JOURNAL OF PERSONALITY ASSESSMENT
2024, VOL. 106, NO. 6, 711–723
https://doi.org/10.1080/00223891.2023.2268199

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Evaluating the Psychometric Properties of the German Self and Interpersonal Functioning Scale (SIFS)

Caroline Macina[1*], André Kerber[2*], Johannes Zimmermann[3], Ludwig Ohse[4], Leonie Kampe[5], Jil Mohr[4], Marc Walter[6], Susanne Hörz-Sagstetter[4*] and Johannes Sebastian Wrege[1*]

[1]Universitäre Psychiatrische Kliniken (UPK), Basel, Switzerland; [2]Freie Universität Berlin, Berlin, Germany; [3]Universität Kassel, Kassel, Germany; [4]Psychologische Hochschule Berlin, Berlin, Germany; [5]Internationale Psychoanalytische Universität, Berlin, Germany; [6]Psychiatrische Dienste Aargau, Windisch, Switzerland

ABSTRACT

The Self and Interpersonal Functioning Scale (SIFS) is a 24-item self-report questionnaire assessing personality functioning according to the alternative DSM-5 model for personality disorders. We evaluated the German SIFS version in a total sample of 886 participants from Germany and Switzerland. Its factor structure was investigated with confirmatory factor analysis comparing bifactor models with two specific factors (self- and interpersonal functioning) and four specific factors (identity, self-direction, empathy, and intimacy). The SIFS sum and domain scores were tested for reliability and convergent validity with self-report questionnaires and interviews for personality functioning, -organization, -traits, -disorder categories, and well-being. None of the bifactor models yielded good model fit, even after excluding two items with low factor loadings and including a method factor for reverse-keyed items. Based on a shortened 22-item SIFS version, models suggested that the g-factor explained 52.9–59.6% of the common variance and that the SIFS sum score measured the g-factor with a reliability of .68–.81. Even though the SIFS sum score showed large test-retest reliability and correlated strongly with well-established self-report questionnaires and interviews, the lack of structural validity appears to be a serious disadvantage of the SIFS compared to existing self-reports questionnaires of personality functioning.

ARTICLE HISTORY
Received 13 February 2023
Revised 10 September 2023
Accepted 26 September 2023

The alternative model for personality disorders (AMPD) in Section III of the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5; APA, 2013) and the International Classification of Diseases (11th ed.; ICD-11; WHO, 2018) both present a dimensional model for the assessment of personality disorders (PDs). Therein, several shortcomings of the categorical approach of PDs in the DSM-IV (APA, 1994), DSM-5 Section II (APA, 2013), and the ICD-10 (WHO, 1992) have been addressed. Empirically, categorical PD models suffer from low reliability (Clark, 2007), low discriminant and convergent validity, overlaps and comorbidity of the PD categories, arbitrary threshold values, and temporal instability of the categorical PD diagnoses (Morey et al., 2015; Widiger & Trull, 2007). In the attempt to overcome some of these weaknesses of the categorical PD approach, the AMPD delineates the common denominator of PDs as impairments in self and interpersonal functioning (Criterion A) and provides a direct assessment of severity with the Level of Personality Functioning Scale (LPFS; Bender et al., 2011).

Criterion A is operationalized in the LPFS by impairments in self-functioning, comprising the domains of identity (ID) and self-direction (SD), and impairments in interpersonal functioning, comprising the domains of empathy (EM) and intimacy (INT). Each of the four domains is defined by three subdomains, resulting in a total of twelve subdomains of personality functioning. Severity levels of the LPFS reach from no or low impairments (0) to extreme impairments (4). For a PD diagnosis, at least moderate impairment (2) in overall personality functioning is required. In addition, the DSM-5 AMPD contains Criterion B, which consists of five pathological personality trait domains as follows: negative affectivity, disinhibition, detachment, antagonism, and psychoticism (APA, 2013). It has been argued that, while Criterion A can be understood as the underlying condition of a PD, Criterion B comprises the consequences of these impaired capacities conceptualized in Criterion A, which may account for the large intercorrelation of the two criteria (Sharp & Wall, 2021). Thus, the LPFS avoids the reductionistic division into healthy and disordered subjects, takes into account the dimensional nature of personality pathology, and efficiently addresses a severity factor common to all PDs (Zimmermann, Hopwood, et al., 2023). In addition, a severity continuum may help raise awareness that a PD is modifiable, thereby reducing stigma against subjects with PDs (Tyrer et al., 2015). Morey et al. (2013) found that

CONTACT Caroline Macina caroline.macina@upk.ch University Psychiatric Clinics (UPK) Basel, Wilhelm Klein-Strasse 27, Basel, 4002, Switzerland.
*These authors share first and last authorship.
Supplemental data for this article is available online at https://doi.org/10.1080/00223891.2023.2268199
© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

clinician ratings of the LPFS are incrementally associated with prognosis, functioning, and treatment intensity needs when controlling for all ten PD categories of the DSM-IV. Therefore, a direct assessment of Criterion A *via* LPFS seems clinically useful, although we acknowledge that other authors are more skeptical about the validity and utility of Criterion A (e.g., Widiger & Hines, 2022; Wright et al., 2022).

Historically, the LPFS was conceptualized as a clinical expert rating and only later structured interviews and self-report questionnaires were developed. Today, there are several interviews for assessing personality functioning, including the Structured Clinical Interview for the AMPD—Module I (SCID-5-AMPD-I; Bender et al., 2018) and the Semi-Structured Interview for personality functioning in DSM-5 (STiP-5.1; Hutsebaut et al., 2017). Furthermore, several self-report questionnaires for the assessment of personality functioning were developed in addition to the existing clinical interviews, including the 80-item LPFS-Self Report (LPFS-SR; Morey, 2017), the 12-item LPFS-Brief Form 2.0 (LPFS-BF 2.0; Weekers et al., 2019), the 12-item LPFS-SR of Criterion A (LPFS-SRA; Roche et al., 2016, 2018), the 132-item DSM-5 Level of Personality Functioning Questionnaire (DLOPFQ; Huprich et al., 2018), and the 23-item DLOPFQ- short form (DLOPFQ-SF; Siefert et al., 2020). Current studies are investigating their reliability, validity, and feasibility in clinical practice (for overviews, see Birkhölzer et al., 2021; Zimmermann, Hopwood, et al., 2023).

Another self-report questionnaire capturing personality functioning is the 24-item Self and Interpersonal Functioning Scale (SIFS), which was developed in French (Gamache et al., 2019). This measure is not quite as brief as the LPFS-BF 2.0 (i.e., twice as many items), but provides four domain scales (ID, SD, EM, INT) instead of only two scales for the dimensions of self- and interpersonal functioning. However, it is still time-efficient for subjects to complete (compared to the LPFS-SR or DLOPFQ) and therefore may fill a gap in the repertoire of self-report assessments of personality functioning. Additionally, the first study with the original version of the SIFS showed excellent internal consistency for the global score, strong associations with relevant personality constructs (e.g., low life satisfaction, aggression, identity diffusion, primitive defense, low empathy, low self-esteem, narcissistic and borderline symptoms, and pathological personality trait domains) and good test-retest reliability in clinical and non-clinical groups for all four SIFS domain scales. Nevertheless, two items (items 6 and 16) seemed to be problematic, one of which (item 16: "I have little interest for other people's feelings or problems") was excluded from testing the factor analytical models because of missing discrimination between clinical and non-clinical groups (Gamache et al., 2019). Even so, Gamache et al. (2019) decided to retain the item for computing scale scores because of its relevance to forensic settings. Additionally, item 6 ("I recognize myself in how others describe me") showed low item-scale correlations but was not excluded from factor-analytic analyses subdomain based on the rationale that it conceptually represents the facet *accuracy of self-appraisal*. Confirmatory factor analytic investigations of the original French version, testing a one-factor model, a two-factor (self and interpersonal) model, a four-factor (ID, SD, EM, INT) model, a second-order model (four factors loading on a g-factor), and a bifactor model with four specific factors (all items loading on a g-factor and the four factors) found that the second-order model showed the best fit (good fit indices; Gamache et al., 2019).

Meanwhile, the SIFS has been applied in several further studies showing apparently promising results. These include relatively strong associations with Criterion B traits (except for antagonism; Leclerc et al., 2023; Roche & Jaweed, 2023), physical aggression (Leclerc et al., 2022), stalking (Gamache et al., 2023), resilience in police officers (Angehrn et al., 2023), discrimination of different severity profiles within patients with Borderline pathology (Gamache, Savard, Leclerc, Payant, Côté, et al., 2021), good content validity with respect to the construct definitions in the AMPD (Waugh et al., 2021), associations with emotional, behavioral, and thought problem risks in pregnant women requiring monitoring (Gamache et al., 2022), and diagnostical value in clinical practice (Samylkin et al., 2023). In addition, the SIFS has been discussed as a screener for assessing PDs according to ICD-11, whereby cutoff values for the five severity levels were empirically determined (Gamache, Savard, Leclerc, Payant, Berthelot, et al., 2021). Nevertheless, there are also problematic results of the original SIFS version, especially in terms of reverse-keyed items and specific item wordings, which has led to an adapted SIFS version, which is not psychometrically validated yet (Leclerc et al., 2022). So far, the SIFS has not been psychometrically evaluated in any other language than French. Therefore, it has not yet been possible to determine whether the problematic factor analytic aspects previously found by Gamache et al. (2019) and Leclerc et al. (2022) are language-specific or whether these issues are more fundamental and affect the instrument itself. Most evident, the SIFS has never been validated using standardized clinical interviews developed for Criterion A, indicating a fundamental gap in the current state of the evaluation of the SIFS.

The current study contributes to the literature on the SIFS by (1) testing the factor structure of the German SIFS items, (2) investigating the internal consistency and test-retest reliability of the scale scores, and (3) testing their convergent validity across three different PD interviews and various associated self-report questionnaires for personality functioning, personality organization, pathological personality trait domains, and well-being in a combined sample covering the full spectrum of personality functioning. Our overarching goal was to evaluate whether the German SIFS represents a self-report measure that is more fine-grained than the LPFS-BF 2.0 but less time-intensive than some other questionnaires assessing the severity of impairments in personality functioning. First, we tested whether the SIFS items indeed represent four differentiable domains as claimed by Gamache et al. (2019), or whether, as in the LPFS-BF 2.0, only the two overarching dimensions of self- and interpersonal functioning are distinguishable. Moreover, since the construct of personality functioning implies a strong general factor (e.g., Bliton et al., 2022; Zimmermann

et al., 2020), we addressed this question by comparing bifactor models with one general factor and two or four specific factors. Second, we expected that internal consistency and test-retest reliability of the SIFS sum score, as well as of the domain scores (except for SD; Gamache et al., 2019) are good. In the case that the four domains are reliably measurable beyond the general factor, they could be used to determine a focus for the therapeutic process, therefore increasing clinical utility (Hopwood et al., 2018). Third, we expected that convergent validity of the SIFS sum score is good as well, as indicated by large correlations with the total scores of the SCID-5-AMPD-I, STiP-5.1, and of the Structured Clinical Interview for DSM-5—PDs (SCID-5-PD; Beesdo-Baum et al., 2019a) categories as well as of self-report measures of personality functioning, personality organization, pathological personality trait domains, and well-being. For the SIFS domain scores, we additionally hypothesized whether they correlate more strongly with the corresponding domain scores of the measures used to test convergent validity (e.g., SIFS ID is expected to correlate more strongly with STiP-5.1 ID than SIFS EM). Note that our study is the first to examine the convergent validity of the SIFS score with a multimethod design, thus excluding common method bias as an explanation for potential large correlations.

Methods

Procedure and participants

Participants were enrolled in a multi-center study, a cooperation between two research groups on the AMPD (Berlin-Basel). The present study includes a total sample of $n = 886$. It is composed of three different subsamples, a German-speaking clinical sample from different cities in Germany ($n = 137$), a Swiss-German mixed clinical and non-clinical sample from Basel ($n = 116$), and a German-speaking non-clinical sample from the panel provider Clickworker ($n = 633$). In total, participants were aged between 18 and 66 years ($M = 37.2$, $SD = 12.5$), thereof 425 (48.0%) women. The participants' mean age was in the Basel sample: $M = 28.3$ years, $SD = 11.6$; in the German clinical sample: $M = 32.6$ years, $SD = 10.5$; and in the Clickworker sample: $M = 39.8$ years, $SD = 12.0$). Furthermore, 63.8% of the Basel, 70.1% of the German clinical sample, and 40.3% of the Clickworker sample were female, and 67.2, 76.6, and 88.8%, respectively were employed. Moreover, 11.2, 17.5, and 27.6%, respectively had children, 17.2, 35.7, and 45.9%, respectively had a university degree, and the majority did not take medication (51.7, 67.9, and 90.7%, respectively).

All participants of the multicenter study gave written informed consent. Two different ethics committees, the Northwestern and Central Swiss Ethics Committee and the Ethics Committee of Psychologische Hochschule Berlin, approved the study for the respective study sites. Participants were included in both clinical samples if they (1) were seeking psychotherapeutic treatment, (2) were at least 18 years old, and (3) had sufficient knowledge of German. Patients with acute suicidality, psychosis, mental retardation, or cognitive deficits due to substance or medication use were excluded. Total data collection took place between July 2020 and April 2022. All self-report questionnaires were collected online using the platform formr (Arslan et al., 2020). This study was not preregistered.

Basel mixed clinical and non-clinical sample

The Basel sample included patients of the outpatient psychotherapeutic department of the University Clinics Basel ($n = 87$) and nonclinical subjects (NC) of the University of Basel marketplace website ($n = 29$). Patients were recruited based on the clinic's standardized clinical procedures. Either they sought therapy by themselves or clinicians sent the patients for treatment. Psychiatric inpatients were already in inpatient, mostly acute treatment for some weeks. After an indicative preliminary interview based on the Operationalized Psychodynamic Diagnostic (OPD-2; OPD Task Force, 2008) in the outpatient clinic for psychotherapy and psychosomatics (ZPP Ambulanz, UPK), patients gave informed consent. In addition to the sociodemographic data and self-report questionnaires, patients and NCs underwent the STiP-5.1. In the Basel patient sample, three patients were previously excluded because of inconsistent responses, resulting in a sample of $n = 87$.[1] The average personality functioning impairment was as follows: 2.3% showed no impairment (STiP-5.1 total <0.5), 8.0% showed mild impairment (STiP-5.1 total ≥0.5 and <1.5), 62.1% showed moderate impairment (STiP-5.1 total ≥1.5 and <2.5), 27.6% showed severe impairment (STiP-5.1 total ≥2.5 and <3.5), and 0% showed extreme impairment (STiP-5.1 total ≥3.5). Overall, 56.3% met Criterion A for a PD according to the AMPD (moderate impairment (cutoff total score = 2.0) in personality functioning). The NC sample from Basel underwent the Structured Clinical Interview for DSM-5 Disorders (SCID-5-CV; Beesdo-Baum et al., 2019b) to exclude any mental disorders. In addition, NCs completed the Self-Report Personality Questionnaire for Structured Clinical Interview for DSM-5 (SCID-5-SPQ; Beesdo-Baum et al., 2019a) to exclude categorically defined PDs. Sixty-nine percent of NCs showed no impairments (STiP-5.1 total ≥0.5), and 31.0% showed mild impairment (STiP-5.1 total ≥0.5 and <1.5) in interview-assessed personality functioning. Data collection was anonymous; none of the Basel patients was financially compensated, but the NCs of the Basel sample received on-site financial compensation of 50 CHF.

German clinical sample

The German clinical part of the study included patients from different centers in Germany (mainly Berlin). Recruitment took place *via* outpatient psychotherapeutic departments ($n = 117$; 85.4%), inpatient psychiatric departments ($n = 16$; 11.7%), social media ($n = 1$; 0.7%), and *via* recommendation from study participants ($n = 3$; 2.2%). Patients were assessed by two interviews—the SCID-5-PD and the SCID-5-AMPD-I, as well as by self-report

[1]Inconsistent responses were determined by splitting the 24 SIFS items randomly in two 12-item parcels and computing the absolute difference between the two parcel mean scores for each person. An absolute difference equal or greater than two points was considered as indicating inconsistent responding.

questionnaires. Psychiatric inpatients were already in psychotherapeutic treatment for some weeks, while the remaining participants were seeking psychotherapy. None of the patients were excluded because of inconsistent responses (see definition of inconsistent responses in the Basel sample). Personality functioning impairments were as follows: 8.0% showed no impairment (SCID-5-AMPD-I total <0.5), 38.0% showed mild impairment (SCID-5-AMPD-I total ≥0.5 and <1.5), 23.4% showed moderate impairment (SCID-5-AMPD-I total ≥1.5 and <2.5), 9.5% showed severe impairment (SCID-5-AMPD-I total ≥2.5 and <3.5), and 2.2% showed extreme impairment (SCID-5-AMPD-I total ≥3.5). Overall, 33.3% met Criterion A for a PD according to the AMPD (moderate impairment (cutoff total score = 2.0) in personality functioning). The German participants received a small financial compensation, excluding the psychiatric inpatients. Data from the German clinical sample has been used for the validation of the German SCID-5-AMPD-I (see Ohse et al., 2023).

German Clickworker sample

A sample of 693 German-speaking subjects from the general population were recruited anonymously *via* the crowdsourcing platform Clickworker. Data was checked for inattentive responses, for which two attention questions were included (e.g., "This is a query to test your attention. Please click the second field from the left"). Fifty-three subjects (7.6%) answered at least one of the two questions incorrectly and were thus excluded. Another four participants (0.6%) were excluded because of questionable codes (e.g., TEST11). Finally, three participants (0.4%) were excluded because of inconsistent responses (see definition of inconsistent response in the Basel sample), resulting in a final sample of $n = 633$. The sample received financial compensation (three euros per questionnaire battery) through the Clickworker portal. For a retest assessment, 200 participants out of the first Clickworker population were invited to participate after a time interval of two weeks, of which 157 participants were identified *via* their code.

Measures

Participants in the different samples completed the various test batteries with a total of six self-report questionnaires. The Basel sample ($n = 116$) completed the SIFS in addition to the STiP-5.1 interview. The German clinical sample ($n = 137$) completed the SCID-5-AMPD-I interview as well as the categorical SCID-5-PD interview and two self-report questionnaires, the SIFS and LPFS-SR. The Clickworker sample ($n = 633$) completed the following self-report questionnaires: the SIFS, the LPFS-BF 2.0, the modified Personality Inventory for the DSM-5 Brief Form Plus (PID-BF + M; Bach et al., 2020), the Inventory of Personality Organization (IPO-16; Zimmermann et al., 2013), and the Brief Inventory of Thriving (BIT; Su et al., 2014). A subsample of the Clickworker sample ($n = 157$) completed the SIFS a second time after an interval of two weeks.

Semi-Structured Interview for Personality Functioning DSM-5 (STiP-5.1; Hutsebaut et al., 2017; German version translated by Zettl et al., 2019)

The STiP-5.1 assesses Criterion A on a scale from "0 = no impairment" to "4 = extreme impairment." It consists of one to four open interview questions for each of the twelve subdomains, e.g., "to what extent are you capable of really being and staying yourself?" In addition, help questions serve to clarify the information given by the participant. Further, there is the possibility to continue with check questions by summarizing and reformulating the participants' responses and adjusting the information to fit the subdomain's description. If still no assessment can be made, the interviewer can ask a test question by providing the participant with two options. The German version of the STiP-5.1 shows excellent interrater reliability (intraclass correlation coefficient, ICC[1,1] = .77; Zettl et al., 2019; Zimmermann, Hopwood, et al., 2023). In the patient sample, the STiP-5.1 was assessed by one of four psychodynamic psychologists with varying levels of experience and training. All psychologists were highly experienced in the assessment of PDs and had completed at least one OPD-2 training course (20 hr training), which includes the concept of structural integration. This covers different functional domains of psychological processes that can be differentiated along severity levels and thus show high relevance for Criterion A (Zimmermann et al., 2012). One psychologist had a brief training (2 hr by the developer Joost Hutsebaut) and taught the other psychologists about the AMPD and the structure of the STiP-5.1. In the NC sample, the STiP-5.1 was assessed by one of two MSc students who had a brief training (2 hr by the first author) and rated two videos independently before the study.

Structured Clinical Interview for the AMPD–Module I (SCID-5-AMPD-I; Bender et al., 2018; German version translated by Hörz-Sagstetter et al., in press)

The SCID-5-AMPD-I assesses Criterion A on a scale ranging from "0 = no impairment" to "4 = extreme impairment." At the beginning, the interviewer asks eight open screening questions and one to five screening questions for the twelve subdomains, respectively. After the participants have answered the screening questions, the interviewer can further elaborate the given questions in a standardized way based on the assessment of the severity of impairment. The SCID-5-AMPD-I has a funnel structure, i.e., the interviewer checks the preliminary assessment of the screening questions using a pool of standardized closed questions. Patients from the German clinical sample answered the interview questions with respect to the last five years. The German version of the SCID-5-AMPD-I shows excellent interrater (ICC[2,1] = .95) and test-retest reliability (ICC[1,1] = .84) for overall personality functioning (Ohse et al., 2023). For details on raters and interrater reliability see Ohse et al. (2023).

Structured Clinical Interview for DSM-5–PDs (SCID-5-PD; Beesdo-Baum et al., 2019a)

The SCID-5-PD assesses the ten categorical PDs of DSM-5 Section II. Each individual PD criterion can be rated on a

3-point scale, ranging from "0 = criterion not met" to "1 = criterion met subclinically "to "2 = criterion met." The SCID-5-PD shows overall excellent interrater reliability (ICC[1, 2] = .84) for the PD mean scores (Lobbestael et al., 2011).

Structured Clinical Interview for DSM-5 Disorders–Clinician Version (SCID-5-CV; Beesdo-Baum et al., 2019b)

The SCID-5-CV evaluates DSM-5 diagnoses according to ten modules. For each DSM-5 module, there are corresponding interview questions to help the interviewer assess the criteria for each disorder. The SCID-5-CV shows overall good interrater-reliability (kappa levels at least .70) for most diagnoses (Osório et al., 2019).

Self- and Interpersonal Functioning Scale (SIFS; Gamache et al., 2019; German version translated by two authors (CM and JW, 2020)

The SIFS assesses self-reported personality functioning according to Criterion A of the AMPD using 24 items rated on a 5-point Likert scale, ranging from "0 = This does not describe me at all" to "4 = This describes me totally." The SIFS was originally developed in French and subsequently translated into English by Gamache et al. (2019). The German translation was done based on the English version. The German translation of the SIFS was done by two authors (CM and JW), and the adequacy of the translated German version was verified through a back-translation by two independent bilingual English-German native speakers who had not seen the original English version. The two authors and the first author of the original SIFS (DG) then checked the back-translation. Regarding the psychometric properties, the internal consistency of the four personality functioning scales (Cronbach's α) was between .71 and .92, and test–retest reliability was between .63 and .92 (Gamache et al., 2019).

Level of personality functioning–self report (LPFS-SR; Morey, 2017; German version translated by Zimmermann et al., 2020)

The LPFS-SR is a self-report questionnaire, assessing the four intercorrelated domains of Criterion A using 80 items rated on a 4-point Likert scale, ranging from "1 = totally false" to "4 = very true." For the evaluation, items are weighted according to their severity within the LPFS conceptualization. Regarding the psychometric properties, the internal consistency (Cronbach's α) of the original version was between .80 and .92, and test–retest reliability was between .81 and .91 for the global score and the four domain scores (Hopwood et al., 2018).

Level of Personality Functioning–Brief Form 2.0 (LPFS-BF 2.0; Weekers et al., 2019; German version translated by Spitzer et al., 2021)

The LPFS-BF 2.0 assesses Criterion A using 12 items rated on a 4-point Likert scale ranging from "1 = completely untrue" to "4 = completely true." The German version of the LPFS-BF 2.0 shows good reliability (McDonalds $\omega \geq .83$) for the global scale and the two subscales (self- and interpersonal functioning; Spitzer et al., 2021).

Personality Inventory for DSM-5 Brief Form Plus Modified (PID5BF + M; Bach et al. 2020; earlier version developed by Kerber et al., 2022; German version translated by Zimmermann et al., 2014)

The PID5BF + M assesses the Criterion B domains negative affectivity, disinhibition, detachment, antagonism, and psychoticism as well as ICD-11 domain anankastia, using 36 items rated on a 4-point Likert scale ranging from "0 = very untrue or often untrue" to "3 = very true or often true." The German version of the PID5BF + M shows satisfactory reliability for all domain scores and a theoretically consistent structure of facet scores (see Bach et al., 2020).

Inventory of Personality Organization (IPO-16; Zimmermann et al., 2013)

The IPO-16 assesses impairments in personality organization (Kernberg & Caligor, 2005). The 16 items are rated on a 5-point Likert scale ranging from "1 = never true" to "5 = always true." Personality dysfunction is measured across three content domains (identity diffusion, primitive defense, and lack of reality testing). The IPO-16 total score shows excellent internal consistency (Cronbach's $\alpha = .91$) and excellent test-retest reliability of .85 over two months (Zimmermann et al., 2015).

Brief Inventory of Thriving (BIT; Su et al., 2014)

The BIT assesses health status of well-being (e.g., "I achieve most of my goals"). The ten items are rated on a 5-point Likert scale ranging from "1 = I totally disagree" to "5 = I totally agree." The BIT total score shows an internal consistency (Cronbach's α) of = .85 (Hausler et al., 2017).

Statistical analyses

The factor structure of the SIFS was investigated by testing a bifactor model with a general factor and two uncorrelated specific factors (self- and interpersonal; model 1), which would correspond to the structure of the LPFS-BF 2.0, and a bifactor model with a general factor and four uncorrelated specific factors (ID, SD, EM, INT; model 2). Therefore, a g-factor was set to explain covariances across all items, and specific factors were set orthogonally to explain shared variance among item clusters. Note that these models are less restrictive than many other confirmatory models (including correlated factor and hierarchical factor models with up to four factors), which is why, in the case of an insufficient fit, these more restrictive models would also appear implausible. Because reverse-keyed items could affect the factor structure, we explored models 1 and 2 with a method factor for reverse-keyed items (models 3 and 4, respectively), assuming that this might increase the fit indices. If the German version of the SIFS is able to capture the four specific domains,

models 2 or 4 should show good fit indices and a consistent loading pattern.

We used maximum likelihood estimation with robust standard errors and a Satorra-Bentler scaled test statistic (Satorra & Bentler, 2001). According to Rhemtulla et al. (2012) scales with five or more answer categories and approximately symmetric category thresholds can be treated as continuous. Nevertheless, robust fit indices are recommended to avoid biases due to non-zero skewness and kurtosis of item distributions (Brosseau-Liard et al., 2012; Brosseau-Liard & Savalei, 2014). Regarding confirmatory factor analysis (CFA), good model fit was assumed when the Root Mean Square Error of Approximation (RMSEA) was close to or below .06, the Standardized Root Mean Square Residual (SRMR) close to or below .08 and the Tucker-Lewis Index (TLI) as well as the Comparative Fit Index (CFI) were close to or above .95 (Hu & Bentler, 1999). Furthermore, we calculated the explained common variance (ECV), an index that can be interpreted as the proportion of common variance due to the g-factor. ECV > .60 indicates essential unidimensionality (Reise et al., 2013). Additionally, we also calculated ECV with respect to the specific factors.

To assess the internal consistency of the SIFS sum score, we used model-based Omega total (ω) and Omega Hierarchical (ω_H, Brunner et al., 2012); a $\omega_H > .70$ indicates unidimensionality (Reise et al., 2013). Additionally, we calculated ω_s and ω_{Hs} to assess the internal consistency of the specific SIFS domains; a $\omega_{Hs} > .50$ indicates the reliability of specific personality functioning factors (Reise et al., 2013). To calculate test-retest reliability after an interval of two weeks, we used bivariate zero-order correlations.

To investigate convergent validity, correlations between personality functioning measured with the SIFS and other self-report questionnaires of personality functioning (LPFS-BF 2.0), personality organization (IPO-16), pathological personality trait domains (PID5BF+M), and well-being (BIT) in the Clickworker sample as well as the self-report questionnaire for assessing personality functioning (LPFS-SR) in the German clinical sample were computed. For convergent validity analyses with interview-based assessments of personality functioning, correlations between the SIFS scales and the twelve subdomains, as well as the four domains of the LPFS (assessed with SCID-5-AMPD-I in the German clinical sample and with the STiP-5.1 in the Basel sample), were calculated. We further assessed convergent validity with the DSM-5 Section II PD model by correlating the SIFS scales with the ten specific PDs (dimensionally operationalized as the number of fulfilled criteria; see SCID-5-PD in the German clinical sample). According to Cohen (1992), we interpreted the size of correlation coefficients as follows: small correlations (= .10), medium correlations (= .30), and large correlations (= .50). For comparing the correlations between SIFS domain scores and external measures with each other we used Zou's confidence intervals (CIs) for dependent correlations (Zou, 2007). Following this approach, we reported correlation difference confidence intervals that did not include zero as significant.

The Basel and the Clickworker sample did not include participants with missing values. In the German clinical sample, missing values were as follows: $n_{\text{SCID-5-AMPD}} = 26$; $n_{\text{SCID-5-PD}} = 8$, $n_{\text{LPFS-SR}} = 7$. All the correlations were estimated with pairwise deletion. In line with West et al. (1995), we verified that the SIFS items were sufficiently normally distributed with a skewness <2 and kurtosis <7. For all the analyses we used the statistical software R version 4.1.0 (R Core Team, 2021) and the R packages psych (Revelle, 2020), lavaan (Rosseel et al., 2019), and tidyverse (Wickham & Wickham, 2017).

Results

SIFS, STiP-5.1, and SCID-5-AMPD-I domain and sum scores across the different samples can be found in Table 1. Personality functioning severity levels differed significantly, with the Basel clinical sample being the most impaired, followed by the German clinical sample, the Clickworker sample, and the Basel non-clinical sample. Thus, the combined sample covered the full spectrum of personality functioning.

Structure

The bifactor models did not fit the data well for the 24-item version of the SIFS (see Table 2). Note that this was also true for further CFA models following Gamache et al. (2019; see Supplementary Table 4). In addition, reversed item 6 and (non-reversed) item 10 consistently showed very low factor loading (<.3) in model 2. Items 6 and 10 were therefore excluded in the following analyses.

Estimating the bifactor models without items 6 and 10 increased the fit indices of all estimated models. Out of the two models, the bifactor model with two specific uncorrelated factors (model 1) showed the best, but still not good fit indices. Actually, the model achieved only good fit according to the cutoff for SRMR, while the remaining three fit indices were clearly not good (see Table 2). In addition, the loading pattern of the items was very heterogeneous: Reversed items 8, 19, and 24 showed very low factor

Table 1. Mean values of SIFS and interview-assessed personality functioning across the different samples.

	Samples			
	Basel clinical ($n = 87$)	Basel non-clinical ($n = 29$)	German clinical ($n = 137$*)	Click-worker ($n = 633$)
Assessments	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
SIFS				
Sum	1.99 (0.6)	0.72 (0.3)	1.87 (0.6)	1.38 (0.6)
Identity	2.55 (0.8)	0.63 (0.5)	2.14 (0.9)	1.15 (0.8)
Self-direction	2.13 (0.7)	0.88 (0.6)	2.64 (0.8)	1.31 (0.7)
Empathy	1.14 (0.7)	0.44 (0.4)	1.21 (0.6)	1.04 (0.7)
Intimacy	1.55 (0.8)	0.48 (0.3)	1.75 (0.7)	1.37 (0.8)
Interviews				
Sum	2.10 (0.6)	0.36 (0.3)	1.49 (0.9)	–
Identity	2.43 (0.6)	0.50 (0.4)	1.80 (1.1)	–
Self-direction	2.16 (0.7)	0.23 (0.3)	1.62 (0.9)	–
Empathy	1.84 (0.7)	0.41 (0.4)	1.13 (1.0)	–
Intimacy	1.96 (0.7)	0.29 (0.4)	1.41 (1.0)	–

SIFS: Self and Interpersonal Functioning Scale; Interview for the Basel clinical and non-clinical sample: Semi-Structured Interview for Personality Functioning DSM-5; Interview for the German clinical sample: Structured Clinical Interview for the Alternative DSM-5 Model for Personality Disorders–Module I.
Notes. *$n = 111$ for the SCID-5-AMPD-I of the German clinical sample.

Table 2. Test statistic and fit indices of CFA of the 24-item, 22-item (without items 6 and 10), and method factor for reversed items version of the SIFS.

Model	χ^2 (*df*)	CFI	TLI	RMSEA	SRMR
24-Item version					
1	1271.804 (228)	.857	.827	.079	.080
2	1459.368 (228)	.832	.796	.085	.081
22-Item version (without item 6 and 10)					
1	1065.231 (187)	.873	.843	.081	.073
2	1250.626 (187)	.847	.811	.088	.080
22-Item version (without item 6 and 10) and a method factor for reversed items					
3	860.109 (181)	.903	.876	.072	.063
4	1007.845 (181)	.882	.850	.079	.075

CFI: comparative fit index; RMSEA: root mean square error of approximation; SRMR: standardized root mean residual; TLI: Tucker-Lewis Index. Model 1: a bifactor model with two uncorrelated specific factors (representing self- and interpersonal functioning). Model 2: a bifactor model with four uncorrelated specific factors (representing identity, self-direction, empathy, and intimacy). Model 3: a bifactor model with two uncorrelated factors, with the items loading on the two specific factors (self and interpersonal) and on a general personality pathology factor including a method factor for reversed items. Model 4: a bifactor model with four uncorrelated specific factors (ID, SD, EM, INT), with items loading on the four specific factors and a general personality pathology factor including a method factor for reversed items.

Notes. $N = 886$. All values are according to the robust indices. All *p*-values <.000.

loadings (<.3) on the g-factor while for the specific factor interpersonal functioning, items 18 and 21 showed negative factor loadings. In contrast, the bifactor model with four specific uncorrelated factors (model 2) showed a more homogenous loading pattern for the g-factor, with an ECV of the g-factor (59.6%) and the reliability of the sum score ($\omega = 92, \omega_H = .81$) being higher than for model 1 (ECV = 52.9%, $\omega = .92$, $\omega_H = .68$). Following these findings, model 2 might be preferred despite its worse fit. Item loadings of this model solution are presented in Figure 1. In this model, the internal consistency for the four domains was poor for ω_{Hs} (ID $\omega_s = .88$, $\omega_{Hs} = .31$; SD $\omega_s = .73$, $\omega_{Hs} = .22$; EM $\omega_s = .79$, $\omega_{Hs} = .25$; and INT $\omega_s = .81$, $\omega_{Hs} = .32$). The specific-factor ECV was 12.8% for ID, 6.4% for SD, 8.0% for EM and 13.2% for INT. Factor loadings on the g-factor and specific factors were all positive, but some reversed items (8, 17, 19, 24) and item 16 showed factor loadings <.40 on the g-factor, and item 21 showed a negative factor loading on the specific factor INT. Estimating the CFA models with a method factor for reversed items increased fit indices (see Table 2), but did not improve the loading pattern.

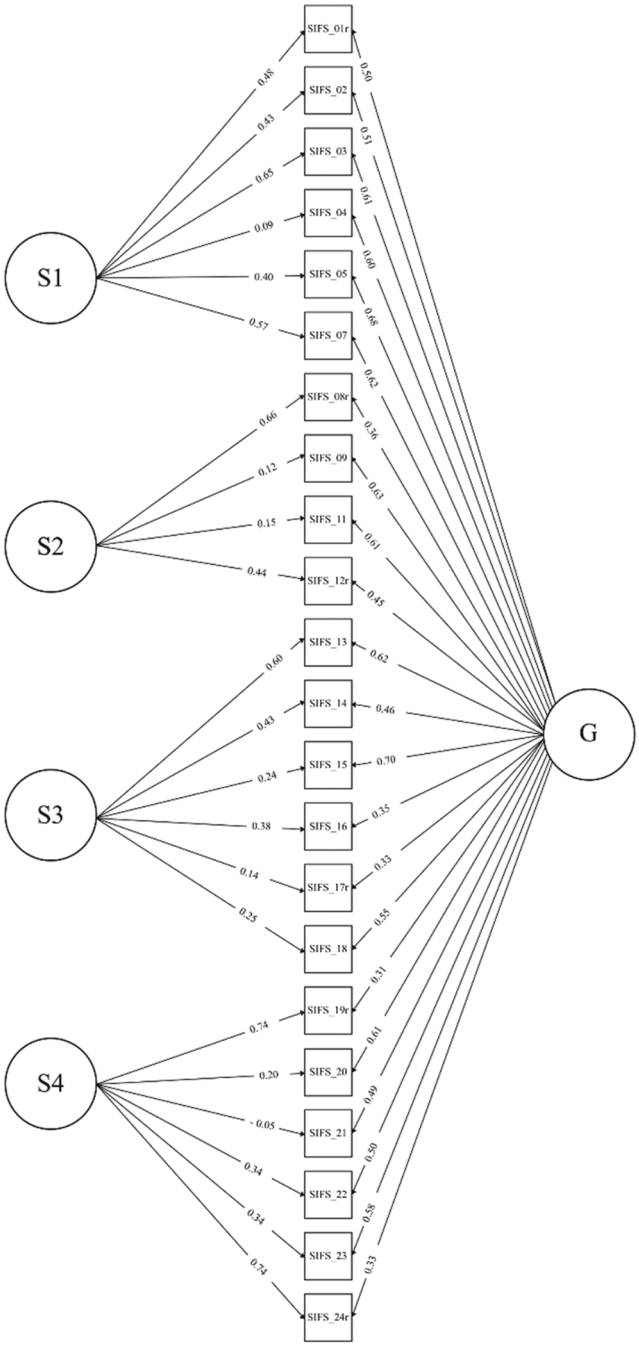

Figure 1. Standardized factor loadings of confirmatory factor analysis of the 22-item version of the SIFS: Bifactor model with four specific uncorrelated factors. *Notes.* $N = 886$. G: general factor (personality functioning); S1: specific factor 1 (identity); S2: specific factor 2 (self-direction); S3: specific factor 3 (empathy); S4: specific factor 4 (intimacy); SIFS: Self and Interpersonal Functioning Scale, representing the four domains of the Level of Personality Functioning Scale.

Test-retest reliability

For the SIFS sum score (22 items) the Pearson correlation for test-retest after two weeks ($n = 157$) was $r = .86$ (95% CI [.81, .89]). At the level of domains, test-retest correlation was $r = .85$ (95% CI [.80, .89]) for ID, $r = .72$ (95% CI [.64, .79]) for SD, $r = .77$ (95% CI [.69, .83]) for EM, and $r = .78$ (95% CI [.71, .84]) for INT. All correlations were statistically significant ($p < .001$).

Convergent validity

Correlations for convergent validity analyses can be found in Table 3. In the Clickworker sample ($n = 633$), correlations between the SIFS sum score (22-item version) and the associated measures were large, including LPFS-BF 2.0 ($r = .82$, $p < .001$) and IPO-16 ($r = .76$, $p < .001$). Additionally, the correlations of the SIFS sum score with the pathological personality trait domain scores were large (ranging from .50 to .73, $p < .001$), with the exception of PID5BF + M anankastia ($r = .38$, $p < .001$). The correlation between the SIFS sum score and the well-being construct was largely negative

Table 3. Correlations between SIFS scales (22-item version without items 6 and 10) and LPFS-BF 2.0, IPO-16, PID5BF + M, BIT subscale scores ($n = 633$); LPFS-SR ($n = 130$); interview-based LPFS scores according to STiP-5.1 and SCID-5-AMPD-I ($n = 253$); and SCID-5-PD ($n = 129$).

	SIFS				
Assessments	ID	SD	EM	INT	Sum
Self-report questionnaires					
LPFS-SR sum	**.74**	**.69**	**.61**	**.66**	**.87**
LPFS-SR ID	**.78**	**.70**	**.52**	**.55**	**.82**
LPFS-SR SD	**.68**	**.66**	**.51**	**.52**	**.76**
LPFS-SR EM	**.55**	**.55**	**.76**	**.57**	**.76**
LPFS-SR INT	**.56**	**.50**	**.49**	**.71**	**.73**
LPFS-BF 2.0 sum	**.82**	**.64**	**.62**	**.63**	**.82**
LPFS-BF 2.0 self	**.84**	**.64**	**.50**	**.51**	**.75**
LPFS-BF 2.0 interpersonal	**.63**	**.52**	**.66**	**.65**	**.75**
IPO-16 sum	**.70**	**.61**	**.67**	**.52**	**.76**
PID-5 Neg. affectivity	**.67**	**.47**	**.36**	**.37**	**.57**
PID-5 detachment	**.58**	**.50**	**.59**	**.72**	**.73**
PID-5 antagonism	**.34**	**.40**	**.54**	**.39**	**.50**
PID-5 disinhibition	**.57**	**.63**	**.52**	**.41**	**.63**
PID-5 psychoticism	**.56**	**.47**	**.57**	**.46**	**.63**
PID-5 anankastia	**.33**	**.18**	**.35**	**.34**	**.38**
BIT sum	**−.67**	**−.57**	**−.40**	**−.58**	**−.67**
Interview-based LPFS (SCID-5-AMPD-I and STiP-5.1)					
Global score	**.69**	**.62**	**.55**	**.63**	**.76**
Self-domain	**.72**	**.65**	**.50**	**.58**	**.75**
Identity	**.75**	**.62**	**.52**	**.60**	**.76**
Sense of self	**.68**	**.54**	**.44**	**.54**	**.67**
Self-esteem	**.68**	**.58**	**.52**	**.51**	**.72**
Emotions	**.55**	**.55**	**.32**	**.57**	**.57**
Self-direction	**.65**	**.62**	**.45**	**.52**	**.68**
Goals	**.55**	**.55**	**.32**	**.43**	**.57**
Standards	**.60**	**.56**	**.43**	**.46**	**.63**
Self-reflection	**.59**	**.56**	**.45**	**.50**	**.64**
Interpersonal domain	**.60**	**.55**	**.57**	**.63**	**.71**
Empathy	**.55**	**.52**	**.57**	**.55**	**.66**
Comprehension	**.56**	**.49**	**.49**	**.53**	**.63**
Tolerance	**.41**	**.40**	**.54**	**.48**	**.55**
Effects	**.50**	**.50**	**.52**	**.48**	**.60**
Intimacy	**.59**	**.53**	**.51**	**.65**	**.70**
Connection	**.50**	**.43**	**.44**	**.57**	**.59**
Closeness	**.61**	**.48**	**.46**	**.60**	**.67**
Mutuality	**.47**	**.48**	**.46**	**.56**	**.59**
SCID-5-PD					
ASPD	.22	**.24**	**.47**	**.32**	**.39**
AVPD	**.53**	**.44**	**.23**	**.39**	**.52**
BPD	**.67**	**.54**	**.36**	**.37**	**.63**
DPD	**.48**	**.43**	.18	**.23**	**.43**
HPD	**.29**	**.28**	**.37**	**.27**	**.39**
NPD	**.25**	**.23**	**.52**	**.29**	**.40**
OCPD	.16	−.01	.17	.13	.15
PPD	**.38**	**.36**	**.50**	**.44**	**.53**
SIPD	.19	.22	**.39**	**.35**	**.36**
STPD	**.31**	**.35**	**.51**	**.45**	**.51**

SIFS: Self- and Interpersonal Functioning Scale; ID: identity; SD: self-direction; EM: empathy, INT: intimacy; LPFS-BF 2.0: Level of Personality Functioning Scale–Brief Form 2.0; LPFS-SR: Level of Personality Functioning–Self Report; IPO-16: Inventory of Personality Organization–16 item version; PID-5BF + M: Personality Inventory for DSM-5 Brief Form–Modified; BIT: Brief Inventory of Thriving; StiP-5.1: Semi-Structured Interview for Personality Functioning DSM-5; SCID-5-AMPD-I: Structured Clinical Interview for the Alternative DSM-5 Model for Personality Disorders–Module I; SCID-5-PD: Structured Clinical Interview for DSM-5 Personality Disorders (PD); AVPD: avoidant PD; DPD: dependent PD; OCPD: obsessive-compulsive PD; PPD: paranoid PD; STPD: schizotype PD; SIPD: schizoid PD; HPD: histrionic PD; NPD: narcissistic PD; BPD: borderline PD; ASPD: antisocial PD.

Notes. Statistically significant *p*-values <.001 are marked in bold.

($r = -0.67$, $p < .001$). The correlation matrix for the convergent validity analyses (conducted with the 22-item version of the SIFS) in the Clickworker sample can be found in Table 3.

To calculate the correlation between the SIFS sum score and the dimensional interviews (STiP-5.1 and SCID-5-AMPD-I), we merged the two interview-samples (in detail, 116 STiP-5.1 and 111 SCID-5-AMPD-I interviews) because both included the same twelve LPFS subdomain ratings. Overall ($n = 227$), the SIFS sum score correlated strongly with all four interview-based LPFS domains (between .66 and .76, $p < .001$), correlating largest with the ID domain ($r = .76$, $p < .001$). The correlation matrix of all the interview-based LPFS ratings (with the 22-item version of the SIFS) can be found in Table 3. Separate correlations of the SIFS with the STiP-5.1 and SCID-5-AMPD-I show no substantial differences (SIFS sum and STiP-5.1 total = .78; SIFS sum and SCID-5-AMPD-I total = .74) and can be found in Supplementary Tables 2 and 3, respectively.

Correlations between the SIFS sum score and DSM-5 Section II PDs (according to SCID-5-PD) were small to medium, with the exception of borderline (BPD, $r = .63$, $p < .001$), paranoid ($r = .53$, $p < .001$), avoidant ($r = .52$, $p < .001$) and schizotypal ($r = .51$, $p < .001$) PD. The smallest correlation was between the SIFS sum score and obsessive-compulsive

PD (OCPD, $r = .15$, $p = .08$). All correlations with OCPD were not significant.

Regarding the four domains of the SIFS, our data showed that the ID domain correlated (compared with the other three domains) more strongly with the sum score of LPFS-BF 2.0 ($r = .82$, $p < .001$) and the BIT ($r = -0.67$, $p < .001$). The SIFS ID domain correlated also significantly more strongly with the corresponding domain in the AMPD interview (STiP-5.1 and SCID-5-AMPD-I merged; $r = .75$, $p < .001$). Regarding the correlations of the four SIFS domain scores and Criterion B, ID correlated significantly more strongly with negative affectivity ($r = .70$, $p < .001$); SD with disinhibition ($r = .56$, $p < .001$); and INT with detachment ($r = .72$, $p < .001$). The correlations between the SIFS domains and DSM-5 Section II PDs showed that the ID domain correlated more strongly with BPD ($r = .67$, $p < .001$). The remaining differences between the correlations of the SIFS domain scores with external measures were not significant according to Zou (2007). For more details on convergent validity for the 22-item version of the SIFS, see Table 3.

Discussion

This study reports the psychometric properties of the German translation of the SIFS for assessing the level of personality functioning according to Criterion A of the AMPD. In line with Gamache et al. (2019), the German SIFS sum score showed promising results regarding test-retest reliability and convergent validity with self-report questionnaires. Additionally, for the first time, this study showed large correlations between the SIFS sum score and the scores of interview-based measures of personality functioning. However, the psychometric structure of the SIFS appears to be rather complex and not in line with theoretical considerations.

In particular, we were not able to establish an appropriate bifactor model for the German 24-item version of the SIFS. This is in contrast to other Criterion A measures (e.g., LPFS-SR, LPFS-BF), which have been shown to conform to a bifactor structure in which a strong g-factor explains the majority of variance and specific factors represent only little to no variance, consistent with the notion that personality functioning is an essentially unidimensional construct (Bliton et al., 2022). In our results, a bifactor model with four uncorrelated specific factors (model 2) indicated less than acceptable fit and included two SIFS items with questionably low factor loadings. These items were the reversed item 6 ("I recognize myself in the way other people describe me") and the rather complicated-worded item 10 ("My actions and decisions are determined by my immediate needs, independently of everything else"). Interestingly, item 6 had already shown problematic psychometric properties in the French version but was still retained (Gamache et al., 2019). An iteration of the CFA models excluding items 6 and 10, a bifactor model with two or four uncorrelated, specific factors (model 1 and 2), also did not show good fit but could perhaps still be interpreted as close to acceptable. Note that, besides model fit, a well-established bifactor model should also consist of substantial loadings on the g-factor, adequate reliability (i.e., ω, ω_H), as well as reasonable ECV (Watts et al., 2019). In comparison to other Criterion A instruments (e.g., Morey, 2017; Weekers et al., 2019), the ECV of the g-factor and ω_H were at least acceptable for the 22-item version when adopting model 2, suggesting that the 22-item SIFS may measure a sufficient amount of general personality functioning variance. The problem with the bifactor model, however, is most evident in the loading pattern. There are significant variations in the standardized loadings on the g-factor in model 2, i.e., five items loaded <.40 on the g-factor, some loadings on the specific factors were very small (<.3) and item 21 showed a negative loading. Moreover, the poor ω_{Hs} for the specific factors confirm the low reliable variance beyond the g-factor. Therefore, we do not recommend the 22-item SIFS for the investigation of specific personality functioning domains, but to use the sum score for a global impression of personality functioning (see implications regarding the scoring system in Supplementary Table 1).

One reason for the unsatisfying psychometric properties may be the inclusion of reversed items within the SIFS. Consequently, there is a risk of reduced reliability when non-reversed and reversed items are included in the same test, as the secondary sources of variance may compromise the unidimensionality of the test (e.g., due to careless responding; Woods, 2006). In line with this, Leclerc et al. (2022) developed a revised version of the original SIFS with 20-items and excluded items with reversed wordings. However, combining non-reversed and reversed items in a test may safeguard against other types of response bias (e.g., acquiescence) and improve the coverage of the domains' content. We addressed the complexities of reversed items by including a method factor in the CFA models. While adding this method factor increased the model fit, the ECV, internal consistency, and factor loadings were unchanged, indicating that besides the reversed items, the German version of the SIFS shows problematic wording compromising its psychometric properties. These mentioned problems do not seem to be unique to the German version but represent fundamental problems of the instrument itself, implying that there are some unresolved issues with the original version of the SIFS.

Although some of the SIFS items showed low factor loadings in the bifactor model (see Figure 1), correlations of the sum score with other self-report questionnaires on Criterion A and the psychodynamic construct of personality organization were still large. This could indicate large convergent validity, but may also result from shared unspecified variance, including momentary distress or common method bias (Podsakoff et al., 2003). Thus, an important strength of this study was demonstrating that the convergent validity of the SIFS sum score was also large when using two (semi-)structured Criterion A interviews, thereby ruling out common method bias. These results are in line with recent studies, which have shown that self-rated and interview-assessed personality functioning can be strongly correlated (e.g., Heissler et al., 2021; Ohse et al., 2023; Somma et al., 2020). Additionally, our results showed some significantly stronger

correlations between the SIFS ID domain external criteria. These results may suggest that self-reported identity disturbances are a particularly strong marker of impaired functioning (e.g., LPFS-BF2.0 sum score). On the other hand, the SIFS ID domain also correlated more strongly (compared to the other three domains) with negative affectivity, suggesting that the ID domain may lack discriminant validity with regard to trait models of personality and personality pathology (Oltmanns & Widiger, 2016). Nevertheless, convergent validity of the specific SIFS domains did not show distinct patterns of correlations with the corresponding external criteria, as we had originally assumed. This substantiates the conclusion that the use of domain scores does not provide much additional benefit.

In line with previous SIFS studies concerning the convergent validity with Criterion B constructs, the SIFS sum score showed large correlations with all pathological personality trait domains (e.g., Gamache et al., 2019; Waugh et al., 2021), except for the additional ICD-11 trait domain anankastia. This is in line with previous studies (McCabe & Widiger, 2020; Zimmermann, Falk, et al., 2023) showing a weak relationship between measures of anankastia and Criterion A. We argue that the relationship between Criterion A and anankastia may be complex, because the construct of anankastia may have adaptive facets that are not necessarily dysfunctional (e.g., deliberativeness) or measures of anankastia may be formulated too adaptively and therefore do not represent the actually maladaptive character of anankastia (Zimmermann, Falk, et al., 2023). Regarding DSM-5 Section II PDs, BPD showed the largest correlations with SIFS sum score across all PD categories, supporting the hypothesis that BPD is a particularly strong marker of general impairments of personality pathology (Sharp et al., 2015). Considering the remaining PD categories, we found that OCPD was almost uncorrelated with the SIFS scores, which parallels the rather small correlations with the phenotypic expression of anankastia (ICD-11 trait). Moreover, the large negative correlation between the SIFS sum and the construct of well-being can be reconciled with the results of low life satisfaction found by Gamache et al. (2019).

Comparing the SIFS with existing self-report measures of Criterion A(e.g., LPFS-SR, LPFS-SRA, LPFS-BF 2.0, DLOPFQ, and DLOPFQ-SF), our study revealed no advantages with respect to its psychometric structural properties. Just as the LPFS-BF (Hutsebaut et al., 2016) was revised (LPFS-BF 2.0) due to its low internal consistency (three items were problematic and therefore reformulated), our study advocates a fundamental revision of the problematic SIFS items (in terms of wording and item reversion) and a careful psychometric evaluation in a multi-method design.

Limitations

The present study has several limitations. First, we translated the SIFS into German according to a standardized back-translation process using an English version provided by Dominique Gamache with two independent bilingual and single-blinded translators. However, the validation of the original SIFS by Gamache et al. (2019) is based on the French version. Nevertheless, we double-checked the German version with the French version and found no significant discrepancies. Second, the Clickworker sample lacked an interview for clinical assessment; therefore, it cannot be ensured that this sample is entirely nonclinical. Moreover, we did not assess clinical diagnoses (other than PDs) for the clinical samples and the NC data (Clickworker and Basel NCs) were not representative of the general population. Third, the German clinical and the Basel sample did not include a careless responding variable in the self-report battery. Fourth, at the Basel study site, we did not investigate the inter-rater reliability between the STiP-5.1 interviewers, however, all interviewers were experienced clinicians trained in dimensionally assessing personality dysfunction with axis IV of the OPD-2. Fifth, we assessed only a small subsample ($n = 157$) for test-retest analyses. Sixth, we did not investigate the discriminant validity of the SIFS sum score. However, because most clinical constructs will be positively correlated with Criterion A, investigating discriminant validity seems challenging (Zimmermann, Hopwood, et al., 2023). Seventh, our results suggest modifications to the item content itself. These are currently taken into account in the development of an adapted version of the SIFS, which was not yet available during our data collection.

Conclusion

The present validation study examined the structure, reliability, and convergent validity of the SIFS. Following our analyses, a bifactor model with one general and four specific uncorrelated factors (ID, SD, EM, INT) seemed most suitable for a shortened 22-item version (without items 6 and 10) of the SIFS, with the g-factor explaining 59.6% of the common variance and showing the reliability of .81. However, fit indices were not fully acceptable and the items' loadings on the g-factor differed considerably, suggesting that several (especially reverse-coded) items cannot be easily integrated into a reliable assessment of personality functioning with the German version of the SIFS. Moreover, factor loadings and explained variance of specific factors were rather small, questioning the utility of domain scores beyond the sum score. Nevertheless, test-retest reliability and convergent validity with other well-established self-reports for Criterion A and personality organization were large for the sum score of the SIFS. In addition, our results show that the SIFS sum score has large convergence with two (semi-)structured interviews measuring Criterion A. Taken together, our study on the psychometric investigation of the German SIFS shows mixed results, and warrants modifications for the use in research (i.e., regarding problematically worded or reversed items), confirming the results found by Leclerc et al. (2022). While in the clinical context, the current 22-item German version may serve as a brief measure for the general severity of PD, it seems currently more advisable to use self-report questionnaires with higher structural validity such as the LPFS-BF 2.0.

Acknowledgments

We acknowledge all interviewers and subjects who participated in our research.

Ethical approval

The Northwestern and Central Swiss Ethics Committee and the Ethics Committee of the Psychologische Hochschule Berlin approved this study independently for the respective study sites.

Disclosure statement

The authors declare there are no conflicts of interest.

Funding

Fundings were received from Schweizerischer Nationalfond (Johannes Wrege), Deutsche Psychoanalytische Gesellschaft (Ludwig Ohse), and Köhler-Stiftung im Stifterverband für die Deutsche Wissenschaft (Susanne Hörz-Sagstetter).

ORCID

Caroline Macina http://orcid.org/0000-0002-3861-0539
André Kerber http://orcid.org/0000-0002-8588-7784
Johannes Zimmermann http://orcid.org/0000-0001-6975-2356
Ludwig Ohse http://orcid.org/0000-0001-6214-2491
Leonie Kampe http://orcid.org/0000-0002-2096-7820
Susanne Hörz-Sagstetter http://orcid.org/0000-0001-6051-5207
Johannes Sebastian Wrege http://orcid.org/0000-0003-3538-2881

Data availability statement

The data of this study are available from the first author, upon reasonable request.

References

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Author.

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

Angehrn, A., Jourdan-Ionescu, C., & Gamache, D. (2023). Resilience among police officers: The role of personality functioning and protective factors. *Policing: An International Journal*, *46*(3), 567–582. https://doi.org/10.1108/PIJPSM-02-2023-0025

Arslan, R. C., Walther, M. P., & Tata, C. S. (2020). formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. *Behavior Research Methods*, *52*(1), 376–387. https://doi.org/10.3758/s13428-|019-01236-y

Bach, B., Kerber, A., Aluja, A., Bastiaens, T., Keeley, J., Claes, L., Fossati, A., Gutierrez, F., Oliveira, S. E. S., Pires, R., Riegel, K. D., Rolland, J., Roskam, I., Sellbom, M., Somma, A., Spanemberg, L., Strus, W., Thimm, J., 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

Beesdo-Baum, K., Zaudig, M., & Wittchen, H. U. (2019a). *SCID-5-PD. Strukturiertes Klinisches Interview für DSM-5®–Persönlichkeitsstörungen. Deutsche Bearbeitung des* [*Structured Clinical Interview for DSM-5®–Personality Disorders von Michael B. First, Janet BW Williams, Lorna Smith Benjamin, Robert L. Spitzer]*. Hogrefe.

Beesdo-Baum, K., Zaudig, M., & Wittchen, H. U. (Eds.). (2019b). *SCID-5-CV Strukturiertes Klinisches Interview für DSM-5-Störungen–Klinische Version: Deutsche Bearbeitung des* [*Structured Clinical Interview for DSM-5 Disorders–Clinician Version von Michael B. First, Janet BW Williams, Rhonda S. Karg, Robert L. Spitzer]*. Hogrefe.

Bender, D. S., Morey, L. C., & Skodol, A. E. (2011). Toward a model for assessing level of personality functioning in DSM–5, part I: A review of theory and methods. *Journal of Personality Assessment*, *93*(4), 332–346. https://doi.org/10.1080/00223891.2011.583808

Bender, D., Skodol, A., First, M., & Oldham, J. (2018). *Module I: Structured Clinical Interview for the Level of Personality Functioning Scale*. American Psychiatric Association.

Birkhölzer, M., Schmeck, K., & Goth, K. (2021). Assessment of Criterion A. *Current Opinion in Psychology*, *37*, 98–103. https://doi.org/10.1016/j.copsyc.2020.09.009

Bliton, C. F., Roche, M. J., Pincus, A. L., & Dueber, D. (2022). Examining the structure and validity of self-report measures of DSM-5 alternative model for personality disorders Criterion A. *Journal of Personality Disorders*, *36*(2), 157–182. https://doi.org/10.1521/pedi_2021_35_531

Brosseau-Liard, P. E., & Savalei, V. (2014). Adjusting incremental fit indices for nonnormality. *Multivariate Behavioral Research*, *49*(5), 460–470. https://doi.org/10.1080/00273171.2014.933697

Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An investigation of the sample performance of two nonnormality corrections for RMSEA. *Multivariate Behavioral Research*, *47*(6), 904–930. https://doi.org/10.1080/00273171.2012.715252

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*(4), 796–846. https://doi.org/10.1111/j.1467-6494.2011.00749x

Clark, L. A. (2007). Assessment and diagnosis of personality disorder: Perennial issues and an emerging reconceptualization. *Annual Review of Psychology*, *58*(1), 227–257. https://doi.org/10.1146/annurev.psych.57.102904.190200

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037//0033-2909.112.1.155

Gamache, D., Cloutier, M. È., Faucher, J., Leclerc, P., & Savard, C. (2023). Stalking perpetration through the lens of the alternative DSM-5 model for personality disorders. *Personality and Mental Health*, *17*(2), 135–146. https://doi.org/10.1002/pmh.1567

Gamache, D., Savard, C., Leclerc, P., & Côté, A. (2019). Introducing a short self-report for the assessment of DSM–5 level of personality functioning for personality disorders: The Self and Interpersonal Functioning Scale. *Personality Disorders*, *10*(5), 438–447. https://doi.org/10.1037/per0000335

Gamache, D., Savard, C., Leclerc, P., Payant, M., Berthelot, N., Côté, A., Faucher, J., Lampron, M., Lemieux, R., Mayrand, K., Nolin, M.-C., & Tremblay, M. (2021). A proposed classification of ICD-11 severity degrees of personality pathology using the Self and Interpersonal Functioning Scale. *Frontiers in Psychiatry*, *12*, 628057. https://doi.org/10.3389/fpsyt.2021.628057

Gamache, D., Savard, C., Leclerc, P., Payant, M., Côté, A., Faucher, J., Lampron, M., & Tremblay, M. (2021). Latent profiles of patients with borderline pathology based on the alternative DSM-5 model for personality disorders. *Borderline Personality Disorder and Emotion Dysregulation*, *8*(1), 1–13. https://doi.org/10.1186/s40479-021-00146-w

Gamache, D., Savard, C., Lemieux, R., & Berthelot, N. (2022). Impact of level of personality pathology on affective, behavioral, and thought problems in pregnant women during the coronavirus disease 2019 pandemic. *Personality Disorders*, *13*(1), 41–51. https://doi.org/10.1037/per0000479

Hausler, M., Huber, A., Strecker, C., Brenner, M., Höge, T., & Höfer, S. (2017). Validierung eines Fragebogens zur umfassenden Operationalisierung von Wohlbefinden. *Diagnostica*, *63*(3), 219–228. https://doi.org/10.1026/0012-1924/a000174

Heissler, R., Doubková, N., Hutsebaut, J., & Preiss, M. (2021). Semi-structured Interview for Personality Functioning DSM-5 (STiP-5.1): Psychometric evaluation of the Czech version. *Personality and Mental Health*, *15*(3), 198–207. https://doi.org/10.1002/pmh.1508

Hopwood, C. J., Good, E. W., & Morey, L. C. (2018). Validity of the DSM–5 levels of personality functioning scale–Self report. *Journal of*

Personality Assessment, *100*(6), 650–659. https://doi.org/10.1080/00223891.2017.1420660

Hörz-Sagstetter, S., Mokros, A., & Zimmermann, J. (in press). *Strukturiertes Klinisches Interview für das DSM–5 – Alternatives Modell für Persönlichkeitsstörungen (SCID-5-AMPD)* [*German adaptation of the Structured Clinical Interview for the DSM–5 – Alternative Model for Personality Disorders by Michael B. First, Andrew E. Skodol, Donna S. Bender, and John M. Oldham]*. Hogrefe.

Hu, L. t., & 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

Huprich, S. K., Nelson, S. M., Meehan, K. B., Siefert, C. J., Haggerty, G., Sexton, J., Dauphin, V. B., Macaluso, M., Jackson, J., Zackula, R., & Baade, L. (2018). Introduction of the DSM-5 levels of Personality Functioning Questionnaire. *Personality Disorders*, *9*(6), 553–563. https://doi.org/10.1037/per0000264

Hutsebaut, J., Feenstra, D. J., & Kamphuis, J. H. (2016). Development and preliminary psychometric evaluation of a brief self-report questionnaire for the assessment of the DSM–5 level of Personality Functioning Scale: The LPFS brief form (LPFS-BF). *Personality Disorders*, *7*(2), 192–197. https://doi.org/10.1037/per0000159

Hutsebaut, J., Kamphuis, J. H., Feenstra, D. J., Weekers, L. C., & De Saeger, H. (2017). Assessing DSM–5-oriented level of personality functioning: Development and psychometric evaluation of the Semi-Structured Interview for Personality Functioning DSM–5 (STiP-5.1). *Personality Disorders*, *8*(1), 94–101. https://doi.org/10.1037/per0000197

Kerber, A., Schultze, M., Müller, S., Rühling, R. M., Wright, A. G., Spitzer, C., Krueger, R. F., Knaevelsrud, C., & Zimmermann, J. (2022). Development of a short and ICD-11 compatible measure for DSM-5 maladaptive personality traits using ant colony optimization algorithms. *Assessment*, *29*(3), 467–487. https://doi.org/10.1177/1073191120971848

Kernberg, O. F., & Caligor, E. (2005). A psychoanalytic theory of personality disorders. In M. F. Lenzenweger & J. F. Clarkin (Eds.), *Major theories of personality disorder* (2nd ed., pp. 114–156). Guilford Press.

Leclerc, P., Savard, C., Sellbom, M., Côté, A., Nolin, M.-C., Payant, M., Roy, D., & Gamache, D. (2023). Investigating the validity and measurement invariance of the Personality Inventory for DSM-5 Faceted Brief Form among French-speaking clinical and nonclinical samples. *Journal of Psychopathology and Behavioral Assessment*, *45*(2), 519–536. https://doi.org/10.1007/s10862-022-10000-0

Leclerc, P., Savard, C., Vachon, D. D., Faucher, J., Payant, M., Lampron, M., Tremblay, M., & Gamache, D. (2022). Analysis of the interaction between personality dysfunction and traits in the statistical prediction of physical aggression: Results from outpatient and community samples. *Personality and Mental Health*, *16*(1), 5–18. https://doi.org/10.1002/pmh.1522

Lobbestael, J., Leurgans, M., & Arntz, A. (2011). Inter-rater reliability of the Structured Clinical Interview for DSM-IV axis I disorders (SCID I) and axis II disorders (SCID II). *Clinical Psychology & Psychotherapy*, *18*(1), 75–79. https://doi.org/10.1002/cpp.693

McCabe, G. A., & Widiger, T. A. (2020). A comprehensive comparison of the ICD-11 and DSM–5 section III personality disorder models. *Psychological Assessment*, *32*(1), 72–84. https://doi.org/10.1037/pas0000772

Morey, L. C. (2017). Development and initial evaluation of a self-report form of the DSM-5 Level of Personality Functioning Scale. *Psychological Assessment*, *29*(10), 1302–1308. https://doi.org/10.1037/pas0000450

Morey, L. C., Bender, D. S., & Skodol, A. E. (2013). Validating the proposed diagnostic and statistical manual of mental disorders, severity indicator for personality disorder. *The Journal of Nervous and Mental Disease*, *201*(9), 729–735. https://doi.org/10.1097/NMD.0b013e3182a20ea8

Morey, L. C., Benson, K. T., Busch, A. J., & Skodol, A. E. (2015). Personality disorders in DSM-5: Emerging research on the alternative model. *Current Psychiatry Reports*, *17*(4), 24. https://doi.org/10.1007/s11920-015-0558-0

Ohse, L., Zimmermann, J., Kerber, A., Kampe, L., Mohr, J., Kendlbacher, J., Busch, O., Rentrop, M., & Hörz-Sagstetter, S. (2023). Reliability, structure, and validity of module I (personality functioning) of the structured clinical interview for the alternative DSM–5 model for personality disorders (SCID-5-AMPD-I). *Personality Disorders: Theory, Research, and Treatment*, *14*(3), 287–299. https://doi.org/10.1037/per0000576

Oltmanns, J. R., & Widiger, T. A. (2016). Self-pathology, the five-factor model, and bloated specific factors: A cautionary tale. *Journal of Abnormal Psychology*, *125*(3), 423–434. https://doi.org/10.1037/abn0000144

OPD Task Force (2008). *Operationalized psychodynamic diagnosis OPD-2: Manual of diagnosis and treatment planning*. Hogrefe & Huber.

Osório, F. L., Loureiro, S. R., Hallak, J. E. C., Machado-de-Sousa, J. P., Ushirohira, J. M., Baes, C. V. W., Apolinario, T. D., Donadon, M. F., Bolsoni, L. M., Guimarães, T., Fracon, V. S., Silva-Rodrigues, A. P. C., Pizeta, F. A., Souza, R. M., Sanches, R. F., Dos Santos, R. G., Martin-Santos, R., & Crippa, J. A. S. (2019). Clinical validity and intrarater and test–retest reliability of the structured clinical interview for DSM-5–clinician version (SCID-5-CV). *Psychiatry and Clinical Neurosciences*, *73*(12), 754–760. https://doi.org/10.1111/pcn.12931

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

R Core Team (2021). *R: A language and environment for statistical computing [computer software]*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, *73*(1), 5–26. https://doi.org/10.1177/0013164412449831

Revelle, W. (2020). *Package 'psych': procedures for psychological, psychometric, and personality research.* R package version 2.1.3.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/a0029315

Roche, M. J., & Jaweed, S. (2023). Comparing measures of Criterion A to better understand incremental validity in the alternative model of personality disorders. *Assessment*, *30*(3), 689–705. https://doi.org/10.1177/10731911211059763

Roche, M. J., Jacobson, N. C., & Phillips, J. J. (2018). Expanding the validity of the Level of Personality Functioning Scale observer report and self-report versions across psychodynamic and interpersonal paradigms. *Journal of Personality Assessment*, *100*(6), 571–580. https://doi.org/10.1080/00223891.2018.1475394

Roche, M. J., Jacobson, N. C., & Pincus, A. L. (2016). Using repeated daily assessments to uncover oscillating patterns and temporally-dynamic triggers in structures of psychopathology: Applications to the DSM-5 alternative model of personality disorders. *Journal of Abnormal Psychology*, *125*(8), 1090–1102. https://doi.org/10.1037/abn0000177

Rosseel, Y. (2019). *Package 'Lavaan'*. R package version 0.6–3. https://cran.r-project.org/web/packages/lavaan/index.html.

Samylkin, D. V., Demidova, L., Gadisov, T. G., Nikitsky, I. S., & Tkachenko, A. A. (2023). Предварительная адаптация и валидизация психометрических методик для дименсиональной оценки расстройств личности в рамках перехода к МКБ-11. *Российский психиатрический журнал*, (1), 51–59.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507–514. https://doi.org/10.1007/BF02296192

Sharp, C., & Wall, K. (2021). DSM-5 level of personality functioning: Refocusing personality disorder on what it means to be human. *Annual Review of Clinical Psychology*, *17*(1), 313–337. https://doi.org/10.1146/annurev-clinpsy-081219-105402

Sharp, C., Wright, A. G. C., Fowler, J. C., Frueh, B. C., Allen, J. G., Oldham, J., & Clark, L. A. (2015). The structure of personality pathology: Both general ('g') and specific ('s') factors? *Journal of Abnormal Psychology*, *124*(2), 387–398. https://doi.org/10.1037/abn0000033

Siefert, C. J., Sexton, J., Meehan, K., Nelson, S., Haggerty, G., Dauphin, B., & Huprich, S. (2020). Development of a short form for the DSM–5 levels of personality functioning questionnaire. *Journal of Personality Assessment*, *102*(4), 516–526. https://doi.org/10.1080/00223891.2019.1594842

Somma, A., Borroni, S., Gialdi, G., Carlotta, D., Emanuela Giarolli, L., Barranca, M., Cerioli, C., Franzoni, C., Masci, E., Manini, R., Luca Busso, S., Ruotolo, G., Krueger, R. F., Markon, K. E., & Fossati, A. (2020). The inter-rater reliability and validity of the Italian translation of the structured clinical interview for DSM-5 alternative model for personality disorders module i and module ii: A preliminary report on consecutively admitted psychotherapy outpatients. *Journal of Personality Disorders*, *34*(Suppl C), 95–123. https://doi.org/10.1521/pedi_2020_34_511

Spitzer, C., Müller, S., Kerber, A., Hutsebaut, J., Brähler, E., & Zimmermann, J. (2021). Die deutsche Version der Level of Personality Functioning Scale-Brief Form 2.0 (LPFS-BF): Faktorenstruktur, konvergente Validität und Normwerte in der Allgemeinbevölkerung. *PPmP: Psychotherapie, Psychosomatik, Medizinische Psychologie*, *71*(07), 284–293. https://doi.org/10.1055/a-1343-2396

Su, R., Tay, L., & Diener, E. (2014). The development and validation of Comprehensive Inventory of Thriving (CIT) and Brief Inventory of Thriving (BIT). *Applied Psychology: Health and Well-Being*, *6*(3), 251–279. https://doi.org/10.1111/aphw.12027

Tyrer, P., Reed, G. M., & Crawford, M. J. (2015). Classification, assessment, prevalence, and effect of personality disorder. *Lancet*, *385*(9969), 717–726. https://doi.org/10.1016/S0140-6736(14)61995-4

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

Waugh, M. H., McClain, C. M., Mariotti, E. C., Mulay, A. L., DeVore, E. N., Lenger, K. A., Russell, A. N., Florimbio, A. R., Lewis, K. C., Ridenour, J. M., & Beevers, L. G. (2021). Comparative content analysis of self-report scales for level of personality functioning. *Journal of Personality Assessment*, *103*(2), 161–173. https://doi.org/10.1080/00223891.2019.1705464

Weekers, L. C., Hutsebaut, J., & Kamphuis, J. H. (2019). The Level of Personality Functioning Scale-Brief Form 2.0: Update of a brief instrument for assessing level of personality functioning. *Personality and Mental Health*, *13*(1), 3–14. https://doi.org/10.1002/pmh.1434

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Sage Publications, Inc.

Wickham, H., & Wickham, M. H. (2017). Package tidyverse. Tidyverse.

Widiger, T. A., & Hines, A. (2022). The diagnostic and statistical manual of mental disorders, fifth edition alternative model of personality disorder. *Personality Disorders*, *13*(4), 347–355. https://doi.org/10.1037/per0000524

Widiger, T. A., & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *The American Psychologist*, *62*(2), 71–83. https://doi.org/10.1037/0003-066x.62.2.71

Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, *28*(3), 186–191. https://doi.org/10.1007/s10862-005-9004-7

World Health Organization (1992). *International classification of diseases ICD-10* (10th ed.). Author.

World Health Organization (2018). *International classification of diseases for mortality and morbidity statistics* (11th rev.). Retrieved from https://icd.who.int/browse11/l-m/en

Wright, A. G. C., Ringwald, W. R., Hopwood, C. J., & Pincus, A. L. (2022). It's time to replace the personality disorders with the interpersonal disorders. *The American Psychologist*, *77*(9), 1085–1099. https://doi.org/10.1037/amp0001087

Zettl, M., Taubner, S., Hutsebaut, J., & Volkert, J. (2019). Psychometrische evaluation der deutschen Version des Semistrukturierten Interviews zur Erfassung der DSM-5 Persönlichkeitsfunktionen (STiP-5.1). *PPmP: Psychotherapie, Psychosomatik, Medizinische Psychologie*, *69*(12), 499–504. https://doi.org/10.1055/a-1010-6887

Zimmermann, J., Altenstein, D., Krieger, T., Holtforth, M. G., Pretsch, J., Alexopoulos, J., Spitzer, C., Benecke, C., Krueger, R. F., Markon, K. E., & Leising, D. (2014). The structure and correlates of self-reported DSM-5 maladaptive personality traits: Findings from two German-speaking samples. *Journal of Personality Disorders*, *28*(4), 518–540. https://doi.org/10.1521/pedi_2014_28_130

Zimmermann, J., Benecke, C., Hörz, S., Rentrop, M., Peham, D., Bock, A., Wallner, T., Schauenburg, H., Frommer, J., Huber, D., Clarkin, J. F., & Dammann, G. (2013). Validierung einer deutschsprachigen 16-item-Version des Inventars der Persönlichkeitsorganisation (IPO-16). *Diagnostica*, *59*(1), 3–16. https://doi.org/10.1026/0012-1924/a000076

Zimmermann, J., Benecke, C., Hörz-Sagstetter, S., & Dammann, G. (2015). Normierung der deutschsprachigen 16-Item-Version des Inventars der Persönlichkeitsorganisation (IPO-16). *Zeitschrift fur Psychosomatische Medizin und Psychotherapie*, *61*(1), 5–18. https://doi.org/10.13109/zptm.2015.61.1.5

Zimmermann, J., Ehrenthal, J. C., Cierpka, M., Schauenburg, H., Doering, S., & Benecke, C. (2012). Assessing the level of structural integration using operationalized psychodynamic diagnosis (OPD): Implications for DSM–5. *Journal of Personality Assessment*, *94*(5), 522–532. https://doi.org/10.1080/00223891.1012.700664

Zimmermann, J., Falk, C. F., Wendt, L., Spitzer, C., Fischer, F., Bach, B., Sellbom, M., & Müller, S. (2023). Validating the German version of the Personality Disorder Severity-ICD-11 Scale using nominal response models. *Psychological Assessment*, *35*(3), 257–268. https://doi.org/10.1037/pas0001199

Zimmermann, J., Hopwood, C. J., & Krueger, R. F. (2023). The DSM-5 level of personality functioning scale. In R. F. Krueger & P. H. Blaney (Eds.), *Oxford textbook of psychopathology* (4th ed., pp. 579–603). Oxford University Press. https://doi.org/10.1093/med-psych/9780197542521.003.0025

Zimmermann, J., Müller, S., Bach, B., Hutsebaut, J., Hummelen, B., & Fischer, F. (2020). A Common metric for self-reported severity of personality disorder. *Psychopathology*, *53*(3–4), 168–178. https://doi.org/10.1159/000507377

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*(4), 399–413. https://doi.org/10.1037/1082-989X.12.4.399