Validation of the Factor Structure and Psychometric and Clinical Properties of the Multidimensional Emotional Disorder Inventory (MEDI) - German Version

Stephan Bartholdy[1], Raphaela I. Zimmer[2], Andrea Hermann[2,3], Rudolf Stark[2,3], Eva-Lotta Brakemeier[1], and Tim Kaiser[4]

[1] Department of Clinical Psychology and Psychotherapy, University of Greifswald, Germany

[2] Department of Psychotherapy and Systems Neuroscience and Bender Institute of Neuroimaging, Justus Liebig University Giessen, Germany

[3] Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig University Giessen, Germany

[4] Department of Education and Psychology, Methods and Evaluation / Quality Assurance, Free University Berlin

Running title: Validation Study of the MEDI - German Version

Abstract length: 250 words

Article length: 6455 words

Figures: none

Tables: 7

Supplementary tables: 6

Author Note

Stephan Bartholdy https://orcid.org/0000-0002-1532-410X

Raphaela Zimmer https://orcid.org/0000-0002-2939-5072

Andrea Hermann https://orcid.org/0000-0003-3186-7777

Rudolf Stark https://orcid.org/0000-0003-4299-5280

Eva-Lotta Brakemeier https://orcid.org/0000-0001-9589-3697

Tim Kaiser https://orcid.org/0000-0002-6307-5347

We have no known conflict of interest to disclose. This research was partially supported by “The Adaptive Mind”, funded by the Excellence Program of the Hessian Ministry of Higher Education, Research, Science and the Arts. Supplemental Material for this study is available at https://osf.io/j4xdc/. We thank Elena Chiara Hake for her contribution to this study, which consisted of conducting a preliminary validation study of the MEDI - German Version within a sample of healthy individuals as a part of her Bachelor´s thesis.

Correspondence concerning this article should be addressed to Stephan Bartholdy, Department of Clinical Psychology and Psychotherapy, University of Greifswald, Franz-Mehring-Str. 47, 17489 Greifswald, Germany. E-mail: stephan.bartholdy@uni-greifswald.de

Abstract

The empirically proven high prevalence and comorbidity of emotional disorders underscore the need for transdiagnostic measures in routine clinical practice and psychotherapy research. The Multidimensional Emotional Disorder Inventory (MEDI) is a brief, transdiagnostic measure assessing nine dimensions of emotional disorders. The aim of this study was to evaluate the psychometric and clinical properties of the German version of the MEDI. The translated version was administered to a large sample including both healthy individuals and patients from two university-based outpatient psychotherapy clinics (N = 1129). Test-retest data from 273 individuals was used to estimate duration-adjusted retest reliabilities over an interval of approximately 7 months. Convergent and discriminant validity of the MEDI scales were evaluated by analysing information on other clinically relevant constructs, including self-report measures for various symptoms, personality traits and disorders, as well as confirmed clinical diagnoses. Exploratory Structural Equation Modeling (ESEM) was employed to examine the factor structure, resulting in a satisfactory overall fit of the original nine-factor model, albeit with reduced consistency for the Avoidance scale. Overall, the MEDI scales showed high estimates of internal consistency (Cronbach´s α between .73-.92) and acceptable test-retest reliabilities (r_{tt} between .58-.78). Correlations with other established symptom and personality measures, as well as clinical diagnoses, were consistent with expectations and demonstrated good convergent and discriminant validity. In conclusion, the German version of the MEDI demonstrated good psychometric properties, making it particularly suitable as a dimensional measure for evaluating differential effects of therapeutic interventions for emotional disorders in both clinical practice and research contexts.

Keywords: emotional disorders; anxiety; depression; transdiagnostic; dimensional assessment

Public Significance Statement

This study aimed to evaluate the psychometric properties of the MEDI – German Version. The questionnaire allows for an efficient and differential assessment of relevant transdiagnostic symptoms of emotional disorders. Based on this study showing its validity, reliability, and clinical utility, as well as on the provided information, the MEDI – German Version can be applied to assess emotional disorders across a wide range of diagnostic contexts and purposes.

Introduction

Emotional disorders (EDs) are characterized by the experience of frequent and intense negative (and in some disorders also positive) emotions (Bullis et al., 2019). The term of EDs often refers to mood and anxiety disorders (e.g., Muñoz-Navarro et al., 2021; Watson et al., 2008; Zvolensky et al., 2014), but is also considered more broadly including, e.g., the posttraumatic stress disorder (PTSD), obsessive-compulsive disorder (OCD; e.g., Dornbach-Bender et al., 2017), or somatic symptom disorder (e.g., Goldberg et al., 2009). EDs include the most prevalent mental disorders in Germany (Jacobi et al., 2014; Otten et al., 2021) and studies show a generally high comorbidity across mental disorders, but especially among EDs (Aldao et al., 2016; Barlow et al., 2014; Caspi et al., 2014; Kessler et al., 2011; Kessler et al., 2012), with 55% of individuals suffering from a primary anxiety or mood disorder fulfilling criteria for a further diagnosis (Kessler et al., 2005). This proportion even increases when considered across the lifespan (Alonso et al., 2007; Kessler et al., 2011). Moreover, on a symptom level, these diagnoses show a strong overlap (e.g., Zbozinek et al., 2012), questioning the validity of current diagnostic systems (Brown & Barlow, 2009; Brown et al., 2001; Rosellini et al., 2015). In fact, EDs do not only share phenomenological characteristics but also supposed biological and psychological mechanisms (Aldao et al., 2016; Barnow, 2012; Lawrence et al., 2009), further questioning purely categorical approaches.

As a result, dimensional classification approaches have emerged from very early on (e.g., Kendell, 1975) and some were also proposed more recently, such as the Research Domain Criteria (RDoC; Insel et al., 2010) or the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017). However, these models have not been adopted into the routine care setting by clinicians so far for several reasons (Rosellini & Brown, 2019), e.g., because of the difficulty to broadly collect data about biomarkers (as would be required in RDoC) or the need to assess an extensive list of components (as given in HiTOP), which would require several lengthy questionnaires. Moreover, diagnostic categories are indispensable especially in the context of health care systems and the integration of research findings (First, 2005).

For this purpose, Barlow et al. (2004) and Brown and Barlow (2009) proposed a hybrid categorical-dimensional classification model that combines the most important ED symptom dimensions, all of which were theoretically derived and associated with the development, expression, and maintenance of EDs. These dimensions are divided into higher and lower-order dimensions: Higher-order dimensions comprise neurotic and positive temperament, which are essential factors in the etiology and course of EDs (Rosellini & Brown, 2019). Lower order dimensions encompass the most central phenomenological constructs in EDs, namely depressed mood, autonomic arousal, somatic anxiety, social anxiety, intrusive cognitions, traumatic re-experiencing and dissociation, as well as emotion-driven avoidance behavior. The approach is efficient in selecting dimensions parsimoniously, but specific enough to assess EDs (Rosellini & Brown, 2019). It even allows for the derivation of interventions in psychotherapy, making it a valuable tool for treatment planning (Böttcher et al., 2020), especially in a routine care setting in which EDs and comorbidities are very common. Following this approach, for instance, Jurado-González et al. (2024) recently found in a large, randomized trial that transdiagnostic group cognitive behavioral therapy (TD-GCBT) in addition to treatment as usual (TAU) was more effective and had more symptom-specific effects in the long term for the treatment of emotional disorders, as compared to TAU alone. These specific long-term effects of the transdiagnostic intervention highlight the importance of viewing and treating emotional disorders on the level of symptoms rather than diagnoses, as well as evaluating symptom changes accordingly.

The Multidimensional Emotional Disorder Inventory (MEDI; Rosellini & Brown, 2019) is a self-report questionnaire, developed to assess the categorical-dimensional transdiagnostic dimensions in EDs mentioned above. It consists of 49 items assessing nine dimensions, i.e., Neurotic Temperament (NT), Positive Temperament (PT), Depressed Mood (DM), Autonomic Arousal (AA), Somatic Anxiety (SOM), Intrusive Cognitions (IC), Social Anxiety (SOC), Traumatic Re-Experiencing (TRM), and Avoidance (AVD). The questionnaire demonstrated good psychometric properties in an English-speaking clinical (outpatient) sample (Rosellini & Brown, 2019), as well as in two Spanish-speaking samples, specifically a

community (Osma et al., 2021) and a clinical sample (Osma et al., 2023). Exploratory analyses in the Spanish community sample suggested a four-factor structure (Osma et al., 2021), which, however, was outperformed by the original 9-factor solution - a finding also replicated in the Spanish clinical sample (Osma et al., 2023). All scales showed an acceptable composite reliability (Rosellini & Brown, 2019: Raykov's ρ = .68 – .93; Osma et al., 2023: Cronbach's α = .66 – .91) and test-retest reliability over one week (Osma et al., 2023: r_{tt} ≥ .71). All three studies provided evidence for its construct validity with convergent and discriminant measures. Osma et al. (2023) also reported percentiles and T-scores for their clinical data. A differentiation in standard values between gender, age, and diagnostic status has not been reported so far. Although a higher prevalence for mental disorders is found in women compared to men (World Health Organization, 2017), these gender differences were not reflected in MEDI scales in the Spanish clinical sample (Osma et al., 2023). In another recent study, the scales of the MEDI were found to be sensitive in detecting changes over the course of treatment, allowing to monitor and target specific symptom changes in psychotherapy (Böttcher et al., 2020). Based on its ability to capture symptom changes for a wide range of emotional disorders, the authors suggest using MEDI profiles of patients to inform treatment planning and identify the most relevant target symptoms to monitor over the course of treatment (Böttcher et al., 2020; Rosellini & Brown, 2019).

The aim of the present study was to conduct a psychometric evaluation of the German version of the MEDI, including an examination of the internal structure of the instrument, an assessment of scale reliability, and an evaluation of its validity in both clinical and community samples.

Method

Samples

Four separate subsamples were included in the validation of the MEDI – German Version: a subsample of adults seeking cognitive-behavioral therapy (CBT) treatment, which

was drawn from the routine outcome measurement and quality assurance system at the University of Greifswald outpatient clinic (*n* = 764), a community subsample of healthy individuals from Greifswald (*n* = 210), a clinical subsample (*n* = 90), as well as a community subsample of healthy individuals (*n* = 65), which were recruited during a large project on emotion regulation at the University of Giessen. These subsamples were combined to a validation sample (*N* = 1129). For an analysis of the test-retest reliability of the MEDI scales, a smaller sample was drawn from the subsample of treatment-seeking individuals from Greifswald, including only individuals who completed the MEDI both at the time of being admitted to the outpatient clinic's waiting list and immediately before starting treatment (*n* = 273). Lastly, another sample was drawn the same subsample, including only patients who underwent cognitive behavioral therapy (including evidence-based third-wave methods) and completed the MEDI before and after the treatment (*n* = 162), in order to calculate different criteria for clinically relevant change.

Overall, 802 (71.0%) of participants were female, 317 (28.1%) were male, and ten (0.9%) were diverse. The mean age of the sample was 29.2 years (*SD* = 11.6, *range*: 18 to 79). While ethnicity was not assessed directly for either of the subsamples from Greifswald and Giessen, participants were predominantly German-speaking and Caucasian. Within the patient sample, 222 patients had already received a confirmed ICD-10 diagnosis, while the remaining 963 patients did not yet undergo the structured clinical interview, which is routinely used to give confirmed diagnoses. A total of 96 participants had been given either a primary or secondary diagnosis for a depressive disorder (57.8%), 21 for a panic disorder / agoraphobia (12.7%), 13 for a somatic symptom disorder (7.8%), 10 for a post-traumatic stress disorder (6%), 13 for a social anxiety disorder (7.8%), 5 for a generalized anxiety disorder (3%), 5 for a specific phobia (3%), and 3 for obsessive-compulsive disorder (1.8%). A complete list of ICD-10 diagnostic codes and their frequencies for the 222 patients can be found in Supplemental Table 1.

Measures

Multidimensional Emotional Disorder Inventory (MEDI)

The MEDI (Rosellini & Brown, 2019) is a transdiagnostic self-report measure of nine dimensions of psychopathology (see Table 1 for scale descriptions). The items are answered on a nine-point Likert-type scale, on which patients rate how characteristic a statement is of them, or how strongly the statement applies to them, ranging from 0 (“not characteristic of me/does not apply to me”) to 8 (“extremely characteristic of me/applies to me very much”). The translated MEDI – German Version, including item scoring, can be found in the Supplemental Materials of this study. In previous validation studies (Osma et al., 2023; Osma et al., 2021; Rosellini & Brown, 2019), strong support for the reliability and validity was found and the MEDI scales correlated with relevant constructs of already established instruments, but also with categorical diagnoses.

Table 1

Summary of Multidimensional Emotional Disorder Inventory subscales, numbers of items per scale, and short descriptions of scale contents

Scale	N_{items}	Description of scale content
Neurotic Temperament	5	General distress, tendency to worry, stress sensitivity
Positive Temperament	5	Cheerful disposition, optimism, motivation and feelings of accomplishment
Depressed Mood	5	Disappointment, lack of interest, hopelessness, suicidal ideation
Autonomic Arousal	5	Physiological symptoms of arousal like breathlessness, trembling, sudden fear and palpitations
Somatic Anxiety	5	Concerns about physical health, sensations and undiagnosed medical conditions
Social Anxiety	5	Feelings of discomfort or nervousness in social settings
Intrusive Cognitions	6	Involuntary entrance of odd, nonsensical, unacceptable, or uncontrollable thoughts or images
Traumatic Re-Experiencing	5	Recurring memories or images of past traumatic events and the emotional response they elicit

Avoidance	8	Strategies or behaviors employed to prevent or escape from distressing emotions, thoughts, places, or situations

Note. N_{items}: number of items per scale.

Clinical Diagnosis

Categorical diagnoses were determined using the DSM-5 version of the Diagnostic Interview for Mental Disorders (DIPS; Margraf, Cwik, Pflug, & Schneider, 2017; Margraf, Cwik, Suppiger, & Schneider, 2017). The DIPS is a semi-structured interview constructed to ascertain diagnoses of DSM-5 and ICD-10 mental disorders. The interviews were conducted by postgraduate therapists in training who are under supervision of licensed psychotherapists. To test the association of MEDI scales with the presence of a DSM-5 diagnoses, we computed polychoric correlations with binary diagnosis variables. We decided against using point biserial correlations due to their high sensitivity to low base rates that would result in possibly severe distortions of effect size estimates, which is not the case for polychoric correlations (Babchishin & Helmus, 2016). The diagnoses included were: generalized anxiety disorder (GAD), panic disorder with and without agoraphobia (PDA), social anxiety disorder (SAD), specific phobia (SPEC), obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), major depression or persistent depressive disorder (DEP), and somatic symptom disorder (SSD).

Beck Depression Inventory-II (BDI-II)

The BDI-II (Beck et al., 1996) is a self-report measure for the severity of depressive symptoms. It consists of 21 items that assess specific problems commonly associated with a depressive disorder. Extensive research has shown high estimates of validity and reliability for the BDI-II across various samples (e.g., see Beck et al., 1988; Wang & Gorenstein, 2013). For the present study, the BDI-II sum score was used as a validity measure for all MEDI scales, with a focus on the Depressed Mood scale.

Brief Symptom Inventory (BSI)

The BSI (Franke, 2000) is a 53-item self-report measure of psychological and physical symptoms includes nine scales: Somatization, Obsession-Compulsion, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobic Anxiety, Paranoid Ideation, and Psychoticism. BSI items refer to the last seven days and are answered on a 0 (“not at all”) to 4 (“extremely”) scale. Previous studies have found high reliability and validity estimates for the BSI scales across different populations and languages. In the original validation study, an internal consistency Cronbach´s Alpha between .71 (Psychoticism scale) and .85 (Depression scale) and estimates of 1-week test-retest reliability between .68 (Somatization scale) and .91 (Phobic Anxiety scale) were reported (Franke, 2000). The BSI was used to validate all MEDI scales, with a primary focus on the Depressed Mood scale.

Short Version of the Big Five Inventory (BFI-K)

The BFI-K (Rammstedt & John, 2005) is a self-report measure consisting of 21 items to economically assess the Big Five personality dimensions (Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness). The items are scored on a five-point Likert-type scale ranging from “very inaccurate” to “very accurate”. Validation studies showed satisfactory psychometric properties of the BFI-K (Kovaleva et al., 2013; Rammstedt & John, 2005). For the present study, all five dimensions of the BFI-K were used as validity measures for the MEDI.

Obsessive-Compulsive Inventory-Revised (OCI-R)

The OCI-R (Foa et al., 2009) is a self-report measure for the severity of symptoms of obsessive-compulsive disorder (OCD). With 18 items, it assesses the most important symptoms of OCD: Washing, Checking, Ordering, Obsessing, Hoarding, and Neutralising. Previous studies have demonstrated the psychometric and clinical validity of the OCI-R for the assessment of OCD symptoms (Abramowitz & Deacon, 2006; Wootton et al., 2015). In this study, the OCI-R was primarily used to validate the Neurotic Temperament, the Somatic Anxiety, the Intrusive Cognitions, and the Avoidance scales of the MEDI.

Panic and Agoraphobia Scale (PAS)

The PAS (Bandelow, 1995) is a comprehensive assessment tool tailored for gauging the severity of panic disorder and agoraphobia symptoms. The PAS encompasses both a clinician-rated and a self-report version, ensuring versatility in its application across diverse settings. In this study, the self-report version was used. Robust internal consistency, inter-rater reliability, and convergent validity was found for the PAS. For the current study, the PAS was employed to validate the Autonomous Arousal scale of the MEDI.

Posttraumatic Diagnostic Scale (PDS)

The Posttraumatic Diagnostic Scale (Foa et al., 1997) is an instrument designed to assist in the assessment and diagnosis of Posttraumatic Stress Disorder (PTSD). The PDS comprises a self-report questionnaire that gauges both the severity of PTSD symptoms and the presence of a qualifying traumatic event. It showed high internal consistency, test-retest reliability, convergent validity, and a strong agreement with diagnostic interviews in previous studies (Foa et al., 2016). The PDS was used to validate the Traumatic Re-Experiencing scale of the MEDI.

Personality Inventory for DSM-5 and ICD-11 – Brief Form, Modified (PID5BF+M)

The PID5BF+M (Bach et al., 2020) is a 36-item scale and was constructed as a short scale to measure maladaptive personality traits from the dimensional models of personality disorders proposed for DSM-5 (Alternative Model of Personality Disorders, AMPD) and ICD-11. The dimensions are Negative Affectivity, Withdrawal, Antagonism, Disinhibition, Anankastia, and Psychoticism. The items are rated on a four-point Likert-type scale ranging from “does not apply at all” to “applies exactly”. The factor structure of the PID5BF+M was validated in samples from eleven countries.

Data Analysis

All statistical analyses were conducted using R. We used the functions available in the *psych* package (Revelle, 2022) for the EFA procedure, as well as the reliability analysis, and *lavaan* (Rosseel, 2012) for model fit evaluation and invariance testing. A complete list of

R packages that were used in this study, as well as the R code for pre-processing and all analyses can be found in the Supplemental Materials.

Validation of the MEDI Factor Structure

In line with Rosellini and Brown (2019), we used Exploratory Structural Equation Modeling (ESEM; e.g., see Marsh et al., 2014) due to the high likelihood of cross-loadings that would result in bad model fit when using the more restrictive confirmatory factor analysis (CFA) procedure. We implemented a two-step ESEM procedure as proposed by Guàrdia-Olmos et al. (2013) to examine the factor structure of the MEDI in our sample. First, an exploratory factor analysis (EFA) was conducted. Because of the nine originally found dimensions of the MEDI (Rosellini & Brown, 2019), we assumed nine factors and used target rotation to replicate the factor structure reported in this publication. Target rotation in EFA is a post-extraction rotation technique designed to simplify the factor pattern matrix and achieve specific factor structures that have been hypothesized or considered meaningful based on prior research or theory (e.g., see Zhang et al., 2019). It represents a hybrid between the exploratory and confirmatory nature of factor analysis. Target rotation ensures that the discrepancy between the loading matrix and the target matrix is minimized. In the second step, we used the loading matrix obtained from this procedure to evaluate model fit, interpretability, and measurement invariance. To evaluate the model fit, we followed the conventional criteria proposed by Hu and Bentler (1999), with a Comparative Fit Index (CFI) and a Tucker-Lewis Index (TLI) close to or above .95, a Standardized Root Mean Squared Residual (SRMR) near or below .08, and a Root Mean Squared Error of Approximation (RMSEA) near or below .06 as indicators for adequate model fit.

Reliability Analysis

We computed Cronbach's α as a measure of internal consistency, as well as the model-based reliability statistics McDonald's hierarchical (ω_h) and total omega (ω_t). McDonald´s ω_h is a measure of the proportion of test score variance attributable to a general factor, while McDonald´s ω_t is the proportion of variance attributable to all common factors (McDonald, 1999).

We also computed the test-retest reliability for all individuals within the subsample of treatment-seeking individuals from Greifswald who completed the MEDI both after being admitted to the outpatient clinic's waiting list and immediately before starting treatment (n = 273). The time intervals between both assessments varied between approximately four and ten months. Because of this considerable variance time periods, we decided to correct the estimates by fitting a Generalized Additive Model (GAM) that regresses the scale values of the second MEDI assessment on the values of the first assessment and applies a cubic spline smoothing function based on the time interval in days. The standardized regression parameter of the first scale value was then interpreted as the test-retest correlation if the time difference was set to its average (M = 205.51 days).

Measurement Invariance Testing

Testing the measurement invariance (MI) of a psychometric instrument ensures that any observed differences between groups (in terms of the construct being measured) reflect true differences rather than artifacts of the measurement process. This is essential for accurate interpretation, meaningful comparisons, and informed decision-making in research and clinical practice. MI can be established between two or more samples by constraining certain properties of a structural equation model to be equal between groups. For configural invariance, no equality constraints are introduced, and the overall factor structure is evaluated. For metric invariance, the factor loadings are constrained to be equal across groups. If this does not lead to a substantially worse model fit, it can be assumed that both groups have similar loading patterns. Finally, for scalar invariance, intercepts are also constrained to be equal. Again, if the model fit does not decrease too strongly, the comparability of latent constructs across the compared groups is supported. We used the cut-off values for change in model fit established by Chen (2007) to determine whether MI holds. Following Chen´s (2007) recommendations, a decrease of 0.01 or more of the CFI, or an increase of 0.015 or more of the RMSEA indicates a lack of MI. We tested measurement invariance between genders (male or female, but not diverse due to the low number of diverse participants), study locations (Greifswald vs. Giessen), and diagnostic status (patient

vs. community). These comparisons should ensure, as much as possible, that possible disparities in MEDI scale expressions between gender, socioeconomic, and cultural, as well as clinically diverse groups can be correctly identified. In particular, the comparison of clinical and healthy samples is essential for the evaluation of treatment outcomes.

Reliable and Clinically Significant Change

A central function of psychometric instruments in clinical psychology is to detect significant changes after treatment. For this purpose, we calculated change scores and cut-off values for clinically significant change. We used Jacobson and Truax´s (1991) approach to calculate reliable and clinically significant changes. The determination of a "reliable" change is a statistical test against a change value of 0. The Reliable Change Index (RCI) was computed as a change score on the respective scales. Because the RCI alone does not inform the practitioner about whether a change is clinically relevant, it should be accompanied by a measure of “clinical significance”. This, according to Jacobson and Truax (1991) is a transition from the clinically relevant (i.e. symptomatic) range of values to the non-clinical range typically achieved by healthy controls. We used scale mean values and standard deviations of the community sample as the clinically unremarkable range of values, whereas those of the patient sample were considered the clinical range. We calculated a cut-off value for the clinical range as the midpoint between the clinical and non-clinical sample means, i.e., the value at which an individual is equally likely to belong to the clinical as to the non-clinical population. If a patient meets both the RCI criterion and moves from the pathological range below the clinical cut-off, this is commonly referred to as “reliable and clinically significant improvement” (RCSI; Jacobson & Truax, 1991).

In addition to the Jacobson-Truax criteria, we calculated an anchor-based change criterion (ACC). We used Receiver Operating Characteristic (ROC) curves to link the percentage change of the MEDI scales to an item answered by the patients at the end of their therapy. The item text was: "How strongly do you currently still feel burdened by your original problem?" and was answered on a five-point Likert-type scale ("not at all", "a little", "moderately", "strongly", "very strongly"). It was dichotomized, so that the answers "not at all"

or "a little" were classified as “changed”, while all other answers were classified as “not changed”. This procedure is intended to establish a connection to the subjective experience of the patient and their problem (e.g., see Bobo et al., 2016; Kounali et al., 2022) and thus offers a possible alternative to the method of Jacobson and Truax, which is based purely on psychometric properties. It results in a value that indicates the percentage by which a scale value must be reduced for patients to rate their condition as significantly improved.

These indicators of change were calculated for a subset of the patient sample (*n* = 162), for which pre- and post-treatment data for the MEDI as well as responses to the anchor item were available. Patients in this subsample underwent individual cognitive behavioral therapy (including evidence-based third-wave methods) according to national treatment guidelines. Therapy ended based on mutual agreement between patient and therapist and was delivered over an average of 32.36 sessions (*SD* = 20.21, *median* = 28, *range*: 1 - 80). As proposed by Froud and Abel (2014), the change score minimizing the sum of squares of 1-sensitivity and 1-specificity was chosen as the optimal cut-off.

Calculation of Norm Values

We calculated norm values for the MEDI scales. Percentiles and T-scores (i.e. standardized scores with a *mean* = 50 and *SD* = 10) were computed for the total sample as well as subsamples by gender, age, and diagnostic status (see Supplemental Material online[1]).

Results

Table 2 summarizes the means, SDs, reliability estimates, and change criteria of the MEDI scales. For an overview of descriptive statistics for the MEDI scales by subgroup, see Supplemental Tables 3 to 6. The calculated norm values, including percentiles and T-scores, can be found within the Supplemental Material online.

[1] https://osf.io/j4xdc/

Table 2

Means, standard deviations, reliability estimates, and change criteria for the MEDI scale sums (N = 1129)

Scale	Clinical (*n* = 854) *M (SD)*	Community (*n* = 275) *M (SD)*	Retest (*n* = 273) r_{tt}	Total Sample (*N* = 1129) α	ω_t	ω_h	RCI	C	Treatment Subsample (*n* = 162) RCSI	ACC	ACC Improved
Neurotic Temperament	23.30 (8.91)	13.12 (8.45)	.66	.82	.85	.66	15	18	20 (16.5%)	37%	60 (37.0%)
Positive Temperament	20.06 (7.26)	25.56 (6.45)	.70	.73	.81	.51	11*	23**	14 (15.2%)	11%*	88 (53.7%)
Depressed Mood	18.57 (9.59)	6.56 (6.54)	.61	.86	.89	.74	17	11	26 (25.2%)	59%	72 (44.4%)
Autonomic Arousal	13.88 (10.21)	4.19 (5.32)	.63	.86	.89	.69	17	8	18 (12.8%)	63%	71 (47.7%)
Somatic Anxiety	14.96 (9.03)	8.98 (6.46)	.74	.76	.82	.52	13	11	15 (21.1%)	22%	81 (51.9%)
Social Anxiety	18.49 (11.74)	13.29 (10.14)	.78	.92	.94	.64	15	16	13 (17.3%)	21%	90 (58.1%)
Intrusive Cognitions	19.96 (11.02)	10.45 (8.94)	.69	.84	.87	.84	18	15	17 (22.1%)	45%	86 (53.8%)

Traumatic Re-Experiencing	12.91 (10.80)	6.15 (6.14)	.70	.87	.90	.69	16	9	23 (16.2%)	57%	74 (52.9%)
Avoidance	25.64 (11.81)	15.18 (9.17)	.58	.76	.80	.54	21	20	19 (20.0%)	51%	48 (29.3%)

Note. *Clinical:* means and standard deviations referring to the patient sample; *Community:* means and standard deviations referring to the community sample; *Retest:* test-retest subsample of treatment-seeking individuals from the outpatient clinic´s waiting list (*n* = 273); *Treatment Subsample:* subset of the clinical sample, for which pre- and post-treatment data was available (*n* = 162 patients undergoing cognitive behavioral therapy, including evidence-based third-wave methods); *M:* mean; *SD:* standard deviation; r_{tt}*:* duration-adjusted test-retest reliability for a time interval of approximately 7 months; *α:* Cronbach's alpha; ω_t*:* McDonald's total omega; ω_h*:* McDonald's hierarchical omega; *RCI:* Reliable Change Index, i.e. changes in sum scores that need to be surpassed for a change that is significantly different from zero; *C:* Criterion C (Jacobson & Truax, 1991), i.e. the midpoint between the score distributions of the clinical and the community sample; *RCSI:* number and proportion of patients who achieved reliable and clinically significant improvement (note that the percentages refer to the total number of individuals who could theoretically achieve RCSI on the respective scale, i.e. who showed clinically relevant scores before treatment); *ACC:* anchor-based change criterion, i.e. the percentage by which the scale value must change in order to be classified as meaningful; *ACC Improved:* number and percentage of patients who improved according to the ACC; *: Positive Temperament scores should increase; **: Positive Temperament scores must be below the clinical cut-off value in order to be clinically relevant.

Factor Structure in the Validation Sample

An evaluation of the original 9-factor solution resulted in an acceptable model fit, $\chi^2(1131) = 2{,}294.875$, $p < .001$, RMSEA = 0.030 ($p = 1.0$), TLI = 0.957, CFI = 0.958, SRMR = 0.034. Following the conventional criteria proposed by Hu and Bentler (1999), i.e. RMSEA near or below 0.06, TLI and CFI close to or above .95, and SRMR near or below .08, the ESEM model fit was good. For a complete overview of the resulting factor loadings, see Supplemental Table 2.

Factor and Scale Correlations

Correlations between the nine MEDI scales are displayed in Table 3. In concordance with the original validation study, the raw composite MEDI scales were correlated with each other in the expected directions. The negative correlation between the Neurotic Temperament and the Positive Temperament scales was slightly higher than in the original validation study (Rosellini & Brown, 2019), $r = .30$. Neurotic Temperament was highly correlated with Depressed Mood ($r = .60$), Autonomic Arousal ($r = .57$), Social Anxiety ($r = .58$), Intrusive Cognitions ($r = .63$), and Avoidance ($r = .60$). Positive Temperament was negatively correlated with all other dimensions, but most highly with Depressed Mood ($r = -.52$). Depressed Mood was also notably correlated with Intrusive Cognitions ($r = .56$), Avoidance ($r = .50$), and Social Anxiety ($r = .50$), but less closely correlated with Somatic Anxiety ($r = .29$). Autonomic Arousal was highly correlated with Avoidance ($r = .58$), Somatic Anxiety ($r = .57$), and Intrusive Cognitions ($r = .53$). Somatic Anxiety was also highly correlated with Avoidance ($r = .52$). Social Anxiety was most highly correlated with Neurotic Temperament ($r = .58$) and Avoidance ($r = .51$). The Intrusive Cognitions scale was most highly correlated with Neurotic Temperament ($r = .63$) and Avoidance ($r = .60$). The scales with the closest intercorrelations with other scales were Neurotic Temperament ($.46 \leq r \leq .63$, excluding Positive Temperament) and Avoidance ($.50 \leq r \leq .60$, excluding Positive Temperament).

Table 3

Correlations between the nine observed MEDI scales

MEDI Scale	1	2	3	4	5	6	7	8
1. Neurotic Temperament								
2. Positive Temperament	-.30**							
3. Depressed Mood	.60**	-.52**						
4. Autonomic Arousal	.57**	-.16**	.48**					
5. Somatic Anxiety	.46**	-.06*	.29**	.57**				
6. Social Anxiety	.58**	-.38**	.50**	.40**	.28**			
7. Intrusive Cognitions	.63**	-.25**	.56**	.53**	.46**	.46**		
8. Traumatic Re-Experiencing	.48**	-.19**	.43**	.49**	.39**	.31**	.57**	
9. Avoidance	.60**	-.16**	.50**	.58**	.52**	.51**	.60**	.55**

Note. N = 1129. * indicates $p < .05$. ** indicates $p < .01$.

Correlations between MEDI latent factors are displayed in Table 4. The inter-correlations resulting from the ESEM nine-factor model were also according to expectations and generally slightly less pronounced than the raw scale correlations.

Table 4

Correlations between the nine MEDI latent factors

MEDI Latent Factor	1	2	3	4	5	6	7	8
1. Neurotic Temperament								
2. Positive Temperament	-.34							
3. Depressed Mood	.51	-.60						
4. Autonomic Arousal	.43	-.14	.39					
5. Somatic Anxiety	.38	-.02	.20	.58				
6. Social Anxiety	.56	-.26	.46	.49	.45			
7. Intrusive Cognitions	.50	-.37	.38	.35	.29	.41		
8. Traumatic Re-Experiencing	.44	-.20	.38	.43	.46	.56	.28	
9. Avoidance	.59	-.16	.50	.53	.52	.57	.41	.54

Note. N = 1129. All correlations $|r| \geq .14$ are significant, $p < .05$.

Reliability and Measurement Invariance

Reliability

The scale reliability was estimated through the internal consistency Cronbach´s α, McDonald's hierarchical (ω_h) and total omega (ω_t), as well as through the duration-adjusted test-retest reliability r_{tt} for a time interval of approximately 7 months (see Methods section) within a subsample of treatment-seeking individuals. Overall, the resulting estimates of reliability were acceptable (see Table 2) and similar to those reported by Rosellini and Brown (2019). All of the nine MEDI scales showed acceptable levels of internal consistency (α between .73 and .92), general factor saturation (McDonald´s ω_h between .51 and .84), as well as high levels of total factor saturation (McDonald´s ω_t between .80 and .94). Furthermore, the MEDI scales demonstrated acceptable levels of test-retest reliability, with r_{tt} between .58 (Avoidance) and .78 (Social Anxiety).

Measurement Invariance

Table 5 summarizes the two key fit measures and their change by additional model constraints. Measurement invariance was supported for all three comparisons with only minimal decreases of model fit in the case of scalar invariance. The results support the assumption that comparisons between a patient sample and a community sample, male and female individuals, as well as between individuals from different geographic regions (and thus, socio-economic backgrounds) within Germany are valid.

Table 5

Summary of measurement invariance testing results for the Exploratory Structural Equation Modeling (ESEM) model of the MEDI

Model	**Location**		**Gender**		**Sample**	
	CFI	RMSEA	CFI	RMSEA	CFI	RMSEA
Configural	0.946	0.034	0.950	0.033	0.933	0.036
Metric	0.946	0.034	0.950	0.033	0.933	0.036
Δ_{fit}	0	0	0	0	0	0
Scalar	0.944	0.034	0.948	0.034	0.926	0.037
Δ_{fit}	-0.002	0	-0.002	0.001	0.007	0.001

Note. Δ_{fit}: difference of model fit compared to the previous (less constrained) model. Location: measurement invariance between the study locations Giessen (*n* = 155) vs.

Greifswald (*n* = 974). Gender: measurement invariance between male (*n* = 317) and female (*n* = 802) individuals. Sample: measurement invariance between patient (*n* = 854) and community samples (*n* = 275).

Convergent and Discriminant Validity

Self-Report Measures

The convergent and discriminant validity of the MEDI was evaluated by correlating the nine MEDI scales with subscales of other common self-report measures which assess similar constructs. The results were consistent with expectations (see Table 6).

The Neurotic Temperament scale of the MEDI correlated highly with the Neuroticism scale of the BFI-K (r = .82, p < .01), as well as with the Negative Affect scale of the PID-5-BF (r = .71, p < .01), the BDI-II score (r = .57, p < .01), the Global Severity index (GSI; r = .66, p < .01), the Depression scale (r = .51, p < .01), the Anxiety scale (r = .57, p < .01), the Interpersonal Sensitivity scale (r = .62, p < .01), and the Obsession-Compulsion scale of the BSI (r = .51, p < .01). The Positive Temperament scale of the MEDI correlated highly negative with the Neuroticism scale (r = -.59, p < .01) and the Extraversion scale of the BFI-K (r = .56, p < .01). The Depressed Mood scale of the MEDI correlated highly with the Depression scale (r = .84, p < .01), the Interpersonal Sensitivity scale (r = .62, p < .01), the Obsession-Compulsion scale (r = .64, p < .01), and the Global Severity Index (GSI) of the BSI (r = .73, p < .01), as well as with the Neuroticism scale of the BFI-K (r = .65, p < .01), the Detachment scale of the PID-5-BF (r = .53, p < .01), and the BDI-II score (r = .78, p < .01). The Autonomic Arousal scale of the MEDI correlated highly with the Somatization scale (r = .78, p < .01), the Anxiety scale (r = .72, p < .01), and the Global Severity Index (GSI; r = .63, p < .01) of the BSI, as well as with the PAS score (r = .63, p < .01) and the BDI-II score (r = .52, p < .01). The Somatic Anxiety scale of the MEDI correlated highly with the Somatization scale of the BSI (r = .51, p < .01). The Intrusive Cognitions scale of the MEDI correlated highly with the Global Severity Index (GSI; r = .66, p < .01), the Depression scale (r = .51, p < .01), the Anxiety scale (r = .52, p < .01), and the Interpersonal Sensitivity scale of the BSI (r =

.54, $p < .01$), as well as with the Neuroticism scale of the BFI-K ($r = .56$, $p < .01$), the OCI-R score ($r = .56$, $p < .01$), and the BDI-II score ($r = .51$, $p < .01$). The Social Anxiety scale of the MEDI correlated highly with the Extraversion scale ($r = -.63$, $p < .01$) and the Neuroticism scale of the BFI-K ($r = .55$, $p < .01$), as well as with the Global Severity Index (GSI; $r = .56$, $p < .01$) and the Interpersonal Sensitivity scale of the BSI ($r = .54$, $p < .01$), and the Detachment scale of the PID-5-BF ($r = .52$, $p < .01$). The Traumatic Re-Experiencing scale of the MEDI correlated highly with the Global Severity Index (GSI) of the BSI ($r = .59$, $p < .01$) and with the PDS score ($r = .58$, $p < .01$). The Avoidance scale of the MEDI correlated highly with the Global Severity Index (GSI; $r = .60$, $p < .01$) and the Anxiety scale of the BSI ($r = .55$, $p < .01$), as well as with the OCI-R score ($r = .52$, $p < .01$).

Table 6

Correlations between MEDI scales and other self-report measures as estimates of convergent and discriminant validity (with differing subsample sizes ranging from n = 152 to n = 1129)

	PID-5-BF			PAS	OCI-R	BDI-II	BSI						PDS	BFI-K				
MEDI Scale	**Neg. Aff.**	**Detach.**	**Avoid.**				**GSI**	**Depr.**	**Som.**	**Anx.**	**Interpers.**	**Obs.**		**Extrav.**	**Agree.**	**Consc.**	**Neur.**	**Open.**
Neurotic Temperament	.71**	.27**	.10**	.34**	.43**	.57**	.66**	.51**	.40**	.57**	.62**	.51**	.32**	-.35**	-.14	-.21*	**.82****	-.03
Positive Temperament	-.14**	-.43**	-.15**	-.09**	-.08*	-.43**	-.34**	-.43**	-.09**	-.14**	-.29**	-.33**	-.16**	.56**	.35**	.35**	**-.59****	.23**
Depressed Mood	.41**	.53**	.22**	.30**	.26**	.78**	.73**	**.84****	.34**	.42**	.62**	.64**	.35**	-.49**	-.22**	-.42**	.65**	-.19*
Autonomic Arousal	.36**	.19**	.11**	.63**	.31**	.52**	.63**	.37**	**.78****	.72**	.36**	.43**	.40**	-.23**	.06	-.15	.37**	-.16*
Somatic Anxiety	.28**	.09*	.04	.39**	.37**	.34**	.42**	.21**	**.51****	.43**	.24**	.29**	.27**	-.11	-.11	-.12	.32**	-.07
Social Anxiety	.37**	.52**	.19**	.21**	.36**	.37**	**.56****	.46**	.30**	.43**	.54**	.47**	.19**	-.63**	-.13	-.36**	.55**	-.07
Intrusive Cognitions	.49**	.28**	.16**	.35**	.56**	.51**	**.66****	.51**	.38**	.52**	.54**	.48**	.33**	-.26**	-.22**	-.30**	.56**	.01
Traumatic Re-Experiencing	.39**	.20**	.16**	.37**	.41**	.46**	**.59****	.38**	.42**	.48**	.44**	.35**	.58**	-.17*	-.25**	-.19*	.47**	.10
Avoidance	.40**	.32**	.25**	.40**	.52**	.48**	**.60****	.41**	.46**	.55**	.41**	.46**	.35**	-.25**	-.13	-.23**	.47**	-.05

Note. The row-by-row largest coefficient in each case is in bold. * indicates $p < .05$, ** indicates $p < .01$. PID-5-BF = Personality Inventory for DSM-5-Brief Form (n = 661, Neg. Aff. = Negative Affectivity Scale, Detach. = Detachment Subscale, Avoid. = Avoidance Subscale), PAS = Panic and Agoraphobia Scale (n = 1129), OCI-R = Obsessive-Compulsive Inventory-Revised (n = 737), BDI-II = Beck Depression Inventory-II (n = 1129), BSI = Brief Symptom Inventory (n = 919, GSI = Global Severity Index, Depr. = Depression Scale, Som. = Somatization Scale, Anx. = Anxiety Scale, Interpers. = Interpersonal Sensitivity Scale, Obs. = Obsession-Compulsion Scale), PDS = Posttraumatic Diagnostic Scale (n = 1129), BFI-K =

Short Version of the Big Five Inventory (*n* = 152, Extrav. = Extraversion Scale, Agree. = Agreeableness Scale, Consc. = Conscientiousness Scale, Neur. = Neuroticism Scale, Open. = Openness Scale).

DSM-5 Diagnosis

Correlations with DSM-5 diagnoses are reported in Table 7. The correlation pattern reported by Rosellini and Brown (2019) was largely replicated. On average, the resulting correlations deviated by 0.12 from those reported by Rosellini and Brown (2019), and the expected correlations of specific scales were similar.

Table 7

Correlations of MEDI scales with DSM-5 diagnoses

MEDI Scale	GAD	PDA	SAD	SPEC	OCD	PTSD	DEP	SSD
Neurotic Temperament	**.20**	**-.19**	**.19**	.05	**.21**	**.22**	**.23**	**-.20**
Positive Temperament	-.04	**.33**	**-.19**	-.11	-.14	.03	**-.41**	**-.21**
Depressed Mood	-.13	**-.33**	.02	**-.25**	.13	.10	**.55**	-.06
Autonomic Arousal	.05	**.42**	.02	**-.21**	.04	**.20**	0	-.01
Somatic Anxiety	**.19**	**.39**	-.14	-.01	**.21**	.11	.05	**.19**
Social Anxiety	-.11	**-.20**	**.63**	**-.20**	.11	.01	.18	**-.23**
Intrusive Cognitions	-.06	-.17	.05	-.07	**.29**	**.30**	.08	**-.20**
Traumatic Re-Experiencing	.07	-.14	-.14	.10	-.05	**.60**	.06	.02
Avoidance	-.11	.17	.07	-.06	**.22**	**.26**	.12	-.06

Note. N = 222. Polychoric correlations between *Multidimensional Emotional Disorder Inventory (MEDI)* scale mean scores and eight binary *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; Falkai et al., 2018) diagnoses are reported. GAD: generalized anxiety disorder. PDA: panic disorder/agoraphobia. SAD: social anxiety disorder. SPEC: specific phobia. OCD: obsessive-compulsive disorder. PTSD: post-traumatic stress disorder. DEP: depression. SSD: somatic symptom disorder. Correlations printed in bold are significantly different from 0 ($p < .05$).

Reliable and Clinically Significant Change

Comparison of the RCSI criterion with the anchor-based method for classifying change showed that the two approaches differed somewhat in how many patients were classified as "meaningfully changed" (see Table 2). In general, the proportion of ACC-improved patients was larger and, in some cases (Somatic Anxiety scale, Traumatic Re-Experiencing scale), more than twice as large as the proportion of RCSI-classified patients.

Discussion

The present study aimed to validate the factor structure and evaluate the psychometric and clinical properties of the Multidimensional Emotional Disorder Inventory (MEDI) – German Version in a large, German-speaking sample, including both clinical and community subsamples. This study makes three important contributions: (1) the translation of the MEDI into German, thereby making it accessible to a large group of people, (2) a replication of the factor structure of the MEDI found by Rosellini and Brown (2019) in a German-speaking sample, and (3) the development of clinically useful change scores that enable the evaluation of treatment outcomes in clinical practice. The resulting nine-factor solution of the MEDI (including Neurotic Temperament, Positive Temperament, Depressed Mood, Autonomic Arousal, Somatic Anxiety, Social Anxiety, Intrusive Cognitions, Traumatic Re-Experiencing, and Avoidance) replicated the original validation study, as well as the validation studies in Spanish samples (Osma et al., 2023; Osma et al., 2021), and showed good indicators of reliability and validity. Tests of measurement invariance showed that the ESEM model was comparable across male and female individuals, participants from different areas of Germany, as well as subjects from clinical and community samples. The MEDI scales showed the expected correlations with clinical diagnoses, providing further evidence for its validity and clinical utility.

Although the fit of the ESEM model was good overall, some items showed only weak loadings on the Avoidance scale. Satisfactory loadings were observed for items describing distraction from unpleasant thoughts and emotions, which was not the case for items describing behavioral avoidance. We suspect that this could be caused by an overrepresentation of depressed patients, who might be more concerned with avoiding depressive cognitions than situations that lead to anxiety. Similar cross-loadings of anxiety and avoidance items were observed in the original validation study and could be attributed to a clinical sample with high rates of comorbidity (Rosellini & Brown, 2019). Further replication studies should aim for samples with less comorbidity, especially regarding anxiety disorders.

The MEDI scales showed the expected correlation patterns with other questionnaires. High correlations of the Neurotic Temperament scale were observed with PID-5-BF+M Negative Affect and BFI-K Neuroticism. Similar to previous studies (Osma et al., 2023; Osma et al., 2021; Rosellini & Brown, 2019), we found relationships comparable to the structural-categorical hybrid model of Brown and Barlow (2009), namely a negative correlation of Positive and Negative Temperament, moderate associations of Neurotic Temperament with other scales, and significant negative relationships of Positive Temperament with other scales, except Somatic Anxiety. As Böttcher et al. (2020) mentioned that a psychometric validation of the MEDI Somatic Anxiety and Traumatic Re-Experiencing scales with other construct-specific scales was still necessary, two especially relevant results in our study were the moderate correlations between the MEDI Somatic Anxiety and the BSI Somatization scales (r = .51) and between the MEDI Traumatic Re-Experiencing scale and the Posttraumatic Diagnostic Scale total score (r = .58).

The test-retest correlations we obtained were somewhat higher than those reported by Rosellini and Brown (2019), but still resulted in relatively large, and thus conservative, RCI values. These could lead to overly conservative estimates of client change (12.8 - 25.2% reliable and clinically significant improvement in this study). Previous studies with similar samples also reported higher rates of meaningful change. For example, Germer et al. (2022) found a 16% RCSI rate for the RCI-based method and 40% when using a percentage change criterion for the Brief Symptom Inventory (BSI). Many authors have criticized the RCI for several reasons, including the high risk of wrongly classifying patients who improved, and have recommended anchor-based change measures instead (see McAleavey, 2021 for an overview of criticisms). Hence, we also recommend the use of the anchor-based criterion for evaluating individual change in clinical practice. To determine whether the nonclinical value range has been reached after an improvement, the originally proposed criterion C (Jacobson & Truax, 1991) should still be used.

Although previous transdiagnostic measures provide an overall score (such as the “Global Severity Index” of the Symptom Checklist-90, Derogatis & Savitz, 1999), it is

questionable if the unidimensionality of the underlying construct is given. Brown and Barlow (2009) did not assume that the underlying constructs behind emotional disorders merge into one overarching factor, or that there is a "general factor" for psychopathology (“P factor”). The assumption of such a factor is currently subject to detailed criticism (Watts et al., 2019). Thus, we have refrained from proposing a MEDI sum value and encourage practitioners to evaluate client progress as a multidimensional construct, as intended by the MEDI. Nevertheless, we would consider it promising to develop a brief, economic and unidimensional measure from the MEDI items. In many contexts, e.g., within routine outcome monitoring (ROM) systems, it is also necessary to have brief measures available that summarize a patient's overall distress (e.g., see Barkham et al., 2023; McAleavey et al., 2024). Developing a clinically broad, unidimensional short version of the MEDI, which is specifically reliable for repeated assessments, was beyond the scope of this article but is currently being done by one of the co-authors of this study. For examples of the practical use of this recently developed brief version of the MEDI – German Version (Short Emotional Disorder Inventory, SEDI, Kaiser et al., in prep.) within the monitoring and feedback system GPNS, see Demir et al. (2022) and Demir et al. (in prep.).

Limitations

Some limitations of this study need to be acknowledged. First, the sample used for this study was predominantly young with a large proportion of female participants. This can at least in part be explained by the fact that the clinical subsample consisted only of treatment-seeking individuals, and women were found to be more likely to seek psychotherapy in the case of mental health problems (Liddon et al., 2018). A notable limitation is that confirmed diagnoses were not available for the whole patient subsample, but that some included patients did only have suspected diagnoses at the time of the MEDI data collection. Clinically, a major part of the patient subsample was diagnosed with a depressive disorder, while many other disorders were rare. Replication of our results using demographically and clinically representative data is therefore warranted.

In other care systems, e.g., in the Improving Access to Psychological Therapies (IAPT) program, individuals with lower levels of distress would not receive outpatient psychotherapy, but other interventions, like guided psycho-educational groups, computerized CBT, or guided self-help (e.g., see Clark, 2018). Very severely distressed individuals, on the other hand, would probably be treated as inpatients. Future studies with larger and more heterogeneous samples are needed to also ascertain the validity of the MEDI within these populations and contexts, as well as to create more differentiated and clinically useful standard value tables. Another potential limitation was that the inclusion of individuals into the non-clinical subsample was only confirmed by a diagnostic interview procedure for a part of this sample, while the inclusion of the other part was based on self-reports. A possible solution would be to re-evaluate the factor structure and measurement invariance in a larger sample of confirmed healthy individuals. Finally, as Rosellini and Brown (2019) already noted, the temporal invariance still needs to be evaluated to see whether the same factor structure holds across time. As Fried et al. (2016) pointed out, this information would be important for evaluating changes during treatment.

Conclusions

The factor structure of the Multidimensional Emotional Disorder Inventory (MEDI) – German Version was validated in a large, German-speaking sample comprised of both diagnosed patients waiting for outpatient psychotherapy as well as individuals with no diagnosed psychological disorders. The results indicated a good model fit, high convergent and discriminant validity, and adequate indices of measurement invariance. Duration-adjusted estimates of test-retest reliability were acceptable and comparable to previous studies using the English or Spanish versions of the inventory. Its adequate psychometric properties especially enable the MEDI – German Version to be used as a dimensional measure of change in important psychopathological constructs over time, e.g., in order to monitor and interpret differential effects of therapeutic interventions to better personalize psychotherapy in clinical practice and research (e.g., see Demir et al., 2022).

In conclusion, based on the present study showing its validity and clinical utility, as well as on the provided information (e.g., criteria for reliable and clinically significant change), the MEDI – German Version can be applied to assess emotional disorders across a wide range of diagnostic contexts and purposes, including randomized controlled trials and process-outcome research in routine clinical settings.

CRediT Statement

SB, RZ, AH, and TK prepared the original draft of the manuscript. TK and ELB supervised the research activities at the University of Greifswald. RS and AH supervised the research activities at the University of Giessen. SB and RZ led the data curation process. SB and TK conducted the formal analysis of the data. SB, RZ, AH, and TK developed the design of the study and conducted the investigation process. All authors reviewed and edited the final version of the manuscript.

Statement of Ethics

Hereby, we confirm that we complied with the guidelines for human studies and that the research was conducted ethically following the World Medical Association Declaration of Helsinki of 1975, as revised in 2008. Ethical approval was obtained for a patient subsample recruited for a study at the University of Greifswald outpatient clinic (internal registration number: BB177/23), as well as for the community and patient subsamples recruited at the University of Giessen (internal registration number: 2018-0037). However, the present data analysis was retrospective in nature and only included patient data that were either collected within other research projects or within the routinely administered clinical diagnostic procedure at the University of Greifswald outpatient clinic. Nevertheless, all patients gave informed consent to anonymous evaluations of their routinely collected data.

References

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. *Journal of Anxiety Disorders*, *20*(8), 1016–1035. https://doi.org/10.1016/j.janxdis.2006.03.001

Aldao, A., Gee, D. G., Los Reyes, A. de, & Seager, I. (2016). Emotion regulation as a transdiagnostic factor in the development of internalizing and externalizing psychopathology: Current and future directions. *Development and Psychopathology*, *28*(4pt1), 927–946. https://doi.org/10.1017/S0954579416000638

Alonso, J., Lépine, J.-P., & others (2007). Overview of key data from the European Study of the Epidemiology of Mental Disorders (ESEMeD). *Journal of Clinical Psychiatry*, *68*(Suppl 2), 3–9.

Babchishin, K. M., & Helmus, L.-M. (2016). The influence of base rates on correlations: An evaluation of proposed alternative effect sizes with real-world data. *Behavior Research Methods*, *48*(3), 1021–1031. https://doi.org/10.3758/s13428-015-0627-7

Bach, B., Kerber, A., Aluja, A., Bastiaens, T., Keeley, J. W., Claes, L., Fossati, A., Gutierrez, F., Oliveira, S. E. S., Pires, R., Riegel, K. D., Rolland, J.-P., Roskam, I., Sellbom, M., Somma, A., Spanemberg, L., Strus, W., Thimm, J. C., Wright, A. G. C., & Zimmermann, J. (2020). International Assessment of DSM-5 and ICD-11 Personality Disorder Traits: Toward a Common Nosology in DSM-5.1. *Psychopathology*, *53*(3-4), 179–188. https://doi.org/10.1159/000507589

Bandelow, B. (1995). Assessing the efficacy of treatments for panic disorder and agoraphobia. Ii. The Panic and Agoraphobia Scale. *International Clinical Psychopharmacology*, *10*(2), 73–81. https://doi.org/10.1097/00004850-199506000-00003

Barkham, M., de Jong, K., Delgadillo, J., & Lutz, W. (2023). Routine Outcome Monitoring (ROM) and Feedback: Research Review and Recommendations. *Psychotherapy*

Research: Journal of the Society for Psychotherapy Research, *33*(7), 841–855. https://doi.org/10.1080/10503307.2023.2181114

Barlow, D. H., Allen, L. B., & Choate, M. L. (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy*, *35*(2), 205–230. https://doi.org/10.1016/S0005-7894(04)80036-4

Barlow, D. H., Sauer-Zavala, S., Carl, J. R., Bullis, J. R., & Ellard, K. K. (2014). The Nature, Diagnosis, and Treatment of Neuroticism. *Clinical Psychological Science*, *2*(3), 344–365. https://doi.org/10.1177/2167702613505532

Barnow, S. (2012). Emotionsregulation und Psychopathologie. *Psychologische Rundschau*, *63*(2), 111–124. https://doi.org/10.1026/0033-3042/a000119

Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck depression inventory-II. *Psychological Assessment*.

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, *8*(1), 77–100. https://doi.org/10.1016/0272-7358(88)90050-5

Bobo, W. V., Angleró, G. C., Jenkins, G., Hall-Flavin, D. K., Weinshilboum, R., & Biernacka, J. M. (2016). Validation of the 17-item Hamilton Depression Rating Scale definition of response for adults with major depressive disorder using equipercentile linking to Clinical Global Impression scale ratings: Analysis of Pharmacogenomic Research Network Antidepressant Medication Pharmacogenomic Study (PGRN-AMPS) data. *Human Psychopharmacology*, *31*(3), 185–192. https://doi.org/10.1002/hup.2526

Böttcher, H., Correa, J., Cassiello-Robbins, C., Ametaj, A., Rosellini, A. J., Brown, T. A., Kennedy, K., Farchione, T. J., & Barlow, D. H. (2020). Dimensional Assessment of Emotional Disorder Outcomes in Transdiagnostic Treatment: A Clinical Case Study. *Cognitive and Behavioral Practice*, *27*(4), 442–453. https://doi.org/10.1016/j.cbpra.2019.11.001

Brown, T. A., & Barlow, D.-H. (2009). A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: Implications for assessment and treatment. *Psychological Assessment*, *21*(3), 256–271. https://doi.org/10.1037/a0016608

Brown, T. A., Di Nardo, P. A., Lehman, C. L., & Campbell, L. A. (2001). Reliability of DSM-IV anxiety and mood disorders: Implications for the classification of emotional disorders. *Journal of Abnormal Psychology*, *110*(1), 49–58. https://doi.org/10.1037/0021-843X.110.1.49

Bullis, J. R., Boettcher, H., Sauer-Zavala, S., Farchione, T. J., & Barlow, D. H. (2019). What is an emotional disorder? A transdiagnostic mechanistic definition with implications for assessment, treatment, and prevention. *Clinical Psychology: Science and Practice*, *26*(2), e12278. https://doi.org/10.1111/cpsp.12278

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clinical Psychological Science*, *2*(2), 119–137. https://doi.org/10.1177/2167702613497473

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Clark, D. M. (2018). Realizing the Mass Public Benefit of Evidence-Based Psychological Therapies: The IAPT Program. *Annual Review of Clinical Psychology*, *14*(1), 159–183. https://doi.org/10.1146/annurev-clinpsy-050817-084833

Demir, S., Bartholdy, S., Tewes, M., Brakemeier, E.-L., & Kaiser, T. (in prep.). Advances in Routine Outcome Monitoring and Feedback Systems: Insights from the Greifswald Psychotherapy Navigation System (GPNS). In *Advances in the science and practice of feedback-informed psychological therapy.* (Original work published in prep.)

Demir, S., Schwarz, F., & Kaiser, T. (2022). Therapy from my point of view: A case illustration of routine outcome monitoring and feedback in psychotherapeutic interventions. *Journal of Clinical Psychology*, *78*(10), 2029–2040. https://doi.org/10.1002/jclp.23408

Derogatis, L. R., & Savitz, K. L. (1999). The SCL-90-R, Brief Symptom Inventory, and Matching Clinical Rating Scales.

Dornbach-Bender, A., Ruggero, C. J., Waszczuk, M. A., Gamez, W., Watson, D., & Kotov, R. (2017). Mapping emotional disorders at the finest level: Convergent validity and joint structure based on alternative measures. *Comprehensive Psychiatry*, *79*, 31–39. https://doi.org/10.1016/j.comppsych.2017.06.011

First, M. B. (2005). Clinical utility: A prerequisite for the adoption of a dimensional approach in DSM. *Journal of Abnormal Psychology*, *114*(4), 560–564. https://doi.org/10.1037/0021-843X.114.4.560

Foa, E. B., Cashman, L., Jaycox, L., & Perry, K. (1997). The validation of a self-report measure of posttraumatic stress disorder: The Posttraumatic Diagnostic Scale. *Psychological Assessment*, *9*(4), 445–451. https://doi.org/10.1037/1040-3590.9.4.445

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2009). The Obsessive-Compulsive Inventory-Revised (1st ed.).

Foa, E. B., McLean, C. P., Zang, Y., Zhong, J., Powers, M. B., Kauffman, B. Y., Rauch, S., Porter, K., & Knowles, K. (2016). Psychometric properties of the Posttraumatic Diagnostic Scale for DSM-5 (PDS-5). *Psychological Assessment*, *28*(10), 1166–1171. https://doi.org/10.1037/PAS0000258

Franke, G. H. (2000). Brief Symptom Inventory (BSI) von LR Derogatis-Deutsche Version. *Göttingen: Beltz Test*.

Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time … Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating

scales of depression. *Psychological Assessment*, *28*(11), 1354–1367. https://doi.org/10.1037/pas0000275

Germer, S., Weyrich, V., Bräscher, A.-K., Mütze, K., & Witthöft, M. (2022). Does practice really make perfect? A longitudinal analysis of the relationship between therapist experience and therapy outcome: A replication of Goldberg, Rousmaniere, et al. (2016). *Journal of Counseling Psychology*, *69*(5), 745–754. https://doi.org/10.1037/cou0000608

Goldberg, D. P., Krueger, R. F., Andrews, G., & Hobbs, M. J. (2009). Emotional disorders: Cluster 4 of the proposed meta-structure for DSM-V and ICD-11. *Psychological Medicine*, *39*(12), 2043–2059. https://doi.org/10.1017/S0033291709990298

Guàrdia-Olmos, J., Peró-Cebollero, M., Ben\'\itez-Borrego, S., Fox, J., & others (2013). Using sem library in r software to analyze exploratory structural equation models. In *Proceedings of the 59th ISI World Statistics Congress*. https://www.statistics.gov.hk/wsc/CPS105-P6-S.pdf

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. https://doi.org/10.1080/10705519909540118

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Jacobi, F., Höfler, M., Strehle, J., Mack, S., Gerschler, A., Scholl, L., Busch, M. A., Maske, U., Hapke, U., Gaebel, W., Maier, W., Wagner, M., Zielasek, J., & Wittchen, H.-U. (2014). Psychische Störungen in der Allgemeinbevölkerung: Studie zur Gesundheit Erwachsener in Deutschland und ihr Zusatzmodul Psychische Gesundheit (DEGS1-MH) [Mental disorders in the general population : Study on the health of adults in Germany and the additional module mental health (DEGS1-MH)]. *Der Nervenarzt*, *85*(1), 77–87. https://doi.org/10.1007/s00115-013-3961-y

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Jurado-González, F., García-Torres, F., Contreras, A., Muñoz-Navarro, R., González-Blanch, C., Adrián Medrano, L., Ruiz-Rodríguez, P., Moreno, E. M., Pérez-Dueñas, C., Cano-Vindel, A., & Moriana, J. A. (2024). Comparing psychological versus pharmacological treatment in emotional disorders: A network analysis. *PloS One*, *19*(4), e0301675. https://doi.org/10.1371/journal.pone.0301675

Kendell, R. E. (1975). *The role of diagnosis in psychiatry*. Blackwell Scientific Publications.

Kessler, R. C., Chiu, W. T., Demler, O., Merikangas, K. R., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 617–627. https://doi.org/10.1001/archpsyc.62.6.617

Kessler, R. C., Ormel, J., Petukhova, M., McLaughlin, K. A., Green, J. G., Russo, L. J., Stein, D. J., Zaslavsky, A. M., Aguilar-Gaxiola, S., Alonso, J., Andrade, L., Benjet, C., Girolamo, G. de, Graaf, R. de, Demyttenaere, K., Fayyad, J., Haro, J. M., Hu, C. y., Karam, A., . . . Ustün, T. B. (2011). Development of lifetime comorbidity in the World Health Organization world mental health surveys. *Archives of General Psychiatry*, *68*(1), 90–100. https://doi.org/10.1001/archgenpsychiatry.2010.180

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. https://doi.org/10.1002/mpr.1359

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., . . . Zimmerman, M. (2017). The Hierarchical Taxonomy of

Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kounali, D., Button, K. S., Lewis, G., Gilbody, S., Kessler, D., Araya, R., Duffy, L., Lanham, P., Peters, T. J., Wiles, N., & Lewis, G. (2022). How much change is enough? Evidence from a longitudinal study on depression in UK primary care. *Psychological Medicine*, *52*(10), 1875–1882. https://doi.org/10.1017/s0033291720003700

Kovaleva, A., Beierlein, C., Kemper, C. J., & Rammstedt, B. (2013). Psychometric properties of the BFI-K: A cross-validation study. *The International Journal of Educational and Psychological Assessment*.

Lawrence, A. E., Liverant, G. I., Rosellini, A. J., & Brown, T. A. (2009). Generalized anxiety disorder within the course of major depressive disorder: Examining the utility of the DSM-IV hierarchy rule. *Depression and Anxiety*, *26*(10), 909–916. https://doi.org/10.1002/da.20607

Liddon, L., Kingerlee, R., & Barry, J. A. (2018). Gender differences in preferences for psychological treatment, coping strategies, and triggers to help-seeking. *The British Journal of Clinical Psychology*, *57*(1), 42–58. https://doi.org/10.1111/bjc.12147

Margraf, J., Cwik, J. C., Pflug, V., & Schneider, S. (2017). Structured clinical interviews for mental disorders across the life span: psychometric quality and further developments of the DIPS open access interviews [Strukturierte klinische Interviews zur Erfassung psychischer Störungen über die Lebensspanne: Gütekriterien und Weiterentwicklungen der DIPS-Verfahren]. *Zeitschrift Für Klinische Psychologie Und Psychotherapie*, *46*(3), 176–186.

Margraf, J., Cwik, J. C., Suppiger, A., & Schneider, S. (2017). *DIPS Open Access: Diagnostisches Interview bei psychischen Störungen.* https://doi.org/10.13154/rub.100.89

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory

factor analysis. *Annual Review of Clinical Psychology*, *10*, 85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700

McAleavey, A. (2021). When (not) to rely on the reliable change index. https://osf.io/download/619b94998dbcf80493eda59d/

McAleavey, A., de Jong, K., Nissen-Lie, H. A., Boswell, J. F., Moltu, C., & Lutz, W. (2024). Routine Outcome Monitoring and Clinical Feedback in Psychotherapy: Recent Advances and Future Directions. *Administration and Policy in Mental Health*, *51*(3), 291–305. https://doi.org/10.1007/s10488-024-01351-9

McDonald, R. P. (1999). Test theory: A unified treatment.

Muñoz-Navarro, R., Adrián Medrano, L., González-Blanch, C., Carpallo-González, M., Olave, L., Iruarrizaga, I., Ruiz-Rodríguez, P., Moriana, J. A., & Cano-Vindel, A. (2021). Validity of Brief Scales for Assessing Transdiagnostic Cognitive Emotion Regulation Strategies in Primary Care Patients With Emotional Disorders. *European Journal of Psychological Assessment*, *37*(2), 86–94. https://doi.org/10.1027/1015-5759/a000584

Osma, J., Martínez-Loredo, V., Quilez-Orden, A., Peris-Baquero, O., Ferreres-Galán, V., Prado-Abril, J., Torres-Alfosea, M. A., & Rosellini, A. J. (2023). Multidimensional emotional disorders inventory: Reliability and validity in a Spanish clinical sample. *Journal of Affective Disorders*, *320*, 65–73. https://doi.org/10.1016/j.jad.2022.09.140

Osma, J., Martínez-Loredo, V., Quilez-Orden, A., Peris-Baquero, O., & Suso-Ribera, C. (2021). Validity Evidence of the Multidimensional Emotional Disorders Inventory among Non-Clinical Spanish University Students. *International Journal of Environmental Research and Public Health*, *18*(16). https://doi.org/10.3390/ijerph18168251

Otten, D., Tibubos, A. N., Schomerus, G., Brähler, E., Binder, H., Kruse, J., Ladwig, K.-H., Wild, P. S., Grabe, H. J., & Beutel, M. E. (2021). Similarities and Differences of Mental Health in Women and Men: A Systematic Review of Findings in Three Large

German Cohorts. *Frontiers in Public Health*, *9*, 553071. https://doi.org/10.3389/fpubh.2021.553071

Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K). *Diagnostica*, *51*(4), 195–206. https://doi.org/10.1026/0012-1924.51.4.195

Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research* [Computer software]. https://CRAN.R-project.org/package=psych

Rosellini, A. J., Boettcher, H., Brown, T. A., & Barlow, D. H. (2015). A Transdiagnostic Temperament-Phenotype Profile Approach to Emotional Disorder Classification: An Update. *Journal of Experimental Psychopathology*, *a2*(1), 110–128. https://doi.org/10.5127/pr.036014

Rosellini, A. J., & Brown, T. A. (2019). The Multidimensional Emotional Disorder Inventory (MEDI): Assessing transdiagnostic dimensions to validate a profile approach to emotional disorder classification. *Psychological Assessment*, *31*(1), 59–72. https://doi.org/10.1037/pas0000649

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Wang, Y.-P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: A comprehensive review. *Revista Brasileira De Psiquiatria (Sao Paulo, Brazil : 1999)*, *35*(4), 416–431. https://doi.org/10.1590/1516-4446-2012-1048

Watson, D., O'Hara, M. W., & Stuart, S. (2008). Hierarchical structures of affect and psychopathology and their implications for the classification of emotional disorders. *Depression and Anxiety*, *25*(4), 282–288. https://doi.org/10.1002/da.20496

Watts, A. L., Poore, H. E., & Waldman, I. D. (2019). Riskier Tests of the Validity of the Bifactor Model of Psychopathology. *Clinical Psychological Science*, *7*(6), 1285–1303. https://doi.org/10.1177/2167702619855035

Wootton, B. M., Diefenbach, G. J., Bragdon, L. B., Steketee, G., Frost, R. O., & Tolin, D. F. (2015). A contemporary psychometric evaluation of the Obsessive Compulsive

Inventory-Revised (OCI-R). *Psychological Assessment*, *27*(3), 874–882. https://doi.org/10.1037/pas0000075

World Health Organization. (2017). *Depression and other common mental disorders: global health estimates (No. WHO/MSD/MER/2017.2).* World Health Organization.

Zbozinek, T. D., Rose, R. D., Wolitzky-Taylor, K. B., Sherbourne, C., Sullivan, G., Stein, M. B., Roy-Byrne, P. P., & Craske, M. G. (2012). Diagnostic overlap of generalized anxiety disorder and major depressive disorder in a primary care sample. *Depression and Anxiety*, *29*(12), 1065–1071. https://doi.org/10.1002/da.22026

Zhang, G., Hattori, M., Trichtinger, L. A., & Wang, X. (2019). Target rotation with both factor loadings and factor correlations. *Psychological Methods*, *24*(3), 390–402. https://doi.org/10.1037/met0000198

Zvolensky, M. J., Farris, S. G., Leventhal, A. M., & Schmidt, N. B. (2014). Anxiety sensitivity mediates relations between emotional disorders and smoking. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, *28*(3), 912–920. https://doi.org/10.1037/a0037450