Screening for obsessive-compulsive symptoms: Validation of the Dimensional Obsessive-Compulsive Scale - English and German Short Forms

Franziska Kühne [a,*], Tatjana Paunov [a], Jonathan S. Abramowitz [b], Jakob Fink-Lamotte [c], Bjarne Hansen [d], Gerd Kvale [d], Florian Weck [a]

[a] *Department of Psychology, Clinical Psychology and Psychotherapy, University of Potsdam, Potsdam, Germany*
[b] *Department of Psychology, University of North Carolina at Chapel Hill, USA*
[c] *Department of Clinical Psychology and Psychotherapy, University of Leipzig, Leipzig, Germany*
[d] *Haukeland University Hospital, Bergen, Norway*

ARTICLE INFO

ABSTRACT

Introduction: Obsessive-compulsive disorders (OCD) are common mental disorders, which – especially without treatment – may take a chronic course. Although they often remain undetected, there is a lack of short and valid screening instruments. With the Dimensional Obsessive-Compulsive Scale - Short Form (DOCS-SF) a screening instrument is available in Norwegian. The aims of the current analogue study were to develop and to validate English and German versions of the DOCS-SF.

Methods: In the online surveys, 304 German and 187 English-speaking students participated. In addition to the DOCS-SF, they answered a battery of established questionnaires on OCD, depression and anxiety. For determining retest reliability, 51 German students participated in another survey.

Results: Results confirmed the single factor structure of the DOCS-SF and indicated good internal consistency (α = 0.89) as well as retest reliability (r = 0.75). Concerning convergent validity, the DOCS-SF was strongly correlated with other OCD scales. Although correlations with anxiety were strong as well, correlations with measures of depression, health anxiety, disgust, and health-related well-being were significantly weaker.

Discussion: The DOCS-SF appears to be a short, reliable, and valid screening instrument for OCD. Subsequent studies should further investigate its divergent validity, and its diagnostic accuracy in clinical populations.

1. Introduction

Unwanted intrusive thoughts, images or impulses (obsessions) and repetitive behaviors or mental rituals (compulsions) are a common experience in many individuals (Rachman & de Silva, 1978; Radomsky et al., 2014). However, people differ in frequency, intensity and distress experienced by these thoughts and behaviors. Individuals diagnosed with obsessive-compulsive disorders (OCD) often suffer from intense symptoms for (many) hours each day, and experience obsessions and compulsions as highly stressful (American Psychiatric Association [APA], 2013). Although the contents of obsessions and compulsions differ, research has shown four characteristic and prevalent *symptom dimensions*, i.e., a) contamination obsessions and cleaning compulsions, b) symmetry obsessions and ordering/repeating compulsions, c)

obsessions concerning responsibility for causing harm and checking compulsions, and d) obsessions regarding individually unacceptable or tabooed thoughts regarding sexuality, religion or violence (Abramowitz et al., 2010; Bloch et al., 2008; Mataix-Cols et al., 2005; McKay et al., 2004; Schulze et al., 2018). However, the number of perceived OCD symptom dimensions is not necessarily associated with severity and suffering (Abramowitz et al., 2010). In particular, a patient might experience obsessions and compulsions only in one domain such as contamination, but severely suffer from them. Therefore, OCD measures should differentiate regarding the above-mentioned dimensions, but also cover *qualitative parameters* such as time occupied, avoidance, distress, interference with functioning, and difficulty refraining from obsessions and compulsions (Abramowitz et al., 2010).

Given its lifetime prevalence of 2.3% (Kessler et al., 2012), the basic

aspects in a stepped-care approach to treat OCD are awareness, recognition and assessment (National Institute for Health and Clinical Excellence [NICE], 2006). As OCD runs the risk of underrecognition even in specialist care, screening questions are the first important steps into formal assessment and further diagnostics (NICE, 2006; Wahl et al., 2010). Thus, measurement instruments should not only cover the heterogeneity of OC symptom dimensions, but also provide feasible and economic assessments (Overduin & Furnham, 2012). Several self-report measures have been published for the assessment of OC symptoms (Overduin & Furnham, 2012) and of adjunctive dysfunctional beliefs (Meißner et al., 2020). Still, established instruments are characterized by numerous limitations, for example by confounding severity with the number of symptoms, not covering the heterogeneity of OC symptom dimensions, referring to a unidimensional assessment of severity or disconnecting obsessions and compulsions (Abramowitz et al., 2010).

Considering the criteria of the latest Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA, 2013), the Dimensional Obsessive-Compulsive Scale (DOCS; Abramowitz et al., 2010) is a measure that addresses the above-mentioned points of criticism. During the last years, it has become a widely used severity measure with excellent psychometric properties (Abramowitz et al., 2010; Thibodeau et al., 2015); translations into various languages have been provided (e. g., Chinese: Wang et al., 2012; German: Fink-Lamotte et al., 2020; Mexican: Trevino-de la Garza et al., 2019).

Eilertsen et al. (2017) have developed an efficient short form of the scale (DOCS-SF) which is available in Norwegian. In the first part, the DOCS-SF asks for obsessions according to the four OC symptom dimensions. In the second part of the questionnaire, five items on qualitative parameters are used to provide a severity rating, which is given for all obsessions combined. The DOCS-SF is a feasible measure that showed excellent reliability ($\alpha = 0.94$), was sensitive to treatment change, showed high sensitivity (96%) and specificity (94% at a cut-off score of 16), and was thus suggested as a useful screening tool for OCD (Eilertsen et al., 2017).

In light of the advantages and limitations of the available tools, the aim of the current study was to develop and validate cross-culturally applicable English and German versions of the DOCS-SF. We translated the Norwegian DOCS-SF (Eilertsen et al., 2017), and adapted it to the English (Abramowitz et al., 2010) and German (Fink-Lamotte et al., 2020) long versions of the DOCS. We hypothesized the DOCS-SF to show a unidimensional factor structure, good internal consistency and good retest reliability. We further hypothesized strong correlations with other OCD measures (convergent validity), and discriminant validity defined by weaker correlations with measures of anxiety, depression, disgust, and health-related well-being.

2. Material and methods

2.1. Translation and adaption of the DOCS-SF

The translation from Norwegian into German language was conducted using the back-translation method following good scientific practice (Brislin, 1986). Two German students speaking Norwegian at B2/C1-level conducted the translations, which were then merged into one, and reviewed by the study team. In a second step, the wording was adapted to the English (Abramowitz et al., 2010) and German (Fink-Lamotte et al., 2020) DOCS long versions. For example, the Norwegian "bakterier og smitte" was translated into "Bakterien und Infektionen" first, but then adapted to a) the more general "Keime/germs" and b) in accordance with the long versions to "Verunreinigung/contamination". Third, examples for repetitive behaviors and unpleasant thoughts were added in order to make items more tangible and user friendly (Abramowitz et al., 2010; Fineberg et al., 2008). In contrast to the Norwegian DOCS-SF, where persons may choose for a time frame they refer their symptoms to, we decided for a default time frame (i.e., the last month) in order to enhance standardization. To reduce central tendency bias, the

response format was changed from 8-point (Eilertsen et al., 2017) and 5-point scales (Abramowitz et al., 2010) to a 6-point Likert-scale with verbal anchors at the ends (e.g., from 0 = *not at all to 5 = extremely*). Moreover, a question was added that asks for the symptom dimension perceived as the most stressful.

The final version was proofread by the research team. The first block of questions now consists of a checklist including the most empirically supported OC symptom dimensions and asks for the occurrence of those during the last month (see Appendices A and B). Second, persons are asked for the individually most stressful dimension, and indicate the associated level of distress from zero (*not at all*) to 100 (*extremely distressed*) percent. The last block of questions focuses on the most stressful dimension only, and asks for the associated time expenditure, avoidance, distress, interference with functioning and the difficulty in refraining from obsessions and compulsions. Each of the five items is rated on a 6-point Likert scale as described above. To provide a severity index, the items are summed up, whereas the sum score ranges from zero to 25. In the English version, a 7-point Likert-scale was used (e.g., from *0 = not at all to 6 = extremely*). In order to compare the English with the German sample, we transferred the English version into a 6-point scale (from *0 = not at all to 5 = extremely*) by rescaling the scores as follows: 0 = 0, 1 = 0.83, 2 = 1.66, 3 = 2.49, 4 = 3.32, 5 = 4.15, 6 = 5 (Dawes, 2007).

2.2. Participants and procedure

As the purpose of the scale is screening for OC symptoms in order to initiate more elaborate diagnostic processes subsequently, we decided to include analogue samples.

2.2.1. German sample

German-speaking participants were mainly recruited at the University of xx, Germany from July 2019 to January 2020. They took part in the online survey voluntarily, gave informed consent, and received course credit for participation. There were no prerequisites for study participation. The sample comprised 304 adults with an average age of $M = 24.86$ years ($SD = 6.93$, range 17–70). The majority were students ($n = 260$, 85.5%) and females ($n = 232$, 76.3%). They were highly educated as 69.1% ($n = 210$) indicated higher secondary education, and additionally, 28.3% ($n = 86$) held a university degree.

2.2.2. English sample

English-speaking participants were recruited from October 2019 to April 2020 at the University of xx, USA. They took part in the online survey voluntarily and gave informed consent. Since we explored from preliminary analyses on the German data that mainly persons with little symptom burden had participated so far, we decided to include only participants that indicated at least one symptom dimension in the English survey. The English sample comprised 187 adults with an average age of $M = 19.03$ years ($SD = 1.60$, range 17–28). The majority were students ($n = 185$, 98.9%), and females ($n = 112$, 59.9%). They were highly educated as 85.6% ($n = 160$) indicated higher secondary education, and additionally, 14.4% ($n = 27$) held a college degree.

2.2.3. Retest sample

Furthermore, participants were recruited at the University of xx as described above to participate in two subsequent online surveys between June and September 2020 to determine retest reliability at two weeks. The retest sample ($N = 51$) had an average age of $M = 25.67$ years ($SD = 5.29$, range 18–43). Again, $n = 41$ (80.4%) were students, and the majority was female ($n = 40$, 78.4%). There were no significant differences between the German original and the German retest samples regarding age and gender ($p > .05$).

2.3. Convergent validity measures

2.3.1. Dimensional Obsessive-Compulsive Scale (DOCS; Abramowitz et al., 2010)

The DOCS is a self-report instrument assessing the four main OCD symptom dimensions as described above. Regarding each dimension, respondents rate severity on five items (i.e., time expenditure, avoidance, distress, functional interference and difficulties with refraining). The items are rated on a 5-point Likert-scale (from 0 = *no symptoms* to 4 = *extreme symptoms*), resulting in sum scores from zero to 80. Factor analysis revealed a stable four-factor structure, and a general OCD factor (Abramowitz et al., 2010). Evidence was provided for its stability over time, its construct validity and diagnostic accuracy (Abramowitz et al., 2010; Thibodeau et al., 2015). The German translation showed excellent reliability (α ≥ 0.91), satisfactory to good construct validity and satisfactory diagnostic accuracy (Fink-Lamotte et al., 2020). In the current study, the DOCS was applied in the German sample and showed excellent internal consistency (α = 0.91).

2.3.2. Obsessive-Compulsive Inventory-revised (OCI-R; Foa et al., 2002)

The OCI-R is an 18-item questionnaire assessing distress experienced by common OCD symptoms. Items are rated on a 5-point scale (ranging from 0 = *not at all* to 4 = *very much*), and the sum scores range between zero and 72. The OCI-R includes six subscales, i.e., washing, checking, ordering, obsessing, neutralizing and hoarding. Evidence supports its six-factor structure (Gönner et al., 2008). Foa et al. (2002) report excellent retest reliability, and strong support was found for its divergent validity (Gönner et al., 2008) and its convergent validity with other OCD measures (Abramowitz & Deacon, 2006; Gönner et al., 2007). In the present study, it showed good internal consistency in both, English (α = 0.87) and German (α = 0.88), samples.

2.4. Divergent validity measures

2.4.1. Patient Health Questionnaire-2 (PHQ-2; Kroenke et al., 2009)

The PHQ-2 is an ultra-brief screening tool for depressive symptoms (Löwe et al., 2005). It consists of two items that are rated on a 4-point Likert-scale (ranging from 0 = *not at all* to 3 = *nearly every day*). The sum score ranges from zero to six. The PHQ-2 showed construct validity with measures of self-esteem, life satisfaction and risk factors of depression (Löwe et al., 2010). In the present study, its internal consistency was acceptable to good in the English (α = 0.82) and German (α = 0.78) samples.

2.4.2. Generalized Anxiety Disorder Scale-7 (GAD-7; Spitzer et al., 2006)

The GAD-7 is a 7-item screener for anxiety symptoms. Items are rated on a 4-point Likert-scale (ranging from 0 = *not at all* to 3 = *nearly every day*), and the sum score ranges from zero to 21. Evidence was provided for its one-dimensional factor structure, for construct validity (Hinz et al., 2017; Löwe et al., 2008) and for diagnostic accuracy (Spitzer et al., 2006). In the current study, its internal consistency was good in both, English (α = 0.88) and German (α = 0.85), samples.

2.4.3. Illness Attitude Scale-bodily preoccupation subscale (IAS-BP; Kellner, 1986)

The IAS-BP is the 3-item subscale of the Illness Attitude Scale focusing on bodily preoccupations. Items are rated on a 5-point Likert-scale (ranging from 0 = *never* to 4 = *mostly*). The sum score ranges from zero to 12. There is evidence for its convergent and divergent validity and its diagnostic accuracy (Höfling & Weck, 2013; Weck et al., 2010). In the present study, its internal consistency was acceptable in both, English (α = 0.80) and German (α = 0.78), samples.

2.4.4. Disgust Propensity and Sensitivity Scale-revised (DPSS-R; Fergus & Valentiner, 2009)

The DPSS-R is a 12-item measure on disgust propensity and sensitivity. Items are rated on a 5-point Likert-scale from 0 (*never*) to 4 (*always*). The sum score ranges from zero to 48. There is evidence for its reliability, convergent and incremental concurrent validity (Fergus & Valentiner, 2009). In the current study, internal consistency was good in both, English (α = 0.85) and German (α = 0.83), samples.

2.4.5. World Health Organization Well-Being Index (WHO-5; World Health Organization [WHO], 1998)

The WHO-5 is a short measure on health-related well-being. It comprises five items (e.g., "I have felt calm and relaxed.") rated on a 6-point Likert-scale (from 0 = *at no time* to 5 = *all of the time*). Raw sum scores are converted to a percentage scale. Brähler et al. (2007) provided evidence for its reliability and construct validity. In the current study, internal consistency was good in both, English (α = 0.81) and German (α = 0.85), samples.

3. Data analytic strategy

Statistical analyses were performed using IBM SPSS Statistics for Windows (2016). For confirmatory factor analysis, R (R Core Team, 2014; Rosseel, 2012) was used.

The DOCS-SF data was checked for plausibility. Since the items are based on each other, values were inspected according to logical mistakes that were made. Generally, implausible values were only corrected if they could be interpreted on the basis of the other items of an individual (see below). If this was not possible, the participant was excluded from the analysis (English sample: 5.6%, *n* = 11; German sample: 10.6%, *n* = 37). If data could be interpreted meaningfully, implausible values were corrected: The most common logical mistake was that some participants had rated a symptom dimension as the most stressful which was not mentioned as an experienced symptom dimension before (English sample: 9.6%, *n* = 19; German sample: 19.1%, *n* = 67). If this was the case, the dimension was retrospectively added and the rest of the information was kept unmodified. Logical mistakes, their frequencies and the respective modifications can be found in Appendix C. In the English sample, 85% of the DOCS-SF values (*n* = 168) did not show any plausibility problems, and 10% (*n* = 19) were kept with minor modifications. In the German sample, 60% of the DOCS-SF values (*n* = 209) did not show any plausibility problems, and 30% (*n* = 96) were kept with minor modifications. Plausibility check resulted in final sample size of 187 (English), 304 (German), and 51 (retest) participants. Single missing values in the other questionnaires were replaced with the individual mean of the respective variable.

3.1. Factor structure

To examine the hypothesis on the one-dimensional factor structure of the DOCS-SF, we conducted confirmatory factor analyses (CFA). As the English and German data were not normally distributed, CFA was conducted using robust maximum likelihood estimator (MLR; Field et al., 2012). Although the Chi^2 statistic is commonly used for examining the adequacy of model fit, it overestimates lack of fit with larger sample sizes (Bollen, 2014). Therefore, multiple complementary fit indices are recommended to evaluate model fit (Bühner, 2011). Following Hu and Bentler (1999), goodness of fit was evaluated using standard root-mean-square residual (SRMR), root-mean-square error of approximation (RMSEA), comparative fit index (CFI) and Tucker Lewis Index (TFI). Good model fit was defined as follows: SRMR ≤0.08, RMSEA ≤0.06, CFI ≥0.95 and TLI ≥0.95 (Hu & Bentler, 1999).

3.2. Reliability

For determining internal consistencies, we calculated Cronbach's α. According to Bland and Altman (1997), coefficients are interpreted as follows: 0.70-0.80 acceptable, >0.80 good, and >0.90 excellent. Regarding retest reliability, paired *t*-tests and Pearson correlations were

computed for the DOCS-SF sum scores between both measurements. Test-retest correlations were then compared to the retest coefficients of other OCD measures (Abramowitz et al., 2010; Foa et al., 2002).

3.3. Validity

Pearson correlations were computed to test the hypotheses on convergent and divergent validity. To evaluate if correlations between the DOCS-SF and the convergent measures were significantly stronger than correlations with the divergent measures, t-tests for comparing dependent correlations were conducted following Weaver and Wuensch (2013). For example, to determine whether the correlation between DOCS-SF and OCI-R was stronger than between DOCS-SF and PHQ-2, three correlation coefficients were compared with each other (i.e., r_{xy}: DOCS-SF and OCI-R; r_{xz}: DOCS-SF and PHQ-2; r_{yz}: OCI-R and PHQ-2).

4. Results

Descriptive statistics for all measures are presented in Table 1. Regarding the DOCS-SF, 63% ($n = 117$) of the English participants indicated one OCD dimension, 26% ($n = 49$) indicated to perceive two dimensions and 11% (n = 21) perceived three or more dimensions. Fourteen percent ($n = 43$) of the German participants perceived no OCD dimension, 53% ($n = 161$) perceived one OCD dimension, 23% ($n = 69$) indicated to perceive two dimensions and 10% (n = 31) perceived three or more dimensions (Appendix E). English participants perceived on average 1.53 ($SD = 0.83$, range 1–5) dimensions, whereas the German participants experienced on average 1.32 ($SD = 0.92$, range 0–5) dimensions, a difference that became significant ($t(489) = 2.55$, $p < .05$).

In comparison to the English sample, German participants were significantly older ($t(489) = -11.32$; $p < .001$) and more likely to be female ($X^2(1, N = 491) = 14.89$, $p < .001$). Participants in the English sample ($M = 40.37\%$, $SD = 24.40\%$, range 0–95) were significantly more distressed ($t(489) = 5.09$, $p < .001$) than the German participants were ($M = 23.85\%$, $SD = 24.36\%$, range 0–99). In line, the difference between the DOCS-SF means between the two samples was significant and moderate to large (see Table 1).

4.1. Factor structure and reliability

A one factor solution was indicated by the CFA on both, the English ($\chi^2(5) = 0.84$, SRMR = 0.01, RMSEA = 0.00, CFI = 0.1.00, TLI = 1.01) and the German data ($\chi^2(5) = 0.10$, SRMR = 0.02, RMSEA = 0.05, CFI = 0.99, TLI = 0.98), indicating excellent model fit.

Cronbach's α was 0.88 (English sample) and 0.89 (German sample) respectively, and both coefficients may be interpreted as good (Bland & Altman, 1997; Nunnally & Bernstein, 1984).

Concerning retest reliability at two weeks, the DOCS-SF total scores did not change significantly ($M_{t1} = 4.16$, $SD_{t1} = 5.92$; $M_{t2} = 3.73$, $SD_{t2} = 5.67$; $t(50)0.758$, $p = .452$). The Pearson correlation was $r_{t1,t2} = 0.75$ (p

< .01), indicating a high correlation. Considering the COVID-19 situation in 2020, the DOCS-SF total score at t_1 ($t(50)$-1.681, $p = .099$) did not differ significantly from the total score in the original German sample which was recruited mainly in 2019 ($M = 5.55$, $SD = 5.21$, see Table 1), whereas the total score at t_2 was significantly lower ($t(50)$-2.299, $p = .026$). Limitations regarding retest reliability are debated within the discussion section.

4.2. Convergent and divergent validity

In line with our hypothesis, all correlations with OCD-related measures were strong (i.e., $r \geq 0.53$) and significant for both samples (see Table 2). In the German sample, we additionally used the DOCS long version, whereas the correlation between the DOCS and DOCS-SF was significantly higher than the correlation between the DOCS-SF and the OCI-R ($t(301) = 3.97$, p < .001).

Following Weaver and Wuensch (2013), for each of the divergent measures (PHQ-2, GAD-7, IAS-BP, DPSS-R, WHO-5), the correlations with the DOCS-SF were tested against the correlations of the DOCS-SF with the convergent measures (DOCS, OCI-R). Results of significance testing can be found in Appendix D. In the German sample, all convergent associations were significantly stronger than the divergent associations, except for anxiety as measured by the GAD-7. In the English sample, the convergent associations were significantly stronger than the divergent associations with disgust and well-being as measured by the DPSS-R and the WHO-5. Still, the convergent associations were not significantly stronger than the divergent associations with anxiety (GAD-7), depression (PHQ-2) and health anxiety (IAS-BP).

5. Discussion

Within the current study, we developed and validated cross-culturally applicable English and German versions of the Dimensional Obsessive-Compulsive Scale - Short Forms (DOCS-SF). In accordance with our hypothesis and for both language versions, there was evidence for a unidimensional factor structure. Moreover, the DOCS-SF showed good internal consistency in both samples, which was comparable to previous data (Abramowitz et al., 2010; Eilertsen et al., 2017; Fink-Lamotte et al., 2020).

In the following sections, we will discuss the results in more detail. As hypothesized, the DOCS-SF was significantly and strongly correlated with OC symptoms as measured by the OCI-R for both language versions, indicating convergent validity. In the German sample, we additionally utilized the DOCS long version. First of all, the correlation between both DOCS versions was significantly higher than the correlation between the DOCS-SF and the OCI-R. Secondly, the correlations between each of the German DOCS versions and the other measures were very similar. Thus, we conclude that the DOCS short form measures a similar construct as the DOCS long form does.

In contrast to the results on convergent validity, results on

Table 1
Descriptive data for the German and English samples and all measures.

Measure	German sample ($n = 304$)			English sample ($n = 187$)			$t(489)$	p	d
	M	SD	range	M	SD	range			
DOCS-SF	5.55	5.21	0–20	10.87	5.73	1–23	10.57	.001	.96
DOCS	10.91	9.26	0–40	–	–	–	–	–	–
OCI-R	14.06	9.84	0–45	16.88	10.52	0–50	3.00	.003	.28
PHQ-2	1.74	1.40	0–6	2.31	1.64	0–6	4.10	.001	.38
GAD-7	5.79	4.12	0–20	6.80	4.69	0–21	2.50	.013	.23
IAS-BP	3.54	2.38	0–12	3.73	2.68	0–11	0.82	.414	.08
DPSS-R	14.39	6.76	0–39	14.91	7.93	0–37	0.77	.439	.07
WHO-5	49.71	18.70	8–96	52.72	17.80	12–100	1.76	.078	.17

Notes. DOCS-SF Dimensional Obsessive-Compulsive Scale Short Form, DOCS Dimensional Obsessive-Compulsive Scale, OCI-R Obsessive-Compulsive Inventory – Revised, PHQ-2 Patient Health Questionnaire-2, GAD-7 Generalized Anxiety Disorder Scale-7, IAS-BP Illness Attitudes Scale – Subscale Bodily Preoccupations, DPSS-R Disgust Propensity and Sensitivity Scale – Revised, WHO-5 WHO-5 Wellbeing Index.

Table 2
Pearson correlations for convergent and divergent validity.

Measure	German sample ($n = 304$)							English sample ($n = 187$)					
	DOCS	OCI-R	PHQ-2	GAD-7	IAS-BP	DPSS-R	WHO-5	OCI-R	PHQ-2	GAD-7	IAS-BP	DPSS-R	WHO-5
DOCS-SF	.73**	.62**	.45**	.62**	.33**	.35**	-.34**	.53**	.56**	.60**	.43**	.35**	-.47**
DOCS		.75	.45**	.66**	.39**	.47**	-.35**	–	–	–	–	–	–
OCI-R			.39**	.57**	.42**	.45**	-.30**		.41**	.56**	.53**	.57**	-.34**
PHQ-2				.65**	.11	.33**	-.66**			.83**	.48**	.32**	-.58**
GAD-7					.32**	.45**	-.58**				.49**	.39**	-.60**
IAS-BP						.45**	-.13*					.37**	-.27**
DPSS-R							-.26**						-.28**

Notes. DOCS-SF Dimensional Obsessive-Compulsive Scale Short Form, DOCS Dimensional Obsessive-Compulsive Scale, OCI-R Obsessive-Compulsive Inventory – Revised, PHQ-2 Patient Health Questionnaire-2, GAD-7 Generalized Anxiety Disorder Scale-7, IAS-BP Illness Attitudes Scale – Subscale Bodily Preoccupations, DPSS-R Disgust Propensity and Sensitivity Scale – Revised, WHO-5 WHO-5 Wellbeing Index.** $p < .01$ (two-tailed), * $p < .05$ (two-tailed).

discriminant validity yielded mixed results. In the German sample, all convergent associations were significantly stronger than the divergent associations, except for general anxiety as measured by the GAD-7. In the English sample, the convergent associations were significantly stronger than the divergent associations for disgust and health-related well-being, the most distant constructs, but not for general anxiety, health anxiety and depression. In these respects, the results are very comparable to those of former studies, in which discrimination from anxiety and/or depression proved difficult too (Abramowitz et al., 2010; Eilertsen et al., 2017; Fink-Lamotte et al., 2020). On one hand, this finding is unsurprising as OCD is conceptually related to the anxiety disorders (Abramowitz, 2018). On the other, the GAD-7 is a rather general instrument that also covers symptoms of general distress (such as "Trouble relaxing" or "Feeling afraid as if something awful might happen"). As both German DOCS versions use the wordings "slightly different" (whereas in English it says "experiences might be different than the examples") in their introductions, participants in the German sample might have had more trouble distinguishing anxiety from OC symptoms. Still, bivariate correlations between OC symptoms and general anxiety were comparable between the two samples.

For measuring depressive symptoms, we used the PHQ-2, which encompasses two items on "Little interest or pleasure in doing things" and "Feeling down, depressed or hopeless". Depressive disorders are very prevalent among the general population (Kessler et al., 2012), and also around 50% of OCD patients suffer from comorbid depression (Abramowitz, 2018). Altogether, conceptual overlap is rather difficult to disentangle within a short screening instrument, and should be part of more elaborated diagnostic processes. In addition, it proved more suitable to determine discriminant validity by referring to more distant constructs, such as overall disgust, health-related well-being, and (at least partially) health anxiety.

As proposed by Eilertsen et al. (2017) we investigated the retest reliability of our short measure. Over two weeks, the correlation proved to be comparable to the previous study on the complete German version of the DOCS (Fink-Lamotte et al., 2020: $r_{tt} = 0.8$ over 3–12 weeks), but higher than the correlation over 12 weeks in the original publication (Abramowitz et al., 2010: $r_{tt} = 0.66$). Although our high retest correlation speaks for the utility of the DOCS-sf especially for screening purposes, it may also indidcate the possibility of a memory effect (Moosbrugger & Kelava, 2012) which is why future studies should investigate larger re-test periods.

In the next sections, we will highlight limitations and future research demands. During the translation process, we decided to adapt the wordings of both DOCS-SF language versions to the English and German long versions. As a consequence of the experiences made within the Norwegian study (Eilertsen et al., 2017), we further gave examples of obsessions and compulsions in order to enhance feasibility. We also predefined the time frame asked for to the last month to improve standardization. According to our results, the examples contributed to a better understanding of items which is reflected in the lower use of the 'other' category compared to the Norwegian results (Eilertsen et al.,

2017). Still, plausibility problems were evident predominantly in the German sample, and between the first and second part of DOCS-SF items. Further studies should investigate whether offering the opportunity to indicate no dimension as stressful at all might contribute to less missing and ambiguous answers.

As outlined above, another difference of the DOCS-SF as compared to the Norwegian and the long versions refers to the type of scale used. In order to reduce the possibility of a central tendency bias, we changed the response format to a 6-point Likert-scale with verbal anchors. As a limitation of the current study, the response format used in the English survey differed (i.e., 7-point scale), and was rescaled arithmetically to the intended format to enable comparability of the sample means. Whereas the English participants in the second survey were instructed to indicate at least one OCD dimension, German participants in the first survey were not. Besides, our multi-step study included convenience samples, but no patient data yet. Due to the type of recruitment, all three samples mainly comprised young female students. Therefore, further studies including different patient groups, persons with elevated OC symptoms as well as more heterogeneous (nonstudent) community samples from different age and education groups are crucial to assess the diagnostic accuracy of the measure. In doing so, subsequent studies should always include the DOCS long version to assess convergent validity.

Several reasons may have contributed to the more pronounced divergent associations (around $r = .10$) in the English sample. First, the three surveys were conducted subsequently, which is why the English sample was recruited between October 2019 (Corona pandemic not yet officially registered in the USA) and April 2020 (first Corona wave). Second, the English sample comprised more male participants. Furthermore, in reaction to the first German survey, the English participants had to indicate at least one symptom dimension in order to be included in the sample. Although the English participants seemed more affected by OC symptoms, the contents of the most stressful dimension did not differ between the two groups. While our first investigation suggests rough comparability, it remains an empirical question whether the German and the English questionnaires are equally sound screening tools for detecting OC symptoms.

In conclusion, we recommend the DOCS-SF as a reliable and valid self-report screening instrument covering the most prevalent OC symptom dimensions, but clearly suggest further research on its psychometric properties. Future studies could address the following aspects: use another instrument for the measurement of general anxiety than the GAD-7; determine discriminant validity by referring to more distant constructs, such as overall disgust or health-related well-being; include longer re-test periods; investigate re-test reliability in English-speaking samples; examine whether offering the opportunity to indicate no dimension as stressful might contribute to less missing values; investigate the comparability of different language versions; include more heterogeneous as well as patient samples, and determine the sensitivity and specificity of the measures. By proposing these research goals, we hope to stimulate more research on the psychometric properties of OC

screening measures, in order to enhance early detection and thus, adequate treatment of OCD patients."

CRediT authorship contribution statement

Franziska Kühne: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing. **Tatjana Paunov:** Methodology, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Jonathan S. Abramowitz:** Investigation, Resources, Writing - review & editing, Supervision. **Jakob Fink-Lamotte:** Writing - review & editing, Supervision. **Bjarne Hansen:** Writing - review & editing, Supervision. **Gerd Kvale:** Writing - review & editing, Supervision. **Florian Weck:** Resources, Writing - review & editing, Project administration.

Declaration of competing interest

The authors have no conflicts of interest to declare.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jocrd.2021.100625.

Author disclosure.

Role of funding sources

Ethical approval

The study was approved by the Ethics review boards of the University of North Carolina at Chapel Hill and of the University of Potsdam.

References

Abramowitz, J. S. (2018). Presidential address: Are the obsessive-compulsive related disorders related to obsessive-compulsive disorder? A critical look at DSM-5's new category. *Behavior Therapy, 49*(1), 1–11.

Abramowitz, J. S., Deacon, B. J., Olatunji, B. O., Wheaton, M. G., Berman, N. C., Losardo, D., Timpano, K. R., McGrath, P. B., Riemann, B. C., Adams, T., Björgvinsson, T., Storch, E. A., & Hale, L. R. (2010). Assessment of obsessive-compulsive symptom dimensions: Development and evaluation of the dimensional obsessive-compulsive scale. *Psychological Assessment, 22*, 180–198. https://doi.org/10.1037/a0018260

Abramowitz, J. S., & Deacon, B. J. (2006). Psychometric properties and construct validity of the Obsessive-Compulsive Inventory – revised: Replication and extension with a clinical sample. *Anxiety Disorders, 20*, 1016–1035.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Bland, J. M., & Altman, D. G. (1997). Cronbach's alpha. *BMJ, 314*, 572. https://doi.org/10.1136/bmj.314.7080.572

Bloch, M. H., Landeros-Weisenberger, A., Rosario, M. C., Pittenger, C., & Leckman, J. F. (2008). Meta-analysis of the symptom structure of obsessive-compulsive disorder. *American Journal of Psychiatry, 165*, 1532–1542. https://doi.org/10.1176/appi.ajp.2008.08020320

Bollen, K. A. (2014). *Structural equations with latent variables* (John Wiley & Sons).

Brähler, E., Mühlan, H., Albani, C., & Schmidt, S. (2007). Teststatistische Prüfung und Normierung der deutschen Versionen des EUROHIS-QOL Lebensqualität-Index und des WHO-5 Wohlbefindens-Index. [Testing and standardization of the German version of the EUROHIS-QOL and WHO-5 quality of life-indices]. *Diagnostica, 53*, 83–96. https://doi.org/10.1026/0012-1924.53.2.83

Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner, & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137–164). Sage.

Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion. [Introduction to test and questionnaire construction]*. Pearson.

Dawes, J. (2007). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research, 50*(1), 61–104.

Eilertsen, T., Hansen, B., Kvale, G., Abramowitz, J. S., Holm, S. E. H., & Solem, S. (2017). The dimensional obsessive-compulsive scale: Development and validation of a short form (DOCS-SF). *Frontiers in Psychology, 8*, 1503. https://doi.org/10.3389/fpsyg.2017.01503

Fergus, T. A., & Valentiner, D. P. (2009). The disgust propensity and sensitivity scale – revised: An examination of a reduced item version. *Journal of Anxiety Disorders, 23*, 703–710.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage.

Fineberg, N. A., Krishnaiah, R. B., Moberg, J., & O'Doherty, C. (2008). Clinical screening for obsessive-compulsive and related disorders. *Israel Journal of Psychiatry and Related Sciences, 45*(3), 151–163.

Fink-Lamotte, J., Jahn, I., Stierle, C., Kühne, F., Lincoln, T., Stengler, K., & Exner, C. (2020). Die Validierung der Dimensional-Obsessive Compulsive Scale (DOCS) an einer deutschsprachigen Stichprobe [Validation of the Dimensional-Obsessive Compulsive Scale (DOCS) in a German sample]. *Verhaltenstherapie*. https://doi.org/10.1159/000510093

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment, 14*(4), 485.

Gönner, S., Leonhart, R., & Ecker, W. (2007). Das zwangsinventar OCI-R - die deutsche version des obsessive-compulsive inventory-revised - ein kurzes selbstbeurteilungsinstrument zur mehrdimensionalen messung von Zwangssymptomen. [The German version of the obsessive-compulsive inventory-revised: A brief self-report measure for the multidimensional assessment of obsessive-compulsive symptoms]. *Psychotherapie, Psychosomatik, Medizinische Psychologie, 57*(9–10), 395–404. https://doi.org/10.1055/s−2007−970894

Gönner, S., Leonhart, R., & Ecker, W. (2008). The Obsessive-Compulsive Inventory-Revised (OCI-R): Validation of the German version in a sample of patients with OCD, anxiety disorders, and depressive disorders. *Journal of Anxiety Disorders, 22*(4), 734–749. https://doi.org/10.1016/j.janxdis.2007.07.007

Hinz, A., Klein, A. M., Brähler, E., Glaesmer, H., Luck, T., Riedel-Heller, S. G., Wirkner, K., & Hilbert, A. (2017). Psychometric evaluation of the Generalized Anxiety Disorder Screener GAD-7, based on a large German general population sample. *Journal of Affective Disorders, 210*, 338–344. https://doi.org/10.1016/j.jad.2016.12.012

Höfling, V., & Weck, F. (2013). Assessing bodily preoccupations is sufficient: Clinically effective screening for hypochondriasis. *Journal of Psychosomatic Research, 75*(6), 526–531. https://doi.org/10.1016/j.jpsychores.2013.10.011

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

IBM SPSS Statistics for Windows, Version 24.0 [Computer software] (2016). (IBM Corporation).

Kellner, R. (1986). *Somatization and hypochondriasis* (Praeger).

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H. U. (2012). Twelve- month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International Journal of Methods in Psychiatric Research, 21*(3), 169–184.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: The PHQ-4. *Psychosomatics, 50*(6), 613–621. https://doi.org/10.1176/appi.psy.50.6.613

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. *Medical Care, 46*(3), 266–274. https://doi.org/10.1097/MLR. 0b013e318160d093

Löwe, B., Kroenke, K., & Gräfe, K. (2005). Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *Journal of Psychosomatic Research, 58*(2), 163–171. https://doi.org/10.1016/j.jpsychores.2004.09.006

Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., Schneider, A., & Brähler, E. (2010). A 4-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders, 122*(1–2), 86–95. https://doi.org/10.1016/j.jad.2009.06.019

Mataix-Cols, D., do Rosario-Campos, M. C., & Leckman, J. F. (2005). A multidimensional model of obsessive-compulsive disorder. *American Journal of Psychiatry, 162*(2), 228–238.

McKay, D., Abramowitz, J. S., Calamari, J. E., Kyrios, M., Radomsky, A., Sookman, D., Taylor, S., & Wilhelm, S. (2004). A critical evaluation of obsessive-compulsive disorder subtypes: Symptoms versus mechanisms. *Clinical Psychology Review, 24*(3), 283–313. https://doi.org/10.1016/j.cpr.2004.04.003

Meißner, C., Weck, F., & Kühne, F. (2020). Screening dysfunktionaler Überzeugungen bei Zwangsstörungen. [Screening for dysfunctional beliefs in obsessive-compulsive disorders. *Psychotherapeut*, 1–8.

Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion. [Test theory and questionnaire construction]* (Springer).

National Institute for Health and Care Excellence. (2016). Obsessive-compulsive disorder: Core interventions in the treatment of obsessive-compulsive disorder and body dysmorphic disorder (NICE Guideline Number 31) https://www.nice.org.uk/guidance/cg31/evidence/full-guideline-pdf-194883373.

Overduin, M. K., & Furnham, A. (2012). Assessing obsessive-compulsive disorder (OCD): A review of self-report measures. *Journal of Obsessive-Compulsive and Related Disorders, 1*(4), 312–324. https://doi.org/10.1016/j.jocrd.2012.08.001

R: A language and environment for statistical computing [Computer software]. (2014). R Foundation for Statistical Computing. http://www.R-project.org/.

Rachman, S., & de Silva, P. (1978). Abnormal and normal obsessions. *Behaviour Research and Therapy, 16*(4), 233–248. https://doi.org/10.1016/0005-7967(78)90022-0

Radomsky, A. S., Alcolado, G. M., Abramowitz, J. S., Alonso, P., Belloch, A., Bouvard, M., Clark, D. A., Coles, M. E., Doron, G., Fernandez-Alvarez, H., Garcia-Soriano, G., Ghisi, M., Gomez, B., Inozu, M., Moulding, R., Shams, G., Sica, C., Simos, G., & Wong, W. (2014). Part 1—you can run but you can't hide: Intrusive thoughts on six continents. *Journal of Obsessive-Compulsive and Related Disorders, 3*(3), 269–279.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. http://www.jstatsoft.org/v48/i02.

Schulze, D., Kathmann, N., & Reuter, B. (2018). Getting it just right: A reevaluation of OCD symptom dimensions integrating traditional and bayesian approaches. *Journal of Anxiety Disorders, 56*, 63–73. https://doi.org/10.1016/j.janxdis.2018.04.003

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166* (10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Thibodeau, M. A., Leonard, R. C., Abramowitz, J. S., & Riemann, B. C. (2015). Secondary psychometric examination of the Dimensional Obsessive-Compulsive Scale: Classical testing, item response theory, and differential item functioning. *Assessment, 22*(6), 681–689. https://doi.org/10.1177/1073191114559123

Trevino-de la Garza, B., Berman, N., Fisak, B., Ruvalcaba-Romero, N., & Gallegos-Guajardo, J. (2019). Validation of the dimensional obsessive-compulsive scale for Mexican population. *Journal of Obsessive-Compulsive and Related Disorders, 21*, 13–17.

Wahl, K., Kordon, A., Kuelz, K. A., Voderholzer, U., Hohagen, F., & Zurowski, B. (2010). Obsessive-compulsive disorder (OCD) is still an unrecognised disorder: A study on the recognition of OCD in psychiatric outpatients. *European Psychiatry: The Journal of the Association of European Psychiatrists, 25*(7), 374–377. https://doi.org/10.1016/j.eurpsy.2009.12.003

Wang, C., Wang, J., Tang, T., Zhang, S., & Bjorgvinsson, T. (2012). Psychometric validation of dimensional obsessive-compulsive scale (DOCS) in Chinese college students. *Chinese Journal of Clinical Psychology, 20*, 315–319. https://doi.org/10.16128/j.cnki.1005-3611.2012.03.013

Weaver, B., & Wuensch, K. L. (2013). SPSS and SAS programs for comparing Pearson correlations and OLS regression coefficients. *Behavior Research Methods, 45*(3), 880–895. https://doi.org/10.3758/s13428-012-0289-7

Weck, F., Bleichhardt, G., & Hiller, W. (2010). Screening for hypochondriasis with the illness attitude scales. *Journal of Personality Assessment, 92*(3), 260–268. https://doi.org/10.1080/00223891003670216

World Health Organization. (1998). *The use of well-being measures in primary health care - the DepCare project*.