (Web-based) to facilitate data collection and analyses.

In addition, high test-retest reliability for a 2-week period could be largely demonstrated, except for SP-D, which may be explained by inconsistent contact with the phobic situation or object at the 2 points of measurement. Similarly, ratings of specific phobic situations might be inexact in cases with more than one phobia, caused by a lack of reference to one specific and concrete situation.⁴ In Study 2, and despite rather low case numbers, sensitivity to change approximately 1 year after initial assessment was found significant for the dimensional scales PD-D, GAD-D, and Cross-D, and trends for SAD-D, SP-D and AG-D emerged.

Expanding previous studies, we also included the PAS and FSS-III in Study 1 to account for the severity of panic disorder and agoraphobia as well as a broader spectrum of specific fears. In previous investigations, validity of the respective dimensional scales was demonstrated, but was modest, particularly for SP-D. Concurrent and discriminant validity was supported for AG-D and PD-D, but remained poor for SP-D—even when alternative validity measures covering the blood/injury and the animal subtype of specific phobia were considered. Hence, poor validity for the SP-D scale may be explained by imprecise introduction and/or item formulation. For AG-D, concurrent validity was supported using the PAS and FSS-Agoraphobia subscale as alternative validity measures. Of note, correlations of AG-D were highest for BSI phobia, PAS and FSS-Agoraphobia, questioning the FQ-Agoraphobia as a validity measure in a nonclinical sample.

From the variety of validity measures, future studies should however consider which scales to use to investigate the psychometric properties, acknowledging time constraints and motivation of participants. Our studies indicate preference of the Mini-SPIN against the FSS-III for SAD-D, the BSI phobia, PAS, or FSS-III over the FQ-Agoraphobia for AG-D, the BSI anxiety or PAS for PD-D, and the GAD-7 for GAD-D. We further suggest carefully considering use of the FSS-III because of its length, and because results did not turn out to be superior to other validity measures. For SP-D, validity needs to be further evaluated, probably regarding both item formulation and selection of validity measures.

To also investigate the dimensional scales' sensitivity to change, a subsample of adult attendees to a German outpatient clinic⁴ was re-examined approximately 1 year after initial assessment. Evidence was found at least for PD-D, GAD-D, and Cross-D. Nevertheless, confidence intervals were relatively narrow despite fairly small sample sizes, tentatively suggesting underlying effects despite failed significance for SAD-D, SP-D, and AG-D.

Overall utility of the Cross-D as a more brief cross-cutting measure has been difficult to evaluate so far. One study observed insufficient correlations between disorder-specific scales and Cross-D,³ while others⁴

found moderate to high correlations to the disorder-specific scales. Study 1 thus re-examined psychometric properties of the Cross-D, demonstrating again concurrent validity of the Cross-D because of its correlations with disorder-specific scales (except for SP-D) and with the PROMIS-Anx. However, correlations between Cross-D and dimensional scales ranged between .60 for AG-D and .84 for PD-D, suggesting that the Cross-D may be more related to some anxiety disorders than to others. If so, then use of the Cross-D as an independent measure across the anxiety disorders would be problematic. The Cross-D was, however, the only scale to differentiate between probable cases with and without help-seeking. Also, test-retest reliability and sensitivity to change was as good for the Cross-D as for the dimensional anxiety scales, arguing for its clinical utility at least on a broad level (ie, any anxiety disorder).

Of note, dimensional anxiety measures have been used for decades to reflect the frequency, severity, or variety of symptoms, thereby accounting for the dimensional nature of the underlying psychological phenomena.^{2,31,32} From that, one may ask why should we put so much emphasis on the dimensional anxiety scales for DSM-5? Is it not just simply a repetition of what is already known? Broadly speaking, this new dimensional approach is highly of interest for clinical practice, because it assesses symptoms of disorders nearer to their real nature of psychopathology. More precisely, the major advantages of the “new” scales are that they are brief, concise, and based on a consistent template to measure the multiple symptoms of fear and anxiety, covering cognitive, physiological, and behavioral aspects, across a range of diagnostic constructs. In particular for respondents with more than one anxiety disorder, which is the rule in the majority of cases, the scales may facilitate assessment by reflecting what kind of symptoms are present for which anxiety disorder and to what degree.

The strengths of our approach include the recruitment and reassessment of a relatively large student and clinical sample for psychometric and test-retest evaluation of the dimensional scales, demonstrating applicability of the scales also in online surveys, and the application of widely used self-report measures for validation purposes. Drop-out analyses revealed that differences between respondents who participated at Time 1 only in contrast to those participating at both assessments did not affect findings in either study.

Our findings need, however, to be considered with regard to some caveats. In both studies, participants were relatively young, well educated, and single, which limits generalizability of results to older or less educated samples. There was a preponderance of females in both samples. Given epidemiological and frequent clinical observations that females are more often affected than males by anxiety disorders, and that anxiety disorders are

more disabling in females than in males in terms of comorbidity patterns and illness burden,³³ our sample and study findings likely map relations in clinical practice, without necessarily limiting generalizability of our findings. Comorbidity was not evaluated here. However, most recent findings on the dimensional anxiety scales' sensitivity to clinical severity indicated that self-reports of adult patients were largely unaffected by the number of anxiety disorders or depressive disorders.⁶

In Study 1 (student sample), we used the SSQ to screen for the probable diagnostic status. The SSQ is part of the DIA-X/M-CIDI,⁹ for which reliable and valid assessment of mental disorders was established.^{34,35} Acknowledging the limits of screening inventories,¹⁴ help-seeking was considered to indicate severity or impairment due to (probable) diagnostic status. However, as this may have limited group sizes for comparisons of disorder-specific dimensional scales and the Cross-D for participants with vs without an indication for anxiety disorder, help-seeking did not validate caseness.

In Study 2, the low response rate of 53.9% (55/102 participants) at second assessment resulted in a small sample size available for analyses, particularly concerning specific diagnoses. Analyses with more statistical power might have yielded more significant findings. Diagnoses are limited to self-report and clinician ratings. Measurement of change is based on patient's self ratings, which may be subject to bias. Clinicians' ratings of change were not available for all patients for the exact time of dimensional scale reassessment. In addition, the most optimal design to evaluate the scales' sensitivity to change would require including a non-treated comparison group,³⁶ which was not feasible in our study. Future studies should therefore include objective change ratings that are related to actual diagnoses and replicate analyses in larger samples when examining the dimensional scales' sensitivity to change.

Conclusion

Good psychometric properties of the brief, consistent, and generally formulated dimensional anxiety scales for DSM-5 have been established in terms of unidimensionality, convergent and discriminant validity, classification performance, and sensitivity to clinical severity. Using a larger sample of students and reassessment of a clinical sample, these most recent studies replicate previous findings and, in addition, demonstrate high test-retest reliability and sensitivity to change. Hence, evidence from different samples (students vs clinical sample) and across different ages (young vs middle-age adults) increases for the clinical utility of the dimensional anxiety scales in clinical settings and relative fields (ie, prospective observations). However, revision of SP-D for

the assessment of specific phobias is implied given the modest validity and reliability. For the Cross-D, there is similar evidence for its clinical utility, its sensitivity to change, and the utility of the Cross-D as an instrument to assess anxiety disorders, although further evaluation concerning beneficial effects beyond the specific dimensional scales is implied. Overall, these findings strongly argue for the use of the new dimensional anxiety scales for routine use in clinical practice.

Disclosures

Susanne Knappe has nothing to disclose. Katja Beesdo-Baum has nothing to disclose. Jens Klotsche has nothing to disclose. Franziska Heyde has nothing to disclose. Sarah Hiob has nothing to disclose. Jürgen Hoyer has the following disclosure: Astra Zeneca, independent contractor, speaking honoraria. Anja Strobel has nothing to disclose. Richard LeBeau has nothing to disclose. Michelle Craske has the following disclosures: UCLA, employee, salary; NIMH, grant funding, research support; APA Books, type of activity: other, book royalties; Oxford University Press, type of activity: other, book royalties. Hans-Ulrich Wittchen has the following disclosures: Servier, advisor, honoraria; Pfizer, advisor, honoraria; Lundbeck, advisor, honoraria; Novartis, advisor, honoraria. Jens Siegert has nothing to disclose.

Supplementary materials

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/S1092852913000710>

REFERENCES:

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Arlington, VA: American Psychiatric Association; 2013: 5-25.
2. Shear KM, Bjelland I, Beesdo K, Gloster AT, Wittchen H-U. Supplementary dimensional assessment in anxiety disorders. *Int J Methods Psychiatr Res*. 2007; **16**(Suppl 1): S52-S64.
3. LeBeau RT, Glenn DE, Hanover L, et al. A dimensional approach to measuring anxiety for DSM-5. *Int J Methods Psychiatr Res*. 2012; **21**(4): 258-272.
4. Beesdo-Baum K, Klotsche J, Knappe S, et al. Psychometric properties of the dimensional anxiety scales for DSM-5 in an unselected sample of treatment seeking patients. *Depress Anxiety*. 2012; **29**(12): 1014-1024.
5. Möller E, Majdandzic M, Craske MG, Bögels S. Dimensional assessment of anxiety disorders in parents and children for DSM-5. *Int J Methods Psychiatr Res*. In press.
6. Knappe S, Klotsche J, Strobel A, et al. Dimensional anxiety scales for DSM-5: sensitivity to clinical severity. *Eur Psychiatry*. 2013; **28**(7): 448-456.
7. Beesdo-Baum K, Knappe S. Developmental epidemiology of anxiety disorders. *Child Adolesc Psychiatr Clin N Am*. 2012; **21**(3): 457-478.
8. Fava GA, Ruini C, Rafanelli C. Psychometric theory is an obstacle to the progress of clinical research. *Psychother Psychosom*. 2004; **73**: 145-148.

9. Wittchen HU, Pfister H. *DIA-X-Interviews: Manual für Screening Verfahren und Interview; Interviewheft Längsschnittuntersuchung (DIA-X-Lifetime); Ergänzungsheft (DIA-X-Lifetime); Interviewheft Querschnittuntersuchung (DIA-X-12 Monate); Ergänzungsheft (DIA-X-12 Monate); PC-Programm zur Durchführung des Interviews (Längs- und Querschnittuntersuchung); Auswertungsprogramm*. Frankfurt: Swets & Zeitlinger; 1997.
10. Derogatis LR, Melisaratos N. The Brief Symptom Inventory: an introductory report. *Psychol Med*. 1983; **13**(3): 595–605.
11. Marks IM, Mathews AM. Brief standard self-rating for phobic patients. *Behav Res Ther*. 1979; **17**(3): 263–267.
12. Connor KM, Kobak KA, Churchill LE, Katzelnick D, Davidson JRT. Mini-SPIN: a brief screening assessment for generalized social anxiety disorder. *Depress Anxiety*. 2001; **14**(2): 137–140.
13. Spitzer RL, Kroenke K, Williams JBW, Lowe B. A brief measure for assessing generalized anxiety disorder—the GAD-7. *Arch Intern Med*. 2006; **166**(10): 1092–1097.
14. Knappe S, Hoyer J. Clinical assessment of anxiety disorders. In: Emmelkamp PMG, Ering T, eds. *International Handbook of Anxiety Disorders: Theory, Research and Practice*. Vol II: Wiley; In press.
15. Pilkonis PA, Choi SW, Reise SP, et al. Item banks for measuring emotional distress from the Patient-Recorded Outcomes Measurement Information System (PROMIS): depression, anxiety, and anger. *Ann Behav Med*. 2011; **41**: S145–S145.
16. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol*. 2010; **63**(11): 1179–1194.
17. Bandelow B. *Panic and Agoraphobia Scale (PAS)*. Seattle, WA: Hogrefe & Huber Publishers; 1999.
18. Wolpe J, Lang PJ. A fear survey schedule for use in behavior therapy. *Behav Res Ther*. 1964; **2**: 27–30.
19. Beck JG, Carmin CN, Henninger NJ. The utility of the fear survey schedule—III: an extended replication. *J Anxiety Disord*. 1998; **12**(3): 177–182.
20. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)*. 2007; **4**(7): 28–37.
21. Berk M, Ng F, Dodd S, et al. The validity of the CGI severity and improvement scales as measures of clinical effectiveness suitable for routine clinical use. *J Eval Clin Pract*. 2008; **14**(6): 979–983.
22. Forkmann T, Scherer A, Boecker M, et al. The clinical global impression scale and the influence of patient or staff perspective on outcome. *BMC Psychiatry*. 2011; **11**: 83.
23. Stata Statistical Software: Release 12 [computer program]. College Station, TX: StataCorp; 2011.
24. Muthen LK, Muthen BO. *Mplus User's Guide*. Vol. 6. Los Angeles, CA: Muthen & Muthen; 1998–2010.
25. Savitz DA, Olshan AF. Multiple comparisons and related issues in the interpretation of epidemiologic data. *Am J Epidemiol*. 1995; **142**(9): 904–908.
26. Muthén B. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*. 1984; **49**(1): 115–132.
27. Hu L, Bentler PM. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. *Psychol Methods*. 1998; **3**(4): 424–453.
28. Meng XL, Rosenthal R, Rubin DB. Comparing correlated correlation-coefficients. *Psychol Bull*. 1992; **111**(1): 172–175.
29. Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat Med*. 2000; **19**(9): 1141–1164.
30. Kubinger KD, Rasch D, Moder K. Zur Legende der Voraussetzungen des t-Tests für unabhängige Stichproben. *Psychologische Rundschau*. 2009; **1**(60): 26–27.
31. Regier D. Merging categorical and dimensional diagnoses of mental disorders. *Epidemiol Psychiatr Sci*. 2012; **21**(3): 267–269.
32. Balon R. Measuring anxiety: are we getting what we need? *Depress Anxiety*. 2005; **22**(1): 1–10.
33. McLean CP, Asnaani A, Litz BT, Hofmann SG. Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *J Psychiatr Res*. 2011; **45**(8): 1027–1035.
34. Reed V, Gander F, Pfister H, et al. To what degree does the Composite International Diagnostic Interview (CIDI) correctly identify DSM-IV disorders? Testing validity issues in a clinical sample. *Int J Methods Psychiatr Res*. 1998; **7**(3): 142–155.
35. Wittchen H-U, Lachner G, Wunderlich U, Pfister H. Test-retest reliability of the computerized DSM-IV version of the Munich-Composite International Diagnostic Interview (M-CIDI). *Soc Psychiatry Psychiatr Epidemiol*. 1998; **33**(11): 568–578.
36. Perini SJ, Slade T, Andrews G. Generic effectiveness measures: sensitivity to symptom change in anxiety disorders. *J Affect Disord*. 2006; **90**(2–3): 123–130.