Received: 1 May 2021 | Revised: 9 February 2022 | Accepted: 9 February 2022

DOI: 10.1002/eat.23694

ORIGINAL ARTICLE

International Journal of EATING DISORDERS WILEY

Psychometric properties of the Compulsive Exercise Test in a large sample of female adolescent and adult inpatients with anorexia nervosa and bulimia nervosa

Sandra Schlegl PhD[1,2] | Larissa Vierl MSc[1] | David R. Kolar PhD[1] | Nina Dittmer PhD[2] | Ulrich Voderholzer MD[1,2,3]

[1]Department of Psychiatry and Psychotherapy, University Hospital, Ludwig-Maximilians-University (LMU), Munich, Germany

[2]Schoen Clinic Roseneck, Prien am Chiemsee, Germany

[3]Department of Psychiatry and Psychotherapy, University Hospital, Freiburg, Germany

Correspondence
Sandra Schlegl, PhD, University Hospital, LMU Munich, Department of Psychiatry and Psychotherapy, Nussbaumstrasse 7, 80336 München, Germany.
Email: sandra.schlegl@med.uni-muenchen.de

Action Editor: Ruth Striegel Weissman

Abstract

Objective: The Compulsive Exercise Test (CET) was developed to assess compulsive exercise in patients with eating disorders (EDs), but originally validated in a non-clinical sample, and psychometric properties were only investigated in small clinical samples. Therefore, the aim of this study was to examine its psychometric properties in a large clinical sample of adolescent and adult inpatients with anorexia nervosa and bulimia nervosa.

Method: A sample of 2,535 German female inpatients with EDs completed the CET and other instruments at admission and discharge. Factor structure (confirmatory [CFA] and exploratory factor analyses [EFA]), internal consistency and construct validity, measurement invariance across age and diagnostic groups, group comparisons of means, as well as sensitivity to change during treatment were assessed.

Results: The CET showed high internal consistency, very good construct validity, and sensitivity to change. CFA indicated a better fit of four-factor and three-factor solutions compared to the original five-factor model. However, subsequent EFA identified an optimum for a five-factor model. Only three subscales were satisfactorily invariant to measurement, but not the CET total score. Only small differences in scores between patient groups were observed.

Discussion: Results support internal consistency, construct validity, and sensitivity to change, whereas factor structure remains inconclusive, questioning the theoretical basis of the CET. There is limited support for using the lack of enjoyment subscale, and only moderate support for using the rigidity subscale in patients with EDs. It is recommended to further explore and/or revise the original CET, including investigation in other samples, for example, male samples.

KEYWORDS

anorexia nervosa, bulimia nervosa, Compulsive Exercise Test, factor analysis, inpatient, psychometric properties, validation

1 | INTRODUCTION

Compulsive exercise is a common symptom in patients with eating disorders (EDs). Depending on the sample and assessment methods, its prevalence ranges from 23.1% to 81% in individuals with anorexia nervosa (AN; Davis, 1997; Dittmer et al., 2020; Stiles-Shields, Bamford, Lock, & Le Grange, 2015) and from 20% to 66.7% in patients with bulimia nervosa (BN; Binford & Le Grange, 2005; Davis, 1997; Shroff et al., 2006). Compulsive exercise associates with a higher severity of ED symptomatology (Monell, Levallius, Forsen Mantilla, & Birgegard, 2018), worse therapy outcomes (Stiles-Shields et al., 2015), longer inpatient stays (Bratland-Sanda et al., 2010; Solenberger, 2001), higher rates of inpatient dropout (El Ghoch et al., 2013), and a higher risk of relapse/chronification (Carter, Blackmore, Sutandar-Pinnock, & Woodside, 2004).

According to the cognitive-behavioral maintenance model of compulsive exercise by Meyer, Taranis, Goodwin, and Haycraft (2011), patients with EDs engage in compulsive exercise to control weight and shape, to improve their mood, and to prevent or regulate negative affect. Furthermore, perfectionism and rigidity were suggested as maintaining factors. Based on this model, Taranis, Touyz, and Meyer (2011) developed a questionnaire for assessing compulsive exercise, the Compulsive Exercise Test (CET) that includes 24 items answered on a 6-point Likert scale. A principal component analysis (PCA) in a nonclinical sample of 367 young female exercisers (from the United Kingdom and Australia) revealed five components: avoidance and rule-driven behavior, weight control exercise, mood improvement, lack of exercise enjoyment, and exercise rigidity. The CET demonstrated good psychometric properties with high internal consistency ($\alpha = .85$), a mean item-total correlation of .48 and good-to-excellent concurrent and convergent validity. However, this dimensional structure of compulsive exercise does not map directly onto clinical samples, as a recent review on the psychometric properties of exercise assessments suggests. Evaluating nine studies that analyzed the validity of the CET, it concluded that "the factor structure […] was not confirmed in the majority of the studies" and that "further research is needed to confirm a factor structure and validate the […] in more diverse clinical samples" (Harris, Hay, & Touyz, 2020). For example, Swenne (2016), Limburg et al. (2021), and Plateau et al. (2014) discussed whether four factors, three factors, or a single factor better explained the factorial structure of the questionnaire. The identification of different latent factor structures for the CET might indicate that it does not fully capture all latent constructs proposed by Meyer et al. (2011), that additional latent constructs impact compulsive exercise in EDs more than in the general population, or that the instrument is not invariant between, for example, clinical and nonclinical populations.

Besides true differences in latent constructs underlying compulsive exercise between nonclinical samples and patients with EDs, several limitations of the existing research on the CET might have contributed to divergent findings in previous studies. For example, psychometric properties of the CET were mostly examined in samples with less than optimal size (Formby, Watson, Hilyard, Martin, & Egan, 2014; Sauchelli et al., 2016; Vrabel & Bratland-Sanda, 2019). In addition, in psychometric studies of the CET with clinical samples, patients with Eating Disorders Not Otherwise Specified (EDNOS) often dominated (Formby et al., 2014; Sauchelli et al., 2016; Swenne, 2016).

The few studies comparing compulsive exercise between patients with AN and BN found mixed results: Sauchelli et al. (2016) found significantly higher CET total scores and most subscale scores (except exercise rigidity) in patients with BN ($n = 56$) compared to individuals with AN ($n = 40$ Spanish participants, 80% female, adults). Cunningham, Pearman, and Brewerton (2016) also found that patients with BN ($n = 18$) showed higher CET total scores than those with AN ($n = 6$) (U.S. participants, 58% female, adults). In contrast, we did not find differences between individuals with AN ($n = 151$) and patients with BN ($n = 75$; participants: from Germany, 100% female, 69% adults), neither for the CET total score nor for the subscales (Schlegl, Dittmer, Hoffmann, & Voderholzer, 2018). We are not aware of any study that has compared adolescents and adults. However, it is important to establish whether the CET is measurement invariant between age and diagnostic groups before further investigating group differences.

Finally, no study has investigated the sensitivity to change of the CET, that is, its ability to show change after an intervention or across time, and whether its sensitivity is similar in different age or diagnostic groups. Since compulsive exercise is a concerning ED symptom, whose trajectory should be tracked during treatment, it is essential to assess to what extent change can be captured using the CET.

Thus, the aims of our study were (a) to examine the factor structure of a German translation of the CET in a large clinical sample ($N = 2{,}535$), (b) to assess the internal consistency and construct validity of the measure, (c) to explore measurement invariance of the CET across age and diagnostic groups, (d) to compare the CET scores of different diagnostic and age groups, and (e) to investigate sensitivity to change during treatment.

2 | METHODS

2.1 | Participants and procedure

We investigated a sample of female inpatients hospitalized at Schoen Clinic Roseneck in Prien am Chiemsee, Germany, between 2014 and 2018. Inclusion criteria were a diagnosis of AN or BN according to ICD-10 (F50.0–F50.3). Patients were diagnosed by experienced clinicians from the specialized ED unit during a standard intake interview using the International Diagnostic Checklists (Hiller, Zaudig, & Mombour, 1997). Furthermore, a minimum age of 12 years was mandatory both for treatment and for study inclusion. Only data of first admissions to Schoen Clinic Roseneck were used. Patients' body mass index (BMI) needed to be <18.5 for AN and <30 for BN (to exclude outliers and potential confounds related to obesity). Exclusion criteria were current drug/alcohol/medication abuse, acute suicidal tendencies, psychotic symptoms, or a severe life-threatening somatic disorder, reflecting treatment admission criteria. In total, we included 2,535 female inpatients ($n = 821$ adolescents with AN, $n = 991$ adults with AN, $n = 250$ adolescents with BN, and $n = 473$ adults with BN).

Measures were completed at admission and discharge. Sociodemographic and clinical characteristics (e.g., duration of inpatient treatment and duration of ED) were available from medical charts. All patients received a multimodal inpatient treatment program based on cognitive-behavioral therapy (CBT) with group and individual psychotherapy. For a more detailed description of our treatment program, see Schlegl, Quadflieg, Lowe, Cuntz, and Voderholzer (2014).

2.2 | Measures

2.2.1 | Compulsive Exercise Test

The Compulsive Exercise Test (CET; Taranis et al., 2011) is a self-report inventory assessing aspects of compulsive exercise on five subscales: avoidance and rule-driven behavior, weight control exercise, mood improvement, lack of exercise enjoyment, and exercise rigidity. It consists of 24 items answered on a 6-point Likert scale from 0 (never true) to 5 (always true). Subscales are summed to obtain a CET total score. The CET was translated into German by our research group, applying a translation–backtranslation approach combined with a pretest procedure (Brislin, 1970). First, it was translated into German by two researchers (M.Sc. and PhD level, more than 10 years of experience in ED research) separately. The derived versions were discussed by the translators and piloted in $N = 10$ patients with EDs who gave feedback on the measurement. Second, an English native speaker performed a back translation and compared the result to the original version. The final version was then established through consensus building.

2.2.2 | Commitment to Exercise Scale

The Commitment to Exercise Scale (CES; Davis, Brewer, & Ratusny, 1993; German version: Zeeck et al., 2017) is an 8-item self-rating scale for the assessment of compulsive exercise in patients with EDs. It addresses two core aspects: (a) obligatory exercise implies the strict adherence to a regular and clearly structured exercise routine and (b) pathological exercise refers to the physical or psychological burden of exercising. According to Thome and Espelage (2007), we used a 4-point Likert scale from 1 (never) to 4 (always) instead of the original answering format (visual analog scale). McDonald's Omega for the CES total score was 0.93.

2.2.3 | Eating Disorder Examination-Questionnaire

The Eating Disorder Examination-Questionnaire (EDE-Q; Fairburn & Beglin, 1994; German version: Hilbert & Tuschen-Caffier, 2016) is a self-rating instrument to assess ED cognitions and behaviors experienced during the previous 28 days with 28 items on a 7-point Likert scale from 0 to 6. In this study, the EDE-Q global score was used. Furthermore, the EDE-Q item 18 was used to investigate convergent validity ("Over the past 28 days, how many times have you exercised in a driven or compulsive way as a means of controlling your weight, shape or amount of fat, or to burn off calories?"). McDonald's Omega for the EDE-Q global score was 0.94.

2.2.4 | Eating Disorder Inventory-2

The Eating Disorder Inventory-2 (EDI-2; Garner, 1991; German version: Paul & Thiel, 2005) is a multidimensional self-report questionnaire to assess the specific psychopathology of patients with EDs. It consists of 11 scales with 91 items answered on a 6-point scale from 1 (never) to 6 (always). McDonald's Omega for the EDI-2 total score for this sample was 0.96.

2.2.5 | Beck Depression Inventory-II

The Beck Depression Inventory-II (BDI-II; Beck, Steer, & Brown, 1996; German version: Hautzinger, Keller, & Kühner, 2009) is a self-rating instrument to assess the severity of depressive symptoms. The 21 items can be rated on a 4-point scale from 0 to 3 regarding their occurrence and intensity during the last 2 weeks. McDonald's Omega for the BDI-II total score was 0.91.

2.2.6 | Brief Symptom Inventory

The Brief Symptom Inventory (BSI; Derogatis, 1993; German version: Franke, 2000) is a short version of the symptom checklist (Derogatis, 1979) that assesses the current general psychological distress of patients throughout the last week on the basis of 53 items belonging to nine subscales. Answers are given on a 5-point scale, ranging from 0 (not at all) to 4 (extremely). McDonald's Omega for the BSI global severity index (BSI GSI) for this sample was .96.

2.3 | Data analyses

IBM SPSS 25 (IBM Corporation, 2017) was used for most analyses.

2.3.1 | Confirmatory factor analysis (CFA)

CFA was performed using IBM AMOS 26 (Arbuckle, 2019), using maximum likelihood estimation. CFA was conducted to assess the factorial validity of the original five-factor model (Taranis et al., 2011), alternative four- and three-factor models, and one-factor models of the CET (Limburg et al., 2021; Plateau et al., 2014; Swenne, 2016). Item–participant ratio was >1:100. Multiple goodness-of-fit indices were utilized, including a) the root mean square error of approximation (RMSEA), with values <0.06 indicating a good fit and values between 0.08 and 1 a mediocre fit; (b) Tucker–Lewis index (TLI); (c) the comparative fit index (CFI), all with values >0.90 (ideally >0.95) indicating an acceptable fit of the data; and (d) the standardized root mean square residual (SRMR) with

values <0.08 as acceptable fit and <0.05 as good fit (Hooper, Coughlan, & Mullen, 2008; Hu, & Bentler, 1999). In addition, the Akaike information criterion (AIC; Burnham & Anderson, 2004) was used to compare the fit of the nonnested five-, four-, and three-factor models. Furthermore, factor loadings over 0.40 are considered appropriate (Ford, MacCallum, & Tait, 1986).

2.3.2 | Exploratory factor analysis

As CFA did not yield a sufficient model fit, we conducted parallel analysis (Horn, 1965) to determine the number of factors to retain in exploratory factor analysis (EFA). We used a common factor analysis approach with maximum likelihood as the extraction method and oblimin rotation to fit latent factors to the same data set. Explained variance (ratio between the variance of the factor and the total variance) and corrected item-total correlations (r^{it}) (correlations between each item and the total score that excludes that item) were calculated. Kline (1993) recommends deleting questionnaire items with $r^{it} < .3$. There is debate on whether CFA and EFA should be conducted on the same sample. However, this is of greater concern when CFA is conducted *after* EFA, as any CFA with prespecified factor correlations and/or loadings is a possible solution of EFA. In our case, we decided to conduct EFA only after published factor models failed to fully explain our data in CFA (Schmitt, 2011). Furthermore, McDonald's Omega was calculated as a measure of internal consistency.

2.3.3 | Measurement invariance across diagnosis and age

To test for measurement invariance of the CET across diagnosis and age group, we fitted multigroup CFAs with increasing equality constraints and reported model fit indices (Byrne, 2004). Configural invariance (assuming that number of factors and pattern of loadings are identical for both groups), metric invariance (assuming equal factor loadings), and scalar invariance (assuming equal factor loadings and equal item intercepts) were investigated. CFI, TLI, and RSMEA were reported as goodness-of-fit indices.

2.3.4 | Construct validity

To determine convergent validity, Pearson correlations of the CET with the CES and the EDE-Q item 18 were calculated and for the exploration of the discriminant validity, Pearson correlations with the EDE-Q, the EDI-2, the BDI-II, and the BSI.

2.3.5 | Group comparisons

We calculated multivariate analyses of variance (MANOVAs) to compare CET total score and CET subscales between diagnostic and age groups and report Cohen's *d* to quantify differences.

2.3.6 | Sensitivity to change

To assess sensitivity to change in the CET total score as well as the CET subscales during inpatient treatment, we followed the methods used by Machado, Grilo, Rodrigues, Vaz, and Crosby (2020) and computed standardized change scores (SCS) (difference between an individual's admission and discharge CET total score and subscales, divided by the *SD* value of the entire sample at admission). We followed with MANOVAs where SCS were entered as the dependent variable and age-by-diagnosis group as within-subject factor.

3 | RESULTS

3.1 | Sample description

Participants were 2,535 female inpatients with EDs ($n = 821$ adolescents with AN, $n = 991$ adults with AN, $n = 250$ adolescents with BN, and $n = 473$ adults with BN). For a more detailed sample description, see Table 1.

3.2 | Confirmatory factor analysis

We used confirmatory factor analysis (CFA) to examine the fit of the previously suggested five-factor, four-factor, and three-factors models in the total sample (Table 2). According to our multiple criteria for goodness of fit, all models were found to differ significantly from the observed data (for all $p < .001$), suggesting a poor fit and failed to reach the RMSEA threshold (all values > .08). Regarding CFI, the four-factor and the three-factor models had reasonable fits. Multifactorial models showed consistently smaller AIC values compared to one-factor models. For all models, factor loadings were appropriate for all items. In the five-factor model, the CET subscales were correlated with each other, except for the CET subscale lack of enjoyment which showed no significant correlation with weight control exercise ($r = .021$, $p = .350$) and negative correlations with the other subscales ($-.545 \leq r \leq -.169$).

3.3 | Exploratory factor analysis

Since no model showed acceptable fit in the CFAs, we conducted EFA to further explore the factor structure. This approach has also been used in previous articles, where CFA fit criteria were not met (Darcy, Hardy, Crosby, Lock, & Peebles, 2013; Plateau et al., 2014). Parallel analysis suggested a five-factor solution. The obtained five-factor structure of the CET determined by applying EFA with subsequent oblimin rotation is presented in Table 3. Then, 76.1% of the total variance were explained and the same items as in the original CET study loaded on the same factors. Main factor loadings ranged from 0.39 to 0.89. One item (item 15) cross-loaded onto two factors with a difference of loadings of less than 0.10. Sufficient corrected item-total

TABLE 1 Sample description for the total sample and the four subgroups

	Total sample ($N = 2,535$)	Adolescents with AN ($n = 821$)	Adults with AN ($n = 991$)	Adolescents with BN ($n = 250$)	Adults with BN ($n = 473$)
Age (years)	22.32 (9.44)	15.53 (1.22)	26.82 (9.86)	16.01 (1.01)	28.00 (9.76)
M (*SD*) range	12–73	12–17	18–73	13–17	18–61
BMI (kg/m^2)	16.83 (3.73)	15.06 (1.54)	14.80 (2.06)	21.27 (2.93)	21.80 (2.79)
M (*SD*) range	8.60–29.90	10.70–18.50	8.60–18.50	18.00–29.90	18.03–29.88
BMI percentile					
M (*SD*)	14.42 (25.96)	2.42 (4.66)		53.83 (28.11)	
For adolescents only range	0–98.90	0–38.20		11.50–98.90	
Treatment duration (days)	85.14 (44.60)	92.77 (43.53)	89.66 (48.89)	80.97 (43.83)	64.63 (28.03)
M (*SD*) range	1–316	2–250	1–295	4–316	1–183

Abbreviations: AN, anorexia nervosa; BMI, body mass index; BN, bulimia nervosa.

TABLE 2 Fit statistics for previously suggested factor models of the Compulsive Exercise Test (CET) for the total sample ($N = 2,535$)

Model	CFI	TLI	RMSEA [95% CI]	AIC	SRMR[a]
Five factors (24 items; Taranis et al., 2011)	0.888	0.861	0.096 [0.094–0.098]	6,094.688	0.1012
One factor (24 items; Taranis et al., 2011)	0.709	0.653	0.152 [0.150–0.154]	15,145.905	0.1214
Four factors (21 items; Limburg et al., 2021)	0.903	0.877	0.099 [0.96–.101]	4,829,569	0.0772
One factor (21 items; Limburg et al., 2021)	0.768	0.716	0.150 [0.148–0.153]	1,118.181	0.0938
Four factors (21 items; Swenne, 2016)	0.876	0.845	0.107 [0.104–0.109]	5,640.167	0.1045
One factor (21 items; Swenne, 2016)	0.876	0.845	0.107 [0.104–0.109]	1,210	0.1210
Three factors (15 items; Plateau et al., 2014)	0.912	0.879	0.116 [0.113–0.120]	3,162.535	0.0756
One factor (15 items; Plateau et al., 2014)	0.751	0.668	0.193 [0.189–0.196]	8,643.683	0.1075

Note: Reference values for goodness-of-fit indices: SRMR ≤ 0.08 acceptable fit, SRMR ≤ 0.05 good fit, RMSEA < 0.06 good fit, 0.10 ≤ RMSEA ≥ 0.08 mediocre fit, TLI/CFI < 0.90 bad fit, 0.90 ≥ TLI/CFI < 0.95 good fit, and TLI/CFI ≥ 0.95 very good fit.
Abbreviations: AIC, Akaike information criterion; CFI, comparative fit index; RMSEA, root mean square error of approximation; SRMR, standardized root mean squared residual; TLI, Tucker–Lewis index.
[a]Calculated with imputed data.

correlations (r^{it}) existed for 21 of the 24 items. The three items that did not correlate sufficiently all belong to the CET subscale lack of exercise enjoyment. See also Figure S1 for the path diagram of our EFA model.

3.4 | Measurement invariance

To investigate measurement invariance of the CET total score across age, we fitted a multigroup CFA with age group (adolescents vs. adults) as grouping variable. Model fit indices (CFI = 0.738, TLI = 0.708, RMSEA = 0.104) and Chi-squared test of the unconstrained model (no equality constraints; $\chi^2[494] = 11,227.00$, $p < .001$) indicated that configural measurement invariance across age cannot be assumed. Similar results were obtained for AN and BN groups, where model fit indices and Chi-squared test results also did not support configural measurement invariance (CFI = 0.740, TLI = 0.710, RMSEA = 0.103; $\chi^2[494] = 11,110.76$, $p < .001$). However, when measurement invariance analyses were performed separately for CET subscales with more than three items (at least three items were needed due to degrees of freedom), weight control exercise and mood improvement showed good and avoidance and rule-driven behavior acceptable measurement invariance. Details are shown in Table S1.

3.5 | Internal consistency

Internal consistency of the CET total score and all subscales was good to excellent, ranging from McDonald's $\omega = 0.80$ to $\omega = 0.95$ (see also Table 3).

3.6 | Construct validity

Table 4 presents Pearson correlations of the CET total score, subscales, and related measurements. We found high correlations

TABLE 3 Exploratory factor analysis of the Compulsive Exercise Test (CET) including factor loadings and corrected item-total correlations (r^{it}) for the total sample ($N = 2{,}535$)

Factor	Item	Rotated factor loading					r^{it}
		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	
Factor 1: Avoidance and rule-driven behavior	9: If I cannot exercise I feel low or depressed.	**0.812**	−0.059	−0.077	0.126	−0.003	.735
	10: I feel extremely guilty if I miss an exercise session.	**0.567**	0.050	0.188	−0.017	0.251	.859
	11: I usually continue to exercise despite injury or illness, unless I am very ill or too injured.	**0.523**	−0.041	0.087	−0.113	0.299	.708
	15: If I miss an exercise session, I will try and make up for it when I next exercise.	**0.393**	0.009	0.286	−0.049	0.318	.807
	16: If I cannot exercise I feel agitated and/or irritable.	**0.841**	−0.016	−0.066	0.109	0.010	.778
	20: If I cannot exercise I feel angry and/or frustrated.	**0.868**	−0.012	−0.031	0.079	0.001	.803
	22: I feel like I've let myself down if I miss an exercise session.	**0.534**	0.122	0.265	0.017	0.131	.816
	23: If I cannot exercise I feel anxious.	**0.740**	0.059	0.201	0.007	0.009	.836
Factor 2: Weight control exercise	2: I exercise to improve my appearance.	−0.041	−0.029	**0.685**	0.226	0.135	.729
	6: If I feel I have eaten too much, I will do more exercise.	0.208	−0.016	**0.550**	0.046	0.192	.788
	8 (R): I do not exercise to be slim.	0.003	0.033	**0.653**	−0.103	−0.056	.370
	13: I exercise to burn calories and lose weight.	−0.009	0.064	**0.827**	0.128	0.050	.732
	18: If I cannot exercise, I worry that I will gain weight.	0.293	0.025	**0.604**	0.095	0.038	.808
Factor 3: Mood improvement	1: I feel happier and/or more positive after I exercise.	−0.030	−0.077	0.149	**0.735**	0.069	.595
	4: I feel less anxious after I exercise.	0.170	0.043	0.256	**0.510**	0.081	.751
	14: I feel less stressed and/or tense after I exercise.	0.025	−0.003	0.014	**0.749**	0.018	.556
	17: Exercise improves my mood.	0.054	−0.066	−0.072	**0.818**	0.020	.506
	24: I feel less depressed or low after I exercise.	0.127	0.059	−0.041	**0.770**	−0.008	.564
Factor 4: Lack of exercise enjoyment	5: I find exercise a chore.	−0.012	**0.690**	0.064	0.086	0.077	.217
	12 (R): I enjoy exercising.	0.006	**0.633**	0.046	−0.351	−0.103	−.231
	21: I do not enjoy exercising.	0.001	**0.886**	−0.072	0.018	−0.012	.020
Factor 5: Exercise rigidity	3: I like my days to be organized and structured of which exercise is just one part.	0.207	−0.161	0.060	0.243	**0.326**	.640
	7: My weekly pattern of exercise is repetitive.	0.017	0.019	−0.076	0.022	**0.890**	.714
	19: I follow a set routine for my exercise sessions, for example, walk or run the same route, particular exercises, same amount of time, and so on.	0.034	0.033	0.084	0.032	**0.746**	.739
Factor values	Eigenvalues	11.96	3.30	1.19	1.13	0.68	
	Variance explained (%)	49.84	13.73	4.95	4.72	2.82	
	McDonald's omega	0.95	0.90	0.90	0.80	0.84	

Note: (R) Items 8 and 12 were reversed. Salient factor loadings are in bold.

between the CET total score and its subscales except for lack of exercise enjoyment. They also correlated highly with the CES, again except for lack of exercise enjoyment. Furthermore, the EDE-Q item 18 correlated highly with the CET total score and avoidance and rule-driven behavior, and moderately with weight control exercise, mood improvement, and exercise rigidity.

TABLE 4 Pearson correlations between the Compulsive Exercise Test (CET) and other instruments for the total sample ($N = 2{,}535$)

	CET total	CET avoidance and rule-driven behavior	CET weight control exercise	CET mood improvement	CET lack of exercise enjoyment	CET exercise rigidity
CET total	1					
CET avoidance and rule-driven behavior	0.913[a]	1				
CET weight control exercise	0.862[a]	0.737[a]	1			
CET mood improvement	0.708[a]	0.655[a]	0.492[a]	1		
CET lack of exercise enjoyment	0.165[a]	−0.025	0.158[a]	−0.307[a]	1	
CET exercise rigidity	0.860[a]	0.805[a]	0.635[a]	0.638[a]	−0.109[a]	1
CES	0.893[a]	0.918[a]	0.742[a]	0.614[a]	0.005	0.827[a]
EDE-Q item 18	0.528[a]	0.543[a]	0.467[a]	0.319[a]	0.027	0.482[a]
EDE-Q global	0.521[a]	0.442[a]	0.623[a]	0.252[a]	0.205[a]	0.324[a]
EDI-2 total	0.434[a]	0.367[a]	0.482[a]	0.186[a]	0.279[a]	0.237[a]
BDI-II total	0.326[a]	0.298[a]	0.336[a]	0.118[a]	0.236[a]	0.326[a]
BSI GSI	0.339[a]	0.321[a]	0.333[a]	0.136[a]	0.228[a]	0.196[a]

Note: EDE-Q Item 18: Over the past 28 days, how many times have you exercised in a driven or compulsive way as a means of controlling your weight, shape, or amount of fat, or to burn off calories?

Abbreviations: BDI-II, Beck Depression Inventory-II; BSI, Brief Symptom Inventory; GSI, Global Severity Index, CES, Commitment to Exercise Scale; CET, Compulsive Exercise Test; EDE-Q, Eating Disorder Examination–Questionnaire; EDI-2, Eating Disorder Inventory-2.

[a]The correlation is significant at the level of 0.01 (two sided).

Regarding the EDE-Q global score, the EDI-2 total score, the BDI-II total score, and the BSI GSI, medium-to-high correlations were found with the CET total score, avoidance and rule-driven behavior as well as weight control exercise, and low correlations with mood improvement and lack of exercise enjoyment.

3.7 | Group comparisons

Table S2 presents group means. Taken together, there were no or only small differences between groups. Individuals with AN scored somewhat higher on compulsivity (avoidance and rule-driven behavior as well as exercise rigidity) compared to patients with BN. Group comparisons should be considered with caution, as measurement invariance was not established for all subscales (see above).

3.8 | Sensitivity to change

$N = 1{,}939$ of the 2,535 patients (76.5%) provided data at admission and discharge. Table 5 shows mean SCS and *SDs* of the CET total score and the CET subscales. MANOVA revealed a significant effect of the four groups on the CET ($F_{[18,5,796]} = 2.16$, $p = .003$), suggesting differences in the four groups regarding sensitivity to change. However, post hoc tests only corroborated differences regarding exercise rigidity in adults with AN versus adults with BN, suggesting that the CET is sensitive to change to a similar degree in all groups for all other subscales.

4 | DISCUSSION

Previous psychometric research on the CET was primarily conducted in nonclinical or small mixed ED samples, generating mixed findings on the factor structure of the CET. In this study, we investigated its psychometric properties in a large clinical sample of female adolescent and adult inpatients with AN and BN. In summary, we found that the German CET showed good internal consistency, construct validity, and sensitivity to change. However, questions regarding the factorial structure of the CET remain.

Our CFA did not confirm the originally suggested five-factor structure (Taranis et al., 2011) nor fully support the previously suggested alternative four-factor or three-factor solutions. However, EFA yielded a five-factor solution that identified an almost identical model with sufficient model fit. Differences were mostly in size of factor loadings, and size and direction of factor correlations. When we probed measurement invariance across age and diagnostic groups, full model measurement invariance could not be assumed, but all subscales with more than three items were invariant (i.e., avoidance and rule-driven behavior, weight control exercise, and mood improvement). This indicates that the CET total score does not measure the same overall construct across age and diagnostic groups, and comparisons between, for example, patients with AN and BN should be interpreted with caution. Our results further question the original theoretical five-factor model of compulsive exercise, especially regarding its universal applicability to clinical and nonclinical populations, as well as females of all ages. Several explanations for the divergent findings from our study and previous literature are

TABLE 5 Results regarding sensitivity to change: Standardized change scores in the Compulsive Exercise Test (CET) from baseline to discharge of inpatient treatment in the total sample ($N = 2{,}535$) as well as the four subgroups

		SCS	SD	N
CET total score	Adolescents with AN	0.53	0.90	648
	Adults with AN	0.56	0.84	736
	Adolescents with BN	0.47	0.77	199
	Adults with BN	0.43	0.81	356
	Total group	0.52	0.85	1,939
CET avoidance and rule-driven behavior	Adolescents with AN	0.50	0.90	648
	Adults with AN	0.53	0.83	736
	Adolescents with BN	0.38	0.71	199
	Adults with BN	0.43	0.80	356
	Total group	0.49	0.84	1,939
CET weight control	Adolescents with AN	0.48	0.92	648
	Adults with AN	0.49	0.89	736
	Adolescents with BN	0.57	0.96	199
	Adults with BN	0.39	0.85	356
	Total group	0.48	0.90	1,939
CET mood improvement	Adolescents with AN	0.30	0.91	648
	Adults with AN	0.30	0.85	736
	Adolescents with BN	0.19	0.89	199
	Adults with BN	0.23	0.90	356
	Total group	0.28	0.88	1,939
CET lack of exercise enjoyment	Adolescents with AN	0.13	0.88	648
	Adults with AN	0.19	0.80	736
	Adolescents with BN	0.19	0.85	199
	Adults with BN	0.15	0.79	356
	Total group	0.16	0.83	1,939
CET exercise rigidity	Adolescents with AN	0.48	0.95	648
	Adults with AN	0.55	0.90	736
	Adolescents with BN	0.40	0.79	199
	Adults with BN	0.33	0.89	356
	Total group	0.47	0.91	1,939

Abbreviations: CET, Compulsive Exercise Test; AN, anorexia nervosa; BN, bulimia nervosa; SCS, standardized change score.

conceivable: First, methods differed considerably. For example, Taranis et al. (2011) used PCA to identify a factor structure for the CET, whereas we used EFA to account for item-level measurement error. Second, the theoretical model of compulsive exercise proposed by Meyer et al. (2011), which is the basis of the CET, might not be adequate for clinical populations. For example, we found insufficient corrected item-total correlations of the CET subscale lack of exercise enjoyment and low correlations between this subscale and other CET subscales, as well as another measure (CES) for compulsive exercise. This suggests that lack of exercise enjoyment might not be relevant in compulsive exercise for AN and BN. Furthermore, recent theoretical proposals also suggest improving models for a better understanding of compulsive exercise in EDs (e.g., Coniglio, Cooper, & Selby, 2022; Kolar & Gorrell, 2020).

Internal consistency of the CET was good across all subscales and the total score. Except for lack of exercise enjoyment, both the total score and all other subscales correlated satisfactorily with another measure assessing compulsive exercise (i.e., CES) and engagement episodes in compulsive exercise. They furthermore correlated well, albeit slightly lower, with ED-specific and broad psychopathology measures, except for mood improvement and lack of exercise enjoyment. Thus, the CET generally measures compulsive exercise, while certain factors, especially lack of exercise enjoyment, may be less related to compulsive exercise or constitute a less clinically important factor at least in patients with AN and BN, compared to nonclinical populations.

As measurement invariance was only established for three subscales, group comparisons must be interpreted cautiously. Our

analyses were the first to compare adolescents and adults, but even on measurement invariant subscales we only found small or no differences between age and diagnostic groups. This is, however, in line with previous findings regarding AN and BN (Schlegl et al., 2018).

Our findings did not reveal significant differences between groups on SCS, suggesting comparable sensitivity to change in adolescents and adults as well as patients with AN and BN. The subscales avoidance and rule-driven behavior, weight control exercise, as well as exercise rigidity were the most sensitive to change. Thus, it is reasonable to use these subscales to capture change over time in treatment and research. However, we did not compare the sensitivity to change in the CET with other measures of the same construct, which should be considered in future studies.

Our study is based on an excellent clinical sample size. Both adult and adolescent patients with AN and BN were assessed, and sample size in each of the four subgroups was sufficient for assessing measurement invariance. Furthermore, longitudinal data (admission and discharge) were available for 76.5% of patients. Still, some limitations should be noted: (a) All participants were inpatients with severe and often long-standing illness; therefore, results may not be applicable to other populations such as outpatients. (b) Due to low rates of males with AN or BN in inpatient settings, we investigated only girls and women and findings may vary for boys and men. We also did not differentiate between AN subtypes as these were not reliably recorded at admission. (c) Investigating only AN and BN, our findings may not be applicable to other specified EDs. Therefore, the psychometric properties and factor structure of the CET should be further explored in individuals with other EDs, male patients, individuals in outpatient care, and participants reflecting demographic diversity. (d) Due to admission criteria, patients with comorbid current drug/alcohol/medication abuse, acute suicidal tendencies, psychotic symptoms, or a severe life-threatening somatic disorder were not represented. (e) A substantial part of patients may have received a group therapy designed to specifically address compulsive exercise, called "Healthy Exercise Behavior" (see Dittmer et al., 2020), that our clinic offers in routine care since 2015. Therefore, sensitivity to change might have been overestimated. (f) Our study did not investigate whether the CET is useful for measuring the predictive impact of compulsive exercise on treatment effects or other important outcomes such as BMI or general ED symptomatology. Currently, only one study addressed this research question (Young et al., 2017), finding that the CET better predicted ED psychopathology compared to the CES in a small adult AN sample. Thus, future research should replicate this study in a larger and more diverse ED sample. (g) Finally, the constructs suggested to investigate discriminant validity seem to be part of the ED symptomatology and therefore are probably somehow correlated with compulsive exercise. We did not assess something completely unrelated to EDs/compulsive exercise (e.g., music preference). Construct validity should also be further examined by objectively assessing exercise using accelerometers to determine how the CET relates to exercise intensity. In conclusion, our data suggest that three CET subscales (avoidance and rule-driven behavior, weight control exercise, and mood improvement) can be reliably used in patients with EDs, are suited to compare age and diagnostic groups (measurement invariance was shown), and to assess treatment effects (sensitivity to change was shown). While the total score and the exercise rigidity subscale should be interpreted cautiously, they may still be used for tracking individual severity and change in compulsive exercise. Overall, our data suggest that the CET primarily does a sufficient job in assessing weight control and affect-regulating functions of compulsive exercise, but contrary to its intention performs less well on compulsiveness, raising questions about its theoretical basis. Thus, we argue for proposing new and empirically testable models of compulsive exercise that add other potential facets (e.g., trait compulsivity or interpersonal difficulties, see Kolar et al., 2021) to the established valid and important factors of the CET.

5 | CONCLUSION

The CET showed good internal consistency, construct validity, and sensitivity to change, whereas psychometric properties such as factor structure remain inconclusive. Especially, the support of the lack of enjoyment subscale for use in female individuals with AN and BN is limited, and only moderate for the rigidity subscale. Further replicating and/or revising the original CET may be indicated. Also, evaluating the maintenance model of compulsive exercise before refining the CET is recommended.

ACKNOWLEDGEMENTS

We thank Felicitas Hirler, Julia Maier, and Antonia Sasko for data entry and data management and all the patients for consenting to participate in routine data assessment. This research received no specific grant from any funding agency, nor from commercial entities or the not-for-profit sector. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Sandra Schlegl https://orcid.org/0000-0003-1791-7379
David R. Kolar https://orcid.org/0000-0002-8649-5467

REFERENCES

Arbuckle, J. L. (2019). *Amos (version 26.0) [computer program]*. Chicago: SPSS.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck depression inventory-II*. San Antonio, TX: Psychological Corporation.

Binford, R. B., & Le Grange, D. (2005). Adolescents with bulimia nervosa and eating disorder not otherwise specified-purging only. *International Journal of Eating Disorders*, *38*(2), 157–161. https://doi.org/10.1002/eat.20167

Bratland-Sanda, S., Sundgot-Borgen, J., Rø, Ø., Rosenvinge, J. H., Hoffart, A., & Martinsen, E. W. (2010). "I'm not physically active—I only go for walks": Physical activity in patients with longstanding eating disorders. *International Journal of Eating Disorders*, *43*(1), 88–92. https://doi.org/10.1002/eat.20753

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185–216. https://doi.org/10.1177/135910457000100301

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*, 272–300. https://doi.org/10.1207/s15328007sem1102_8

Carter, J. C., Blackmore, E., Sutandar-Pinnock, K., & Woodside, D. B. (2004). Relapse in anorexia nervosa: A survival analysis. *Psychological Medicine*, *34*(4), 671–679. https://doi.org/10.1017/S0033291703001168

Coniglio, K. A., Cooper, M., & Selby, E. A. (2022). Behavioral reinforcement of pathological exercise in anorexia nervosa. *International Journal of Eating Disorders*, *55*(2), 184–192. https://doi.org/10.1002/eat.23626

Corporation, I. B. M. (2017). *IBM SPSS statistics for windows (version 25.0)*. Armonk, NY: IBM Corporation.

Cunningham, H. E., Pearman, S., & Brewerton, T. D. (2016). Conceptualizing primary and secondary pathological exercise using available measures of excessive exercise. *International Journal of Eating Disorders*, *49*(8), 778–792. https://doi.org/10.1002/eat.22551

Darcy, A. M., Hardy, K. K., Crosby, R. D., Lock, J., & Peebles, R. (2013). Factor structure of the eating disorder examination questionnaire (EDE-Q) in male and female college athletes. *Body Image*, *10*(3), 399–405. https://doi.org/10.1016/j.bodyim.2013.01.008

Davis, C. (1997). Eating disorders and hyperactivity: A psychobiological perspective. *Canadian Journal of Psychiatry*, *42*(2), 168–175. https://doi.org/10.1177/070674379704200207

Davis, C., Brewer, H., & Ratusny, D. (1993). Behavioral frequency and psychological commitment: Necessary concepts in the study of excessive exercising. *Journal of Behavioral Medicine*, *16*(6), 611–628. https://doi.org/10.1007/BF00844722

Derogatis, L. R. (1979). *Symptom Checklist-90-revised (SCL-90-R)*. Bloomington, MN: NCS Pearson.

Derogatis, L. R. (1993). *Brief symptom inventory (BSI): Administration, scoring and procedures manual*. Bloomington, MN: NCS Pearson.

Dittmer, N., Voderholzer, U., Monch, C., Cuntz, U., Jacobi, C., & Schlegl, S. (2020). Efficacy of a specialized group intervention for compulsive exercise in inpatients with anorexia nervosa: A randomized controlled trial. *Psychotherapy and Psychosomatics*, *89*, 161–173. https://doi.org/10.1159/000504583

El Ghoch, M., Calugi, S., Pellegrini, M., Milanese, C., Busacchi, M., Battistini, N. C., Bernabé, J., & Dalle Grave, R. (2013). Measured physical activity in anorexia nervosa: Features and treatment outcome. *International Journal of Eating Disorders*, *46*(7), 709–712. https://doi.org/10.1002/eat.22140

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International Journal of Eating Disorders*, *16*(4), 363–370.

Ford, J., MacCallum, R., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, *39*, 291–314. https://doi.org/10.1111/j.1744-6570.1986.tb00583.x

Formby, P., Watson, H. J., Hilyard, A., Martin, K., & Egan, S. J. (2014). Psychometric properties of the compulsive exercise test in an adolescent eating disorder population. *Eating Behaviors*, *15*(4), 555–557. https://doi.org/10.1016/j.eatbeh.2014.08.013

Franke, G. (2000). *BSI. Brief Symptom Inventory von L. R. Derogatis (Kurzform der SCL-90-R)—Deutsche Version*. Gottingen: Beltz Test.

Garner, D. M. (1991). *Eating disorder inventory-2: Professional manual*. Lutz, FL: Psychological Assessment Resources.

Harris, A., Hay, P., & Touyz, S. (2020). Psychometric properties of instruments assessing exercise in patients with eating disorders: A systematic review. *Journal of Eating Disorders*, *8*, 45. https://doi.org/10.1186/s40337-020-00315-2

Hautzinger, M., Keller, F., & Kühner, C. (2009). *BDI-II. Beck-depressions-Inventar. Revision (Vol. 2)*. Frankfurt, Germany: Pearson Assessment.

Hilbert, A., & Tuschen-Caffier, B. (2016). *Eating disorder examination-questionnaire: German translation [in German]* (2nd ed.). Tübingen: Dgvt-Verlag. Retrieved from. https://www.dgvt-verlag.de/e-books/2_Hilbert_Tuschen-Caffier_EDE-Q_2016.pdf

Hiller, W., Zaudig, M., & Mombour, W. (1997). *Internationale Diagnosen-Checklisten für DSM-IV und ICD-10, IDCL*. Göttingen, Germany: Hogrefe.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. http://dx.doi.org/10.1007/bf02289447

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.

Kline, P. (1993). *The handbook of psychological testing*. London: Routledge. https://doi.org/10.4324/9781315812274

Kolar, D. R., & Gorrell, S. (2020). A call to experimentally study acute affect-regulation mechanisms specific to driven exercise in eating disorders. *International Journal of Eating Disorders*, *54*(3), 280–286. https://doi.org/10.1002/eat.23427

Kolar, D. R., Kaurin, A., Meule, A., Dittmer, N., Schlegl, S., & Voderholzer, U. (2021). Interpersonal, affective and compulsive features of driven exercise in anorexia nervosa [preprint]. *PsyArxiv*. https://doi.org/10.31234/osf.io/pqbsn

Limburg, K., Bodill, K., Watson, H. J., Kane, R. T., Hagger, M. S., & Egan, S. J. (2021). Validity of the compulsive exercise test in regular exercisers. *Eating Disorders: The Journal of Treatment and Prevention*, *29*(5), 447–462. https://doi.org/10.1080/10640266.2019.1677130

Machado, P. P. P., Grilo, C. M., Rodrigues, T. F., Vaz, A. R., & Crosby, R. D. (2020). Eating disorder examination—Questionnaire short forms: A comparison. *International Journal of Eating Disorders*, *53*(6), 937–944. https://doi.org/10.1002/eat.23275

Meyer, C., Taranis, L., Goodwin, H., & Haycraft, E. (2011). Compulsive exercise and eating disorders. *European Eating Disorders Review*, *19*(3), 174–189. https://doi.org/10.1002/erv.1122

Monell, E., Levallius, J., Forsen Mantilla, E., & Birgegard, A. (2018). Running on empty—A nationwide large-scale examination of compulsive exercise in eating disorders. *Journal of Eating Disorders*, *6*, 11. https://doi.org/10.1186/s40337-018-0197-z

Paul, T., & Thiel, A. (2005). *Eating Disorder Inventory-2 (EDI-2): Deutsche Version*. Göttingen, Germany: Hogrefe.

Plateau, C. R., Shanmugam, V., Duckham, R. L., Goodwin, H., Jowett, S., Brooke-Wavell, K. S., Laybourne, A., Arcelus, J., & Meyer, C. (2014). Use of the compulsive exercise test with athletes: Norms and links with eating psychopathology. *Journal of Applied Sport Psychology*, *26*(3), 287–301.

Sauchelli, S., Arcelus, J., Granero, R., Jimenez-Murcia, S., Aguera, Z., Del Pino-Gutierrez, A., & Fernandez-Aranda, F. (2016). Dimensions of compulsive exercise across eating disorder diagnostic subtypes and the validation of the Spanish version of the compulsive exercise test. *Frontiers in Psychology*, *7*, 1852. https://doi.org/10.3389/fpsyg.2016.01852

Schlegl, S., Dittmer, N., Hoffmann, S., & Voderholzer, U. (2018). Self-reported quantity, compulsiveness and motives of exercise in patients with eating disorders and healthy controls: Differences and similarities.

Journal of Eating Disorders, *6*, 17. https://doi.org/10.1186/s40337-018-0202-6

Schlegl, S., Quadflieg, N., Lowe, B., Cuntz, U., & Voderholzer, U. (2014). Specialized inpatient treatment of adult anorexia nervosa: Effectiveness and clinical significance of changes. *BMC Psychiatry*, *14*, 258. https://doi.org/10.1186/s12888-014-0258-z

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304–321. https://doi.org/10.1177/0734282911406653

Shroff, H., Reba, L., Thornton, L. M., Tozzi, F., Klump, K. L., Berrettini, W. H., Brandt, H., Crawford, S., Crow, S., Fichter, M. M., Goldman, D., Halmi, K. A., Johnson, C., Kaplan, A. S., Keel, P., LaVia, M., Mitchell, J., Rotondo, A., Strober, M., Treasure, J., & Bulik, C. M. (2006). Features associated with excessive exercise in women with eating disorders. *International Journal of Eating Disorders*, *39*(6), 454–461. https://doi.org/10.1002/eat.20247

Solenberger, S. E. (2001). Exercise and eating disorders: A 3-year inpatient hospital record analysis. *Eating Behaviors*, *2*(2), 151–168.

Stiles-Shields, C., Bamford, B., Lock, J., & Le Grange, D. (2015). The effect of driven exercise on treatment outcomes for adolescents with anorexia and bulimia nervosa. *International Journal of Eating Disorders*, *48*(4), 392–396. https://doi.org/10.1002/eat.22281

Swenne, I. (2016). Evaluation of the compulsive exercise test (CET) in adolescents with eating disorders: Factor structure and relation to eating disordered psychopathology. *European Eating Disorders Review*, *24*(4), 334–340. https://doi.org/10.1002/erv.2439

Taranis, L., Touyz, S., & Meyer, C. (2011). Disordered eating and exercise: Development and preliminary validation of the compulsive exercise test (CET). *European Eating Disorders Review*, *19*(3), 256–268. https://doi.org/10.1002/erv.1108

Thome, J. L., & Espelage, D. L. (2007). Obligatory exercise and eating pathology in college females: Replication and development of a structural model. *Eating Behaviors*, *8*(3), 334–349. https://doi.org/10.1016/j.eatbeh.2006.11.009

Vrabel, K., & Bratland-Sanda, S. (2019). Exercise obsession and compulsion in adults with longstanding eating disorders: Validation of the Norwegian version of the compulsive exercise test. *Frontiers in Psychology*, *10*, 2370. https://doi.org/10.3389/fpsyg.2019.02370

Young, S., Touyz, S., Meyer, C., Arcelus, J., Rhodes, P., Madden, S., … Pike, K., Attia, E., Crosby, R. D., Wales, J., & Hay, P. (2017). Validity of exercise measures in adults with anorexia nervosa: The EDE, compulsive exercise test and other self-report scales. *International Journal of Eating Disorders*, *50*(5), 533–541. https://doi.org/10.1002/eat.22633

Zeeck, A., Schlegel, S., Giel, K. E., Junne, F., Kopp, C., Joos, A., Davis, C., & Hartmann, A. (2017). Validation of the German version of the commitment to exercise scale. *Psychopathology*, *50*(2), 146–156. https://doi.org/10.1159/000455929

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Schlegl, S., Vierl, L., Kolar, D. R., Dittmer, N., & Voderholzer, U. (2022). Psychometric properties of the Compulsive Exercise Test in a large sample of female adolescent and adult inpatients with anorexia nervosa and bulimia nervosa. *International Journal of Eating Disorders*, *55*(4), 494–504. https://doi.org/10.1002/eat.23694