

University of Groningen

## Evaluating the Construct Validity of the Norwegian Version of the Level of Personality Functioning Scale - Brief Form 2.0 in a Large Clinical Sample

Paap, Muirne C. S.; Pedersen, Geir; Kvarstein, Elfrida; Hummelen, Benjamin

*Published in:*  
Journal of Personality Assessment

*DOI:*  
[10.1080/00223891.2023.2182694](https://doi.org/10.1080/00223891.2023.2182694)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2024

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Paap, M. C. S., Pedersen, G., Kvarstein, E., & Hummelen, B. (2024). Evaluating the Construct Validity of the Norwegian Version of the Level of Personality Functioning Scale - Brief Form 2.0 in a Large Clinical Sample. *Journal of Personality Assessment*, *106*(1), 49-59.  
<https://doi.org/10.1080/00223891.2023.2182694>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

## Evaluating the Construct Validity of the Norwegian Version of the Level of Personality Functioning Scale – Brief Form 2.0 in a Large Clinical Sample

Muirne C. S. Paap, Geir Pedersen, Elfrida Kvarstein & Benjamin Hummelen

**To cite this article:** Muirne C. S. Paap, Geir Pedersen, Elfrida Kvarstein & Benjamin Hummelen (2024) Evaluating the Construct Validity of the Norwegian Version of the Level of Personality Functioning Scale – Brief Form 2.0 in a Large Clinical Sample, *Journal of Personality Assessment*, 106:1, 49-59, DOI: [10.1080/00223891.2023.2182694](https://doi.org/10.1080/00223891.2023.2182694)

**To link to this article:** <https://doi.org/10.1080/00223891.2023.2182694>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC



Published online: 10 Mar 2023.



Submit your article to this journal [↗](#)



Article views: 1003



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

# Evaluating the Construct Validity of the Norwegian Version of the Level of Personality Functioning Scale – Brief Form 2.0 in a Large Clinical Sample

Muirne C. S. Paap<sup>1,2</sup> , Geir Pedersen<sup>3,4</sup> , Elfrida Kvarstein<sup>3,5</sup> , and Benjamin Hummelen<sup>1</sup> 

<sup>1</sup>Department of Research and Innovation, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; <sup>2</sup>Department of Child and Family Welfare, Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands; <sup>3</sup>Section for Personality Psychiatry and Specialized Treatments, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway; <sup>4</sup>The Norwegian Centre of Mental Disorders Research (NORMENT), Institute of Clinical Medicine, University of Oslo, Oslo, Norway; <sup>5</sup>Institute of Clinical Medicine, University of Oslo, Oslo, Norway

## ABSTRACT

The Level of Personality Functioning – Brief Form 2.0 (LPFS-BF 2.0) is a 12-item self-report questionnaire developed to gain a quick impression of the severity of personality pathology according to the DSM-5 Alternative Model for Personality Disorders (AMPD). The current study evaluated the construct validity and reliability of the Norwegian version of the LPFS-BF 2.0 in a large clinical sample ( $N = 1673$ ). Dimensionality was examined using confirmatory factor analysis and bifactor analysis followed by an analysis of distinctiveness of the subscales using the proportional reduction in mean squared error (PRMSE), and the concurrent validity was examined using correlations with self-report questionnaires and clinical interviews assessing PDs according to section II of the DSM-5. Taking the findings of the dimensionality and concurrent validity results together, we found moderate to good support for the use of total scores for the Norwegian version of the LPFS-BF 2.0. We would advise against the use of subscale scores, since the subscales provided only a small amount of reliable unique variance.

## ARTICLE HISTORY

Received 1 November 2021  
Accepted 9 February 2023

## Introduction



In the Alternative Model of Personality Disorders (AMPD) of the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013), the presence and severity of a personality disorder (PD) is assessed using the Level of Personality Functioning Scale (LPFS). This scale focuses on the two core features of personality pathology; i.e., impairment of self-functioning and impairment of interpersonal functioning (Bender et al., 2011). The LPFS is assessed at five severity levels, delineating four areas of personality functioning (identity, self-direction, empathy, and intimacy).

The LPFS taps into a wide array of characteristics pertaining to human personality, such as the capability to maintain positive self-esteem, to live according to one's own values, and to mentalize and cooperate in close relationships. In spite of its complexity, it was a specific intent of the DSM-5 Personality and Personality Disorder Workgroup to create a unidimensional construct when developing the LPFS (Morey, 2019). This position is in accordance with the theoretical position that a good sense of self is highly interdependent with the quality of interpersonal relationships (e.g., Fonagy et al., 2018). Thus, ideally, instruments for the

assessment of the LPFS should represent both the complexity and the unidimensional character of the LPFS.

The inclusion of the AMPD in the DSM-5 has given an impetus to the development of self-report questionnaires assessing personality functioning in accordance with the LPFS. For adults, several self-report questionnaires for the LPFS are now available: the 132-item DSM-5 Levels of Personality Functioning Questionnaire (132 items; Huprich et al., 2018) and its 24-item short form (24 items; Siefert et al., 2020); the 80-item LPFS-Self Report (80 items; Morey, 2017; Sleep et al., 2019); the 24-item Self and Interpersonal Functioning Scale (24 items; Gamache et al., 2019); the 12-item AMPD-Criterion A Scale (12 items; Dowgwillo et al., 2018); and the 12-item Level of Personality Functioning Scale-Brief Form (LPFS-BF, 12 items; Hutsebaut et al., 2016) and its successor, the LPFS-BF 2.0 (Bach & Hutsebaut, 2018; Weekers et al., 2019).

Self-report questionnaires are generally used to obtain a quick, general impression of the severity of personality pathology. They have often been designed for screening purposes, to supplement structured clinical interviews, and to evaluate clinical change in treatment studies (Stanton et al., 2019; Weekers et al., 2019). The brevity of the LPFS-BF makes it an attractive instrument to use in these situations; and, now, the second version of the LPFS-BF has been

**CONTACT** Muirne C. S. Paap  [m.c.s.paap@rug.nl](mailto:m.c.s.paap@rug.nl)  Department of Research and Innovation, Clinic for Mental Health and Addiction, Oslo University Hospital, P.O. Box 4956, Nydalen, Oslo, 0424, Norway.

© 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-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. 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.

included in the Standard Set of Patient-recorded Outcomes for Personality Disorder (Prevolnik Rupel et al., 2021), alongside other instruments assessing aspects of personality pathology, such as self-harm, sense of belonging, social functioning, emotional distress, and aggression. Thus, the LPFS-BF may become one of the most widely used instruments for the assessment of the LPFS, underlining the importance of examining the psychometric properties of this instrument.

As mentioned previously, when the LPFS was developed it was intended to be a unidimensional – yet conceptually broad – construct. Most instruments measuring constructs that are conceptually broad will show a certain degree of multidimensionality (see Calderón Garrido et al., 2019). Therefore, we expect to find evidence both for a strong general factor and some multidimensionality. If the general factor is sufficiently strong, this can be taken as evidence of “essential unidimensionality” (see Reise et al., 2013). A study using a large ( $N = 924$ ) non-clinical sample did indeed find support for a strong general factor underlying the LPFS-BF 2.0 (Zimmermann et al., 2020). Two studies (Bach & Hutsebaut, 2018; Weekers et al., 2019) analyzing data from clinical samples, however, favored a two-factor solution for the LPFS-BF 2.0. This could be taken to indicate the presence of some non-ignorable multidimensionality in the LPFS-BF 2.0, at least when used with patients. However, the authors of both aforementioned studies underlined in their conclusions that the factors were highly correlated (close to .70), which could lend support to the notion that conceptualization of general personality functions as a unidimensional construct.

Weekers et al. (2019) further evaluated the construct validity of the LPFS by investigating the association between the total score of the LPFS-BF 2.0 as well as the two subscales they found, and other dimensional measures of personality pathology. The authors reported meaningful associations between the total score of the LPFS-BF 2.0 and a number of other clinical characteristics (such as self-reported depressive symptoms, borderline PD, and domain scores of the Severity Indices of Personality Problems (SIPP), a self-report questionnaire measuring the core components of (mal)adaptive personality functioning (Verheul et al., 2008)). However, it appeared that differentiating between the two subscales of the LPFS, i.e., impairment of self-functioning and impairment of interpersonal functioning, had some clinical value as well, since self-functioning was significantly more strongly correlated with the identity integration domain of the SIPP ( $r = -.62$  versus  $r = -.25$  for interpersonal functioning), whereas interpersonal functioning was significantly more strongly correlated with the social concordance domain ( $r = -.66$  versus  $r = -.24$  for self-functioning). In the study of Weekers et al. (2019), only the five main domains of the SIPP as postulated by Verheul et al. (2008) were examined. However, the 16 facets of the SIPP seem to have better content and construct validity (Paap et al., 2021; Pedersen et al., 2019). In this respect, it is of clinical and scientific interest to conduct a more detailed examination of the relationship between the LPFS-BF 2.0 and the SIPP facets.

### **Aims of the study**

In the current study, we aim to shed further light on the psychometric properties of the LPFS-BF 2.0 in a large

clinical sample of patients with personality disorders and personality difficulties below the personality disorder threshold. More specifically, we will investigate whether the LPFS-BF 2.0 can be treated as essentially unidimensional. In order to investigate whether the chosen analytic approach has an impact on the results and conclusions based thereon, and to facilitate a comparison of our findings to those found by other authors, we will estimate a bifactor model and various two-factor models using confirmatory factor analysis. We will further examine the concurrent validity of the LPFS-BF 2.0 by analyzing the association between the LPFS-BF 2.0 and other clinical measures of psychopathology including self-report questionnaires for assessing personality functioning, interpersonal problems, symptom distress, and psychosocial functioning; and structured clinical interviews for assessing the PD criteria as described by section II of the DSM-5. Finally, we will examine the reliability of the total and subscale scores.

## **Material and methods**

### **Participants**

In this multi-site, naturalistic and explorative study, the material comprised data from 15 outpatient units within the Norwegian Network for Personality Disorders (Karterud et al., 2003; Pedersen et al., 2022). All units are specialized in treatment of personality disorders and personality-related difficulties. Initially, the participants comprised 1812 patients referred to outpatient units within this network, in the period August 2017 to April 2019. The data represent the initial assessment before admittance to treatment. Of these 1812 patients, 139 had not completed the LPFS-BF 2.0. Thus, the final study sample comprised 1673 patients, of whom 79% were female. Mean age was 30 years ( $SD = 9$ ), and mean level of education after compulsory junior secondary school (ending at age 16) was 4 years ( $SD = 3$ ). Sixty-nine percent had one or more PD diagnoses, and 93% had one or more symptom disorders. See Table 1 for demographics and a more detailed description of diagnostic prevalences.

All participating patients of each treatment unit provided written consents, allowing the anonymous use of clinical data for research purposes. Data collection procedures at each contributing unit were approved by the local Data Protection Officer. Security procedures for the quality register were approved by the Data Protection Officer at Oslo University Hospital. Since data in the quality register are anonymous, formal approval from the Regional Committee for Medical Research and Ethics was not required.

### **Clinical assessment of diagnoses and psychosocial functioning**

Patients were diagnosed according to the DSM-5 (APA, 2013), using the Mini International Neuropsychiatric Interview (MINI, Sheehan et al., 1994) for symptom disorders, and the Structured Clinical Interview for DSM-5

**Table 1.** Demographics and diagnostic prevalence at baseline.

|  | Prevalence |
|--|------------|
| Demographics                           |            |
| Living alone                           | 31%        |
| Self-mutilation last 6 months          | 37%        |
| History of suicide attempts            | 41%        |
| Suicide attempts last 6 months         | 12%        |
| Symptom diagnoses*                     |            |
| One or more symptom disorders          | 94%        |
| Current major depressive disorder      | 48%        |
| Major depressive disorder in remission | 10%        |
| Dysthymic disorder                     | 12%        |
| Bipolar I                              | 2%         |
| Bipolar II                             | 4%         |
| Panic without agoraphobia              | 9%         |
| Panic with agoraphobia                 | 9%         |
| Agoraphobia without panic              | 5%         |
| Social phobia                          | 25%        |
| Obsessive-compulsive disorder          | 5%         |
| Generalized anxiety disorder           | 17%        |
| Post-traumatic stress disorder         | 13%        |
| Alcohol dependence/abuse               | 7%         |
| Substance dependence/abuse             | 4%         |
| Eating disorders                       | 8%         |
| Personality disorders                  |            |
| One or more personality disorders      | 69%        |
| Paranoid PD                            | 9%         |
| Borderline PD                          | 33%        |
| Avoidant PD                            | 36%        |
| Obsessive-compulsive PD                | 7%         |
| Dependent PD                           | 5%         |
| PD NOS**                               | 3%         |

Note: \*Diagnoses occurring in less than 2% not included.

\*\*PD NOS: Personality disorder not otherwise specified.

Personality Disorders (SCID-5-PD) for section II PDs (SCID-5-PD, First et al., 2016). Diagnostic reliability was not investigated. However, diagnostic assessments were performed in each unit by clinical staff who had received systematic training in diagnostic interviews and principles of the LEAD-procedure (Longitudinal, Expert, All-Data; Pedersen et al., 2013; Spitzer, 1983). This means that diagnoses were based on all available information, including referral letters, self-reported history and complaints, overall clinical impression, alongside diagnostic interviews.

Psychosocial functioning was evaluated using a revised version of the Global Assessment of Functioning Scale (GAF, American Psychiatric Association, 1994), the Global Functioning Scale (GFS; Pedersen et al., 2018). The GFS is an observer-based single score ranging from 1 to 100, representing symptom severity and social and occupational impairment. In the current study, the GFS was scored according to the split-version (Pedersen & Karterud, 2012), in which the two scores, i.e., symptom severity and social/occupational impairment, are evaluated and scored separately. The GFS score used in this study was based on the lower of the two scores. In a study by Pedersen et al. (2011), reliability of the original split-version of GAF was found acceptable, with a generalizability coefficient of .84 for relative decisions and approximately .82 for absolute decisions. Conventional interpretations of severity indicated by GFS scores are the same as in the original GAF: Mild (61–70); Moderate (51–60); and Severe (41–50), whereas levels below 41 reflect increased levels of severity (APA, 2000, pp. 32–34).

### Assessment by patient self-report

The 12-item self-report questionnaire LPFS-BF 2.0 was scored using a 4-point ordered response format with a point range of 0 to 3<sup>1</sup>: 0 = “Very False or Often False,” 1 = “Sometimes or Somewhat False,” 2 = “Sometimes or Somewhat True,” and 3 = “Very True or Often True.” The twelve items of the LPFS-BF 2.0 correspond to the twelve indicators of the LPFS (see Table 2). Items 1–6 cover self-functioning and items 7–12 cover interpersonal functioning. The LPFS-BF 2.0 was translated into Norwegian in accordance with the guidelines of Hambleton (2005) by a group of four clinical researchers. The last author, who is fluent in both Dutch and Norwegian, translated the original version of the LPFS-BF 2.0 from Dutch into Norwegian, and the third author (a bilingual English-Norwegian speaker) and second author translated the English version into Norwegian. The two translations were compared and inconsistencies were resolved. Another bilingual researcher back-translated the interim version into Dutch. The first author compared this version with the original Dutch questionnaire and made some minor adaptations, resulting in the final version used in this study. For the original Dutch version of the LPFS-BF 2.0, Cronbach’s alpha (Cronbach, 1951) values of .82, .79 and .71 were found in a sample of treatment-seeking adults for the total score, the self-functioning subscale and the interpersonal functioning subscale, respectively (Weekers et al., 2019).

The Severity Indices of Personality Problems (SIPP) is a 118-item self-report instrument scored on a 4-point Likert scale that ranges from “fully disagree” (1) to “fully agree” (4). The items are organized into 16 facets, or subscales (see Appendix A), that aim to measure aspects of the individual’s levels of (mal)adaptive capacities indicating areas of personality dysfunction (Verheul et al., 2008). The SIPP has been analyzed in community samples (Andrea et al., 2007; Pedersen et al., 2019), clinical PD samples (Andrea et al., 2007; Arnevik et al., 2019; Arnevik et al., 2009), and clinical and non-clinical adolescent samples (Feenstra et al., 2011). The psychometric properties of the Norwegian version of the SIPP have been found to be moderate to good: on the whole, the facets are associated with high levels of distinctiveness and adequate levels of reliability, and these findings have been shown to hold across multiple populations (Paap et al., 2021; Pedersen et al., 2019). For example, Cronbach’s alpha values between 0.67 and 0.84 have been reported for the facets in a Norwegian community sample, and values between 0.70 and 0.85 in a clinical PD sample (Pedersen et al., 2019).

Self-reported interpersonal problems were assessed by means of the Circumplex of Interpersonal Problems (CIP; Pedersen, 2002), a 48-item Norwegian version of the Inventory of Interpersonal Problems – Circumplex version (Alden et al., 1990). The CIP has a 5-point response format ranging from 0 to 4, grading subjective distress from “not at all” to “extremely.” Only the mean score was used in this study. As a part of an earlier study of the CIP (Pedersen et al., 2011), the test-retest stability (3–4 days interval) of the

<sup>1</sup>Note that the original Dutch version is scored on a scale of 1–4.



mean score was estimated at 0.96 (ICC 2.1, 95% CI: 0.93–0.98,  $n = 53$ , unpublished data).

Self-reported level of depression was measured using the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001), a 9-item questionnaire with a 4-point response format (0–3) from “Not at all” to “Nearly every day.” PHQ-9 scores are computed as the sum score of all nine items, ranging from 0–27. Scale reliability of the PHQ-9 items, indicated by Cronbach’s alpha, was 0.82 in the current sample.

Self-reported level of anxiety was measured using the Patient Health Questionnaire (GAD-7; Spitzer et al., 2006), a 7-item questionnaire with a 4-point response format (0–3) from “Not at all” to “Nearly every day.” GAD-7 scores were computed as the sum score of all seven items, ranging from 0 to 21. Scale reliability of the GAD-7 items, indicated by Cronbach’s alpha, was 0.82 in the current sample.

Psychosocial functioning was assessed using the Work and Social Adjustment Scale (WSAS; Mundt et al., 2002), a self-report 5-item scale that measures the level of impairment on a scale from 0 to 8, with 0 indicating no impairment at all and 8 indicating very severe impairment. The scores on the five different items are summarized by a total score of 0–40. The WSAS measures the individual variation in clinically important aspects of impairment, and has shown to be highly reliable in a patient sample similar to the current sample (Pedersen et al., 2017). Scale reliability of the WSAS items, estimated by using Cronbach’s alpha, was 0.77 in the current sample.

### Missing data

The data of the current study are based on ordinary routine assessments, and the occurrence of missing data is therefore unavoidable. At the organizational level, the most common failures were slips during the routine administrative tasks (patients were not administered the tests or did not hand in the tests), delay in providing certain instruments during the implementation phase of this project, and non-simultaneous registration of tests; e.g., “diagnosis deferred.” Diagnostic status for PD, symptom disorders, and GFS scores were not registered for 439, 784, and 263 patients, respectively. Of the 439 cases of patients with missing PD diagnosis, 195 (12% of total sample) were due to deferred diagnosis, and 244 (15%) were due to unknown reasons. Overall, these missing data are due to organizational aspects, and may therefore be considered as completely missing at random. This was supported by a series of independent sample *t*-tests comparing patients with and without a registered PD diagnosis measured on a number of clinical variables, i.e., the GAD-7, PHQ-9, CIP, WSAS, GFS, and LPFS-BF 2.0 total and subscale scores. No significant differences were found.

Of the 784 cases of patients with missing symptom disorder diagnoses, 540 (32% of total sample) were due to deferred diagnosis, and 244 (15%) were due to unknown reasons. Independent *t*-tests showed that there were no significant differences between patients with missing symptom diagnoses and patients without missing symptom diagnoses

with respect to level of GAD-7, PHQ-9, CIP, WSAS, GFS, and LPFS-BF 2.0.

Overall, 97 patients had one or more missing values on the LPFS-BF 2.0. More specifically, 78 patients (4.7%) had one missing value; 11 patients (0.7%) had two missing values; and four patients had four missing values. Due to their low frequency and the lack of a systematic pattern (no specific item had more missing responses than others), they were considered random and of no threat to the validity of the inferences based on the current study. This position was supported by a series of independent sample *t*-tests comparing patient groups with and without LPFS-BF 2.0 values using the GAD-7, PHQ-9, CIP, WSAS, GFS measures and a number of SCID-5-PD criteria. Only GFS scores revealed statistically significant differences between the two groups ( $p < .05$ ). However, the difference was only 1.5 GFS points in magnitude, which we consider negligible.

### Psychometric analyses

We used confirmatory bifactor analysis (Cai, 2010; Reise, 2012) to investigate whether we could find support for essential unidimensionality of the LPFS-BF 2.0. In a bifactor model, all items load directly on their respective subscale as well as on the general factor. This sets it apart from a correlated-trait model, where items only load on their own respective factors, but these factors are allowed to correlate. Using the results from the bifactor analysis, one can calculate the explained common variance (ECV). For a confirmatory bifactor analysis where all factors are orthogonal, the ECV for a factor equals the sum of all squared loadings for that factor divided by the sum of all squared loadings across all factors. This index can be seen as an indicator of relative factor strength (Reise, 2012). Reise, Scheines, Widaman, and Haviland (2013) tentatively proposed that when the ECV for the general factor in a bifactor model is larger than 60%, the factor loading estimates for a unidimensional model are close to the true loadings on the general factor in the bifactor model, and the measure can be interpreted as essentially unidimensional. More recently, O’Connor Quinn (2014) proposed a cutoff of 70%. As the authors of both studies stated, these numbers should not be used as strict cutoff values, but instead as general guidelines. In addition to the ECV, we evaluated the percentage of uncontaminated correlations (PUC), Omega, and Omega-hierarchical (Omega-H). It has been suggested that when both ECV and PUC values are sufficiently high, the bias associated with ignoring any multidimensionality present tends to be low (Rodriguez et al., 2016). Omega can be seen as the latent variable analogue to coefficient alpha; it takes the factor structure into account. Omega-H only reflects the variance attributable to a single latent variable (i.e., for a specific factor, it reflects the proportion of reliable systematic variance of that subscale after partitioning out variability that can be attributed to the general factor). As described by Rodriguez et al., when Omega-H for a specific factor is low compared to its Omega estimate, a large part of the reliable variance of the subscale scores can

be attributed to the general factor, rather than what is unique to the specific factor. For a more detailed description of the various Omega coefficients that have been proposed for different types of factor models, see Flora (2020).

As an additional check, we calculated the proportional reduction in mean squared error (PRMSE) to evaluate the incremental value of using subscales (over and above a total score) (Haberman, 2008). A value-added ratio  $> 1.1$  is taken as indicative of a minimally meaningful added value of the subscale score (Feinberg & Jurich, 2017). Note, that these calculations are based on principles of classical test theory. See Paap et al. (2021) for a more detailed description of this method.

In order to ensure comparability with results reported by previous studies, we specified two confirmatory factor models: (1) a one-factor model in which the items loaded on a single factor, and (2) a correlated two-factor model with two correlated factors in which the self items loaded on the first factor, and interpersonal items on the second factor. Initially, we did not include any cross-loadings or correlations between residuals. To explore whether it would improve model fit, we allowed cross-loadings in the second step. The third step also included correlations between residuals.

Goodness of fit was evaluated using Root Mean Square Error of Approximation (RMSEA; Steiger, 1990), the Tucker Lewis Index (TLI; Tucker & Lewis, 1973), also called the Non-Normed Fit Index (NNFI; Bentler & Bonett, 1980), and the Comparative Fit Index (CFI; Bentler, 1990). Scale reliability was estimated using McDonald's Omega ( $\omega$ ) (McDonald, 2011; Trizano-Hermosilla & Alvarado, 2016).

In interpreting the RMSEA values, we applied the following rules of thumb: values of 0.05 or below indicate good model fit; values between 0.05 and 0.08 indicate reasonable fit; values between 0.08 and 0.10 indicate mediocre fit; and values greater than 0.10 indicate unacceptable fit (MacCallum et al., 1996). The general consensus is to use a cutoff value close to 0.06 (Hu & Bentler, 1999) or a stringent upper limit of 0.07 (Steiger, 2007). The CFI and TLI are derived from the chi square statistic, which measures the fit of the model compared to the independence model; the values are supposed to be between 0 and 1. Values greater than 0.90 are normally required for a good fit of a model,

although Hu and Bentler (1999) have suggested  $TLI \geq 0.95$  as the threshold.

## Software

The bifactor analyses and PRMSE-based analyses were performed in the open source software program R version 3.6.1 (R Core Team, 2019). The bifactor model was estimated using a full information maximum likelihood approach in the R package mirt version 1.31 (Chalmers, 2012), which follows the analytic strategy outlined by Cai (2010). Custom coding was used for the PRMSE-based analyses. Confirmatory factor analyses were conducted using Mplus 7.4 (Muthén & Muthén, 1998–2015). Due to the ordinal measurement level of LPFS-BF 2.0 items, estimations were based on the Weighted Least Square Mean and Variance Adjusted (WLSMV) function (Li, 2016). Descriptive and correlational statistics were computed using IBM SPSS Statistics for Windows, Release 26 (IBM, 2019).

## Results

### Bifactor model and subscale distinctiveness

The results of the confirmatory bifactor analysis showed that the general factor explained the vast majority of the variance, i.e., 66.5%. The self factor explained 23.0% of the variance and the interpersonal factor 10.5%. The PUC value equaled 54.5%. The Omega estimates for the general, self and interpersonal factors equaled .87, .77, and .84, respectively. When comparing these values to the Omega-hierarchical (Omega-H) values, there was only a small drop for the general factor (Omega-H = .74), a large drop for the self specific factor (Omega-H = .42) and a very sizeable drop for the interpersonal specific factor (Omega-H = .02). This indicates that a substantial amount of the reliable variance of the subscales is attributable to the general factor. The factor loadings are presented in Table 2. It can be seen that the interpersonal items (items 7–12) loaded more strongly on the general factor than the self items. Item 11 loaded negatively on the interpersonal subsfactor. As is usual in bifactor models, fit indices indicated good model fit (RMSEA=.051 (95% C.I.: .041–.062); CFI=.961; TLI=.909).

**Table 2.** Items of the LPFS-BF 2.0, descriptive statistics, factor loadings of the bifactor model, and factor loadings of the final CFA model.

|  | Mean | SD   | Bifactor model |     |      | CFA |      |
|--|------|------|----------------|-----|------|-----|------|
|  |      |      | G              | S1  | S2   | F1  | F2   |
| 1. I often do not know who I really am   | 1.64 | 1.04 | .48            | .49 |      | .66 |      |
| 2. I often think negatively about myself   | 2.44 | 0.75 | .32            | .52 |      | .82 | -.31 |
| 3. My emotions change without me having grip on them                               | 2.05 | 0.92 | .49            | .38 |      | .64 |      |
| 4. I have no sense of where I want to go in my life                                | 1.71 | 1.05 | .33            | .40 |      | .49 |      |
| 5. I often do not understand my own thoughts and feelings                          | 1.73 | 0.95 | .48            | .55 |      | .70 |      |
| 6. I often make unrealistic demands on myself                                      | 1.82 | 1.02 | .33            | .31 |      | .46 |      |
| 7. I often have difficulty understanding the thoughts and feelings of others       | 1.09 | 0.98 | .62            |     | .29  |     | .67  |
| 8. I often find it hard to stand it when others have a different opinion           | 1.16 | 0.98 | .60            |     | .29  |     | .68  |
| 9. I often do not fully understand why my behavior has a certain effect on others  | 0.95 | 0.96 | .73            |     | .35  |     | .76  |
| 10. My relationships and friendships never last long                               | 1.12 | 1.02 | .61            |     | -.06 |     | .60  |
| 11. I often feel very vulnerable when relations become more personal               | 2.09 | 0.99 | .63            |     | -.48 | .41 | .10  |
| 12. I often do not succeed in cooperating with others in mutually satisfactory way | 0.69 | 0.89 | .66            |     | .18  |     | .67  |

Note: S1/F1= self-functioning; S2/F2= interpersonal functioning. Correlation between F1 and F2 = .66.

Residual correlation between item 8 and 10 in the CFA = -.27; residual correlation between item 10 and 11 = .25.

Factor loadings of the CFA were all significant at the  $<.001$  level, except for item 11 on F2 ( $p=.02$ ).

The value-added ratio equaled 1.04 for the self scale, and 1.07 for the interpersonal scale. Both values are lower than the threshold of 1.1, although not to a large degree.

### Comparing factor models with one factor and two factors

The CFA model in which all items loaded on a single factor yielded poor model fit (RMSEA = .106 (90% C.I.: 0.101–0.112), CFI=.854, TLI=.821), with a substantial amount of item residual covariance. All factor loadings were significant at the <.001 level. The scale reliability estimated by McDonald's Omega ( $\omega_t$ ) was 0.801 (90% C.I.: 0.789–0.813). However, this would be an overestimation of the reliability of true score variance, due to residual covariance contributing to systematic scale variance.

The CFA in which the self factor was defined by items 1–6 and the interpersonal factor by items 7–12, revealed mediocre model fit (RMSEA = .074 (90% C.I.: .069–.080), CFI=.930, TLI=.913). The most prominent reasons for model misfit seemed to be a cross-loading of item 11 on self (.408) and a cross-loading of item 2 on interpersonal (-.312). Adding these cross-loadings to the model improved model fit (RMSEA = .065 (90% C.I.: .060–.071), CFI=.948, TLI=.933). Further inspection of possible model modifications revealed substantial covariance between the residuals of item 8 and 10 ( $r=-.272$ ), and between item 10 and 11 ( $r=.253$ ). By allowing these residuals to covariate, model fit was further improved (RMSEA = .057 (90% C.I.: .051–.063), CFI=.962, TLI=.949). Additional minor modifications (cross-loadings and residual covariance) had low impact on the Chi-Square statistics and were therefore not considered. The final CFA model is given in Table 2. The most striking finding is that item 11 loaded more strongly on the self factor than on the interpersonal factor.

Scale reliability ( $\omega$ ) was .712 (90% C.I.: .664–.730) for the self-dimension and, correcting for residual covariance, .675 (90% C.I.: .654–.696) for the interpersonal dimension. Without this correction, scale reliability was .748 (90% C.I.: .732–.764). These values can be considered acceptable according to conventional rules for group comparisons, but questionable for individual interpretations (DeVellis, 2016; Evers et al., 2015; Nunnally & Bernstein, 1994; Pedhazur & Schmelkin, 1991).

### Concurrent validity

The results in Table 3 show that the LPFS-BF 2.0 total score had strong correlations with similar constructs. The total LPFS-BF 2.0 score was correlated with all SIPP facets, with stable self-image having the strongest correlation and intimacy having the weakest correlation. Some SIPP facets were somewhat more strongly correlated with self-functioning according to the LPFS-BF 2.0; i.e., self-respect, stable self-image, self-reflective functioning, and purposefulness. The SIPP facets respect, cooperation, and feeling recognized were somewhat more strongly associated with interpersonal functioning.

**Table 3.** Descriptive statistics of clinical measures and their correlations with LPFS-BF 2.0 scores.

|                                  | Descriptives |       |       | Correlations |        |               |
|----------------------------------|--------------|-------|-------|--------------|--------|---------------|
|                                  | N            | Mean  | SD    | Total        | Self   | Interpersonal |
| LPFS-BF 2.0 total score          | 1673         | 1.54  | 0.54  |              |        |               |
| LPFS-BF 2.0 self                 | 1673         | 1.90  | 0.61  | 0.856        |        |               |
| LPFS-BF 2.0 interpersonal        | 1672         | 1.19  | 0.64  | 0.869        | 0.490  |               |
| GFS functioning                  | 1505         | 54.49 | 7.72  | -0.137       | -0.162 | -0.076        |
| GFS symptoms                     | 1506         | 53.41 | 5.96  | -0.234       | -0.258 | -0.149        |
| WSAS                             | 1687         | 23.47 | 7.93  | 0.339        | 0.365  | 0.225         |
| PHQ-9                            | 1688         | 18.00 | 5.52  | 0.457        | 0.509  | 0.285         |
| GAD-7                            | 1674         | 13.15 | 4.73  | 0.445        | 0.432  | 0.340         |
| IIP                              | 1684         | 1.67  | 0.53  | 0.607        | 0.502  | 0.546         |
| <i>Number of PD criteria:</i>    |              |       |       |              |        |               |
| Paranoid criteria                | 1316         | 1.24  | 1.50  | 0.313        | 0.194  | 0.347         |
| Schizoid criteria                | 1316         | 0.35  | 0.81  | 0.050        | 0.015  | 0.071         |
| Schizotypal criteria             | 1316         | 0.47  | 0.85  | 0.133        | 0.070  | 0.162         |
| Antisocial criteria <sup>1</sup> | 1316         | 0.27  | 0.83  | 0.084        | 0.028  | 0.111         |
| Borderline criteria              | 1316         | 3.11  | 2.61  | 0.455        | 0.375  | 0.413         |
| Histrionic criteria              | 1316         | 0.37  | 0.83  | 0.185        | 0.119  | 0.198         |
| Narcissistic criteria            | 1316         | 0.29  | 0.81  | 0.230        | 0.116  | 0.280         |
| Avoidant criteria                | 1316         | 2.72  | 2.18  | 0.154        | 0.207  | 0.065         |
| Dependent criteria               | 1316         | 1.32  | 1.52  | 0.180        | 0.206  | 0.110         |
| Obs.-comp. criteria              | 1316         | 1.07  | 1.41  | 0.163        | 0.123  | 0.163         |
| Total PD criteria <sup>2</sup>   | 1316         | 11.21 | 6.64  | 0.463        | 0.378  | 0.425         |
| <i>SIPP facets<sup>3</sup>:</i>  |              |       |       |              |        |               |
| Emotion regulation               | 1666         | 29.64 | 13.12 | -0.619       | -0.539 | -0.528        |
| Effortful control                | 1667         | 37.02 | 12.92 | -0.568       | -0.459 | -0.519        |
| Self-respect                     | 1667         | 24.07 | 12.00 | -0.531       | -0.605 | -0.321        |
| Stable self-image                | 1667         | 30.46 | 12.93 | -0.678       | -0.654 | -0.517        |
| Self-reflective functioning      | 1665         | 32.56 | 12.10 | -0.649       | -0.638 | -0.486        |
| Enjoyment                        | 1667         | 28.51 | 11.31 | -0.457       | -0.479 | -0.313        |
| Purposefulness                   | 1667         | 25.34 | 13.33 | -0.555       | -0.610 | -0.353        |
| Intimacy                         | 1666         | 38.03 | 11.51 | -0.382       | -0.301 | -0.359        |
| Enduring relationships           | 1667         | 32.01 | 11.21 | -0.522       | -0.400 | -0.500        |
| Feeling recognized               | 1667         | 33.74 | 10.66 | -0.619       | -0.483 | -0.580        |
| Responsible industry             | 1667         | 35.98 | 13.59 | -0.420       | -0.351 | -0.371        |
| Trustworthiness                  | 1667         | 36.05 | 15.36 | -0.465       | -0.369 | -0.431        |
| Aggression regulation            | 1667         | 40.50 | 15.59 | -0.507       | -0.322 | -0.545        |
| Frustration tolerance            | 1667         | 31.78 | 10.37 | -0.593       | -0.524 | -0.498        |
| Cooperation                      | 1667         | 34.35 | 13.68 | -0.539       | -0.355 | -0.573        |
| Respect                          | 1667         | 42.52 | 13.39 | -0.522       | -0.300 | -0.593        |

Note: (1) Only antisocial adult criteria; (2) total number of SCID-5-PD criteria without the antisocial adolescence criteria; (3) SIPP facets is measured in T-scores based on Norwegian population norms (Pedersen et al., 2019). Correlations above |.075| were significant at the 0.01 level (2-tailed).

The correlation with self-reported depressive (PHQ-9) and anxiety symptoms (GAD-7) was also quite strong but less strong than for most SIPP facets. It should be noted that self-reported depressive symptoms (PHQ-9) were considerably more strongly correlated with self-functioning than with interpersonal functioning. The relationship between LPFS-BF 2.0 total score and measures of other types of functional impairments was only moderate (WSAS) or weak (GFS Functioning).

The total number of PD criteria, as assessed by the SCID-5-PD, correlated quite strongly with the total LPFS-BF 2.0 score ( $r = .463$ ). However, except for the paranoid and borderline PD criteria, the specific PD criteria correlated weakly with the total LPFS-BF 2.0 score. Of note, the correlation with the avoidant PD criteria was only .154.

## Discussion

In this study, we explored the construct validity of the Norwegian version of the LPFS-BF 2.0 in a large clinical



sample, with a special focus on the dimensionality of the instrument. The results showed moderate to good support for assuming essential unidimensionality for this instrument, when used in a clinical setting. A certain degree of multidimensionality was present; however, the subscales provided only a small amount of reliable unique variance. The suggestion that it might be useful to consider the LPFS-BF 2.0 as a unidimensional scale was further strengthened by the finding that there were meaningful associations between the total LPFS-BF 2.0 score and other clinical measures; notably, self-reported interpersonal problems, self-reported impairment of personality functioning, and total number of PD criteria as assessed by the SCID-5-PD.

Our findings are in line with those found by a previous study using bifactor analysis. In a community sample of 924 participants, Zimmermann et al. (2020) conducted bifactor analyses on a number of self-report instruments for PD, including the LPFS-BF 2.0, and found a strong general factor for these instruments. When we compared our findings to those reported in previous studies that used other types of factor analysis than bifactor analysis, we found both differences and similarities. For instance, like Weekers et al. (2019), who used the original Dutch version of the LPFS-BF 2.0 in a sample of 201 psychiatric patients, we found that a two-factor solution showed better fit than a one-factor solution when using CFA modeling. Furthermore, like Weekers et al., we found that items 10 and 11 still were correlated after the effect of the interpersonal factor was accounted for<sup>2</sup>. Another corresponding finding between the study of Weekers et al. and our study was that model fit improved by allowing a cross-loading of item 11 on the Self factor<sup>3</sup>. This high degree of overlap of findings is significant, since the composition of the current sample is highly similar to the sample used by Weekers et al.; i.e., a clinical sample with high prevalence of avoidant and borderline PD. This would suggest further tentative support for the cross-cultural validity of the instrument. However, the correlation we found between the self and interpersonal factor was substantially larger (0.66) than the one found by Weekers et al. (0.44). This also, at least partially, explains the differences in our conclusions (Weekers et al. favored a two-factor solution rather than treating the LPFS-BF 2.0 as essentially unidimensional). Bach and Hutsebaut (2018), on the other hand, found a correlation of very similar magnitude to ours (0.70) when investigating the dimensionality of the Danish version of the LPFS-BF 2.0. However, they used exploratory rather than confirmatory factor analysis, which makes it difficult to compare results directly. In a recent multi-national

study, Natoli et al. (2022) investigated measurement invariance across seven countries for the 2-factor model in community and student samples using CFA. The authors found support for full scalar invariance in the community samples, and for partial scalar invariance in the student samples. Furthermore, they reported various significant differences in latent means across the samples. Interestingly, the correlation between the self and interpersonal factors reported in this study is very high across all included samples (ranging from .74–.90). This might not be very surprising since measures of psychological functioning are often more unidimensional in samples with relatively low levels of psychopathology than in patient samples (e.g., Paap et al., 2012), but it does beg the question whether a unidimensional approach may not be more appropriate here. The authors did find fairly high reliability estimates across the samples; although it needs to be pointed out that this was predominantly the case for the self scale (10 out of 10 estimates exceeded .8). For the interpersonal scale, only 4 out of 10 estimates exceeded .8, and the lowest was as low as .5. Before drawing firm conclusions regarding the use (or further development) of the subscales across countries and populations, more research – preferably including clinical samples from various countries and employing different types of dimensionality analysis – is needed.

Our results pertaining to concurrent validity were somewhat mixed. Although the correlation between the LPFS-BF 2.0 score and total number of PD criteria was rather large, the correlations between the LPFS-BF 2.0 scores and the number of specific PD criteria were small, except for the borderline PD criteria. The positive correlation between the borderline PD criteria and the LPFS-BF 2.0 does not come as a surprise, since there is substantial overlap between the content of the borderline PD criteria and the descriptions within the LPFS, i.e., identity, emotional dysregulation, and intimacy, making it easier to create instruments that capture borderline PD. The small correlations between the LPFS-BF 2.0 and the number of specific PD criteria pertaining to other PDs may partly be explained by the relatively small range of personality problems defining these individual PDs. This is especially relevant for Avoidant PD, the most common PD in the Norwegian Network for Personality Disorders, since the avoidant PD criteria are rather uniform, describing a person whose is socially inhibited and tends to avoid social situations because of poor self-esteem and/or fear of being denigrated. In clinical samples such as this one, comorbidity can be expected to confound results as well. In this particular sample, however, as much as 65% of patients with avoidant PD were not diagnosed with any other PD. Therefore, we deem it unlikely that comorbidity was the main reason for the low correlations.

A more plausible explanation for the low correlation between the LPFS-BF 2.0 and the number of Avoidant PD criteria, in our view, may be that patients with Avoidant PD are likely to underestimate problems in the interpersonal domain due to their lack of exposure to interpersonal situations. The content of item 9 (“I often do not fully understand why my behavior has a certain effect on others”) is

<sup>2</sup>From a CFA perspective, this could be interpreted as reflecting a misspecification of the model. However, the residual correlation between items 10 and 11 could also be attributed to the fact that the general factor was not accounted for in the CFA model with two factors. This position is supported by the finding that in the bifactor model, items 10 and 11 loaded substantially on the general factor.

<sup>3</sup>From a theoretical point of view, such a cross-loading is not necessarily problematic, since it illustrates the close connection between self-functioning and interpersonal functioning. Similar results were reported by Bliton et al. (2022), who found that items 10 and 11 in the original LPFS-BF loaded strongly on the general factor and negatively on the residual interpersonal factor.

also unlikely to resonate well with these patients. Of note, the content of the corresponding subdomain in the LPFS is not only related to the understanding of effects of own behavior on others but also to lack of awareness of these effects; especially with respect to level 2 and 3 (e.g., “Is generally unaware of or unconcerned about effect of own behavior on others”; a level 2 description). Some patients with severe avoidant PD in our sample may feel that their behavior has no impact on other people at all, which could be considered as equivalent to not being aware of this effect. For persons who are not aware of the impact of own behavior on others, it might be difficult to answer reliably on questions about understanding why their behavior has a certain effect on others.

The total LPFS-BF 2.0 score had strong correlations with the SIPP facets, highlighting the conceptual overlap between the LPFS and the SIPP. These correlations were strongest for stable self-image and self-reflective functioning and weakest for intimacy. There were some notable differences in correlation patterns between self-functioning and interpersonal functioning with regard to the SIPP facets. For instance, compared with interpersonal functioning, self-functioning was more strongly related to the “self-facets” of the SIPP; i.e., self-respect, stable self-image, and self-reflective functioning. Self-functioning was also more strongly related to the SIPP facet purposefulness. This could be explained by overlap in content between self-direction according to the LPFS and purposefulness according to the SIPP. For instance, LPFS-BF 2.0 item 4 (“I have no sense of where I want to go to in my life”) is very similar to SIPP item 114 (“One of my problems is that I lack clear goals in my life”). Likewise, interpersonal functioning according to the LPFS-BF 2.0 was more strongly related to the SIPP facets cooperation and respect, affirming the fact that this aspect of the LPFS revolves around the inability to cooperate with others and respect others’ views and opinions.

The rather low correlation between the intimacy facet of the SIPP and interpersonal functioning according to the LPFS-BF 2.0 is somewhat surprising since intimacy is a central part of the interpersonal component of the LPFS. However, there are notable differences in content between the intimacy descriptions of these two instruments. In the SIPP, intimacy problems are represented by difficulties in expressing affection and sharing thoughts and feelings with others. In the LPFS, the focus is more on the capability to cooperate with others and to establish and maintain long-lasting relationships. It seems that the LPFS-BF 2.0 covers this content quite well. However, it may be advisable to add items in a future revision that capture the preoccupation some patients have with how others may judge or criticize them.

## Conclusion

On the whole, we found moderate to good support for the use of total scores for the Norwegian version of the LPFS-BF 2.0. The ECV associated with the general factor was high, and so was the correlation between the Self and Interpersonal scale in the two-factor CFA. It should be

noted, however, that the evidence in favor of the general factor would ideally have been stronger. It has been suggested that both the ECV and PUC should be high, since this combination leads to the smallest risk of factor loadings being biased when multidimensionality is ignored (Rodriguez et al., 2016). In this study, the PUC value was moderate, which implies that ignoring the multidimensionality present in the LPFS-BF 2.0 may lead to some bias. Furthermore, the Omega-H estimate for the general factor equaled .74, which is somewhat lower than the value recommended by Rodriguez et al. (.80). The CFA analyses were inconclusive; the fit of a 1-factor model was inadequate, and the fit of a 2-factor model only became adequate when we allowed for cross-loadings and covariances among residuals. The distinctiveness of the subscales (and hence their added value over and above a total scale) was too low, but not negligible. The concurrent validity for the LPFS 2.0 total score was generally supported by the strong association between the total score and other self-report measures assessing interpersonal problems, symptom distress, impairment of personality functioning as assessed by the SIPP and reduced psychosocial functioning. There was also a strong correlation between the LPFS-BF 2.0 and total number of PD criteria, as assessed by the SCID-5-PD, and between the LPFS-BF 2.0 and the borderline PD criteria. However, for most other PD criteria, this correlation was weak. This finding is especially relevant for avoidant PD, since this PD was the most common PD in this sample. At present then, identifying patients with avoidant PD from the perspective of the AMPD may warrant the use of instruments that are designed for this purpose such as the Avoidant Personality Disorder Impairment Questionnaire (Liggett et al., 2017).

Based on our findings, we would advise against the use of subscale scores for clinical assessment purposes. Although the Omega values estimated under the bifactor model for the specific factors were adequate (close to .8 for the self factor and slightly larger than .8 for the interpersonal factor), the corresponding Omega-H values were substantially lower for both factors; for the interpersonal factor, Omega-H was even close to 0. This suggests that the subscales provide only a small amount of reliable unique variance. Furthermore, the subscales did not show convincing levels of distinctiveness.

An important next step would be to investigate how the items are interpreted by patients by using think-aloud techniques. One of the drawbacks of using self-report is that the clinician has very little insight into whether the patient actually interprets an item as intended, let alone what their thought process is when they answer the question. Variations in these processes across patients can have impactful consequences for both reliability and validity. This needs to be investigated in future studies. Furthermore, it may be advisable to add items to the LPFS-BF 2.0 that cover crucial aspects of avoidant PD so that these patients are not overlooked.

We would like to note that no clinician-rated instruments for the assessment of the LPFS were available for use in the Norwegian Network for Personality Disorders at the time of data collection for this study. It is of high importance that future studies compare self-reported impairment of personality

functioning as measured with the LPFS-BF 2.0 with clinician-rated impairment, such as measured by for example the SCID-5-AMPD-I (Bender et al., 2018), in order to examine whether self-report instruments and clinical interviews yield comparable outcomes.

## Acknowledgments

We wish to thank the patients and staff of the following units of the Norwegian Network for Personality Disorders for their contribution to this study: Unit for Group Therapy, Øvre Romerike District Psychiatric Center, Akershus University Hospital, Jessheim; Group Therapy Unit, Nedre Romerike District Psychiatric Center, Akershus University Hospital, Lillestrøm; Group Therapy Unit, Follo District Psychiatric Center, Akershus University Hospital, Ski; Group Therapy Unit, Groruddalen District Psychiatric Center, Akershus University Hospital, Oslo; Clinic for Personality disorders, Outpatient Clinic for Specialized Treatment of Personality Disorders, Section for Personality psychiatry and specialized treatments, Oslo University Hospital, Oslo; Group Therapy Unit, Lovisenberg District Psychiatric Center, Lovisenberg Hospital, Oslo; Group Therapy Team, Vinderen Psychiatric Center, Diakonhjemmet Hospital, Oslo; Unit of Personality psychiatry, Vestfold District Psychiatric Center, Sandefjord; Unit for Intensive Group Therapy, Aust-Agder District Psychiatric Center, Sørlandet Hospital, Arendal; Unit for Group Therapy, District Psychiatric Center, Strømme, Sørlandet Hospital, Kristiansand; Group Therapy Unit, Stavanger District Psychiatric Center, Stavanger University Hospital, Stavanger; Section for group treatment, Kronstad District Psychiatric Center, Haukeland University Hospital, Bergen; MBT Team, Department of Substance Abuse Medicine, Haukeland universitetssjukehus, Bergen; MBT-Team, Outpatient Clinic, Rogaland A-senter, Stavanger; and Group Therapy Unit, Ålesund District Psychiatric Center, Ålesund Hospital, Ålesund.

## Disclosure statement

None of the authors has any financial disclosure or other conflict of interest related to this manuscript.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a FRIPRO Young Research Talent grant for the first author (Grant no. NFR 286893) awarded by the Research Council of Norway.

## ORCID

Muirne C. S. Paap  <http://orcid.org/0000-0002-1173-7070>  
 Geir Pedersen  <http://orcid.org/0000-0002-8599-8871>  
 Elfrida Kvarstein  <http://orcid.org/0000-0001-9221-4249>  
 Benjamin Hummelen  <http://orcid.org/0000-0002-8717-8076>

## Data availability statement

Due to restrictions imposed by the Regional Medical Ethics Committee regarding patient confidentiality, data are only available upon reasonable request. Requests for data may be sent to the Privacy and Data Protection Officer at Oslo University Hospital in Oslo, Norway.

## References

- Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the inventory of interpersonal problems. *Journal of Personality Assessment*, 55(3–4), 521–536.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders: DSM-IV*. American Psychiatric Association.
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). American Psychiatric Association.
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders, 5th edition: DSM-5*. American Psychiatric Association.
- Andrea, H., Verheul, R., Berghout, C., Dolan, C., van der Kroft, P., Bateman, A., Fonagy, P., & van Busschbach, J. (2007). *Measuring the core components of maladaptive personality: Severity Indices of Personality Problems (SIPP-118)*. (No. Report 005). Medical Psychology and Psychotherapy: Reports. Retrieved from <http://hdl.handle.net/1765/10066>
- Arnevik, E. A., Pedersen, G., Walderhaug, E., Lien, I., Wilberg, T., & Hummelen, B. (2019). Measuring personality problems in patients with substance use disorders: A cross-sample validation. *Journal of Dual Diagnosis*, 15(4), 324–332. <https://doi.org/10.1080/15504263.2019.1668583>
- Arnevik, E. A., Wilberg, T., Monsen, J. T., Andrea, H., & Karterud, S. (2009). A cross-national validity study of the Severity Indices of Personality Problems (SIPP-118). *Personality and Mental Health*, 3(1), 41–55. <https://doi.org/10.1002/pmh.60>
- Bach, B., & Hutsebaut, J. (2018). Level of Personality Functioning Scale – Brief form 2.0: Utility in capturing personality problems in psychiatric outpatients and incarcerated addicts. *Journal of Personality Assessment*, 100(6), 660–670. <https://doi.org/10.1080/00223891.2018.1428984>
- 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. S., Skodol, A. E., First, M. B., & Oldham, J. M. (2018). *Structured clinical interview for the DSM-5 alternative model for personality disorders (SCID-5-AMPD); module I: Level of Personality Functioning Scale*. American Psychiatric Association Publishing.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588–606. <https://doi.org/10.1037/0033-2909.88.3.588>
- 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](https://doi.org/10.1521/pedi_2021_35_531)
- Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, 75(4), 581–612. <https://doi.org/10.1007/s11336-010-9178-0>
- Calderón Garrido, C., Navarro González, D., Lorenzo Seva, U., & Ferrando Piera, P. J. (2019). Multidimensional or essentially unidimensional? A multi-faceted factor-analytic approach for assessing the dimensionality of tests and items. *Psicothema*, 31(4), 450–457. <https://doi.org/10.7334/psicothema2019.153>
- Chalmers, R. P. (2012). MIRT: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29. <https://doi.org/10.18637/jss.v048.i06>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Sage Publications.
- Dowgwillo, E. A., Roche, M. J., & Pincus, A. L. (2018). Examining the interpersonal nature of criterion A of the DSM-5 section III alternative model for personality disorders using bootstrapped confidence intervals



- for the interpersonal circumplex. *Journal of Personality Assessment*, 100(6), 581–592. <https://doi.org/10.1080/00223891.2018.1464016>
- Evers, A., Lucassen, W., Meijer, R. R., Sijtsma, K. (2015). COTAN review system for evaluating test quality. <https://psynip.nl/wp-content/uploads/2022/05/COTAN-review-system-for-evaluating-test-quality.pdf>
- Feenstra, D. J., Hutsebaut, J., Verheul, R., & Busschbach, J. J. (2011). Severity Indices of Personality Problems (SIPP-118) in adolescents: Reliability and validity. *Psychological Assessment*, 23(3), 646–655. <https://doi.org/10.1037/a0022995>
- Feinberg, R. A., & Jurich, D. P. (2017). Guidelines for interpreting and reporting subscores. *Educational Measurement: Issues and Practice*, 36(1), 5–13. <https://doi.org/10.1111/emip.12142>
- First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2016). *Structured clinical interview for DSM-5 clinical version (SCID-5-PD)*. American Psychiatric Association.
- Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, 3(4), 484–501. <https://doi.org/10.1177/2515245920951747>
- Fonagy, P., Gergely, G., Jurist, E. L., & Target, M. (2018). *Affect regulation, mentalization, and the development of the self*. Routledge.
- 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: Theory, Research, and Treatment*, 10(5), 438–447. <https://doi.org/10.1037/per0000335>
- Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and Behavioral Statistics*, 33(2), 204–229. <https://doi.org/10.3102/1076998607302636>
- Hu, L., & Bentler, P. (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: Theory, Research, & Treatment*, 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>
- IBM (2019). *SPSS Statistics for Windows, Version 26.0*. IBM Corp.
- Karterud, S., Pedersen, G., Bjordal, E., Brabrand, J., Friis, S., Haaseth, O., Haavaldsen, G., Irion, T., Leirvag, H., Torum, E., & Urnes, O. (2003). Day treatment of patients with personality disorders: Experiences from a Norwegian treatment research network. *Journal of Personality Disorders*, 17(3), 243–262. <https://doi.org/10.1521/pedi.17.3.243.22151>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- Liggett, J., Carmichael, K. L., Smith, A., & Sellbom, M. (2017). Validation of self-report impairment measures for section III obsessive-compulsive and avoidant personality disorders. *Journal of Personality Assessment*, 99(1), 1–14. <https://doi.org/10.1080/00223891.2016.1185613>
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>
- McDonald, R. P. (2011). *Test theory: A unified treatment*. Routledge.
- Morey, L. C. (2017). Development and initial evaluation of a self-report form of the DSM-5 level of Personality Functioning Scale. *Psychological Assessment*, 27(10), 1302–1308.
- Morey, L. C. (2019). Thoughts on the assessment of the DSM-5 alternative model for personality disorders: Comment on Sleep et al. (2019). *Psychological Assessment*, 31(10), 1192–1199. <https://doi.org/10.1037/pas0000710>
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. H. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *British Journal of Psychiatry*, 180(5), 461–464. <https://doi.org/10.1192/bjp.180.5.461>
- Muthén, L. K., & Muthén, B. O. (1998–2015). *Mplus Users Guide* (7th ed.). Muthén & Muthén. Available online at <https://www.statmodel.com/>
- Natoli, A. P., Bach, B., Behn, A., Cottin, M., Gritti, E. S., Hutsebaut, J., Lamba, N., Le Corff, Y., Zimmermann, J., & Lapalme, M. (2022). Multinational evaluation of the measurement invariance of the Level of Personality Functioning Scale – Brief form 2.0: Comparison of student and community samples across seven countries. *Psychological Assessment*, 34(12), 1112–1125. <https://doi.org/10.1037/pas0001176>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.
- Paap, M. C. S., Hummelen, B., Braeken, J., Arnevik, E. A., Walderhaug, E., Wilberg, T., Berghuis, H., Hutsebaut, J., & Pedersen, G. (2021). A multi-center bi-factor analysis of the Severity Indices of Personality Problems 118 (SIPP-118): Do we really need all those facets? *Quality of Life Research*, 30(2), 567–575. <https://doi.org/10.1007/s11136-020-02654-8>
- Paap, M. C. S., Meijer, R. R., Cohen-Kettenis, P. T., Richter-Appelt, H., de Cuyper, G., Kreukels, B. P. C., Pedersen, G., Karterud, S., Malt, U. F., & Haraldsen, I. R. (2012). Why the factorial structure of the SCL-90-R is unstable: Comparing patient groups with different levels of psychological distress using Mokken scale analysis. *Psychiatry Research*, 200(2–3), 819–826. <https://doi.org/10.1016/j.psychres.2012.03.012>
- Pedersen, G. (2002). Revised Norwegian version of inventory of interpersonal problems – Circumplex (IIP-C). *Tidsskrift for Norsk Psykologforening*, 39, 25–34.
- Pedersen, G., Arnevik, E. A., Hummelen, B., Walderhaug, E., & Wilberg, T. (2019). Psychometric properties of the Severity Indices of Personality Problems (SIPP) in two samples: A Norwegian community sample and clinical samples of patients with and without personality disorders. *European Journal of Psychological Assessment*, 35(5), 698–711. <https://doi.org/10.1027/1015-5759/a000436>
- Pedersen, G., Hagtvet, K. A., & Karterud, S. (2011). Interpersonal problems: Self-therapist agreement and therapist consensus. *Journal of Clinical Psychology*, 67(3), 308–317. <https://doi.org/10.1002/jclp.20762>
- Pedersen, G., & Karterud, S. (2012). The symptom and function dimensions of the Global Assessment of Functioning (GAF) scale. *Comprehensive Psychiatry*, 53(3), 292–298. <https://doi.org/10.1016/j.comppsy.2011.04.007>
- Pedersen, G., Karterud, S., Hummelen, B., & Wilberg, T. (2013). The impact of extended longitudinal observation on the assessment of personality disorders. *Personality and Mental Health*, 7(4), 277–287. (Published online in Wiley Online Library DOIpar. <https://doi.org/10.1002/pmh.1234>)
- Pedersen, G., Kvarstein, E. H., & Wilberg, T. (2017). The Work and Social Adjustment Scale: Psychometric properties and validity among males and females, and outpatients with and without personality disorders. *Personality and Mental Health*, 11(4), 215–228. <https://doi.org/10.1002/pmh.1382>
- Pedersen, G., Urnes, Ø., Hummelen, B., Wilberg, T., & Kvarstein, E. H. (2018). Revised manual for the Global Assessment of Functioning scale. *European Psychiatry: The Journal of the Association of European Psychiatrists*, 51, 16–19. <https://doi.org/10.1016/j.eurpsy.2017.12.028>
- Pedersen, G., Wilberg, T., Hummelen, B., & Kvarstein, E. H. (2022). The Norwegian network for personality disorders – Development, contributions and challenges through 30 years. *Nordic Journal of Psychiatry*, 1–9. <https://doi.org/10.1080/08039488.2022.2147995>
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Psychology Press.
- Quinn, H. O. C. (2014). *Bifactor models, explained common variance (ECV), and the usefulness of scores from unidimensional item response theory analyses* [Master thesis]. <https://doi.org/10.17615/t6ff-a088>



- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>
- 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>
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223–237. <https://doi.org/10.1080/00223891.2015.1089249>
- Prevolnik Rupel, V., Jagger, B., Fialho, L. S., Chadderton, L. M., Gintner, T., Arntz, A., Baltzersen, Å. L., Blazdell, J., van Busschbach, J., Cencelli, M., Chanen, A., Delvaux, C., van Gorp, F., Langford, L., McKenna, B., Moran, P., Pacheco, K., Sharp, C., Wang, W., Wright, K., ... Crawford, M. J. (2021). Standard set of patient-reported outcomes for personality disorder. *Quality of Life Research*, 30(12), 3485–3500. <https://doi.org/10.1007/s11136-021-02870-w>
- Sheehan, D. V., Lecrubier, Y., Janavs, J., Knapp, E., Weiller, E., & Bonora, L. I. (1994). *Mini International Neuropsychiatric Interview (MINI)*. University of South Florida Institut for Research in Psychiatry and INSERM-Hôpital de la Salpêtrière.
- 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>
- Sleep, C. E., Lynam, D. R., Widiger, T. A., Crowe, M. L., & Miller, J. D. (2019). An evaluation of DSM-5 section III personality disorder criterion A (impairment) in accounting for psychopathology. *Psychological Assessment*, 31(10), 1181–1191. <https://doi.org/10.1037/pas0000620>
- Spitzer, R. L. (1983). Psychiatric diagnosis: Are clinicians still necessary? *Comprehensive Psychiatry*, 24(5), 399–411. [https://doi.org/10.1016/0010-440X\(83\)90032-9](https://doi.org/10.1016/0010-440X(83)90032-9)
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166(10), 1092–1097. <http://www.ncbi.nlm.nih.gov/pubmed/16717171>
- Stanton, K., Brown, M. F., Bucher, M. A., Balling, C., & Samuel, D. B. (2019). Self-ratings of personality pathology: Insights regarding their validity and treatment utility. *Current Treatment Options in Psychiatry*, 6(4), 299–311. <https://doi.org/10.1007/s40501-019-00188-6>
- Steiger, J. H. (1990). Structural model equation and modification: An internal estimation approach. *Multivariate Behavioral Research*, 25(2), 173–180. [https://doi.org/10.1207/s15327906mbr2502\\_4](https://doi.org/10.1207/s15327906mbr2502_4)
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5), 893–898. <https://doi.org/10.1016/j.paid.2006.09.017>
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*, 7, 769. <https://doi.org/10.3389/fpsyg.2016.00769>
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1–10. <https://doi.org/10.1007/BF02291170>
- Verheul, R., Andrea, H., Berghout, C. C., Dolan, C., Busschbach, J. J., van der Kroft, P. J., Bateman, A. W., & Fonagy, P. (2008). Severity Indices of Personality Problems (SIPP-118): Development, factor structure, reliability, and validity. *Psychological Assessment*, 20(1), 23–34. <https://doi.org/10.1037/1040-3590.20.1.23>
- 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>
- 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>

## Appendix A. Descriptions of the SIPP facets

*Emotion regulation* – The capacity to tolerate and manage the emotions you have and to control their intensity, course, and expression; i.e., “I usually have adequate control over my feelings”

*Effortful control* – The ability to focus concentration and direct impulses through conscious effort; i.e., “I seldomly get so excited that I lose control over myself”

*Self-respect* – The capacity to feel that you are worthy and to know that others or yourself have no right to harm you physically or emotionally; i.e., “I feel proud of some things I have accomplished in my life”

*Stable self-image* – To experience an inner sense of continuity/sameness of self across time and situations; i.e., “I know exactly who I am and what I am worth”

*Self-reflective functioning* – The capacity to understand the possible meanings of, and causal connections between, internal and external experiences, as well as the ability to identify reasons for things happening within yourself rather than constantly trying to find answers in the world outside; i.e., “Most of the time, I understand why I do the things I do”

*Enjoyment* – The capacity to enjoy without feeling guilty; i.e., “Overall I feel that my activities are enjoyable to me”

*Purposefulness* – The capacity to make life meaningful by creating the means as well as the opportunities for achievement and organizing time in line with one's goals; i.e., “I strongly believe that life is worth living”

*Intimacy* – The ability to share sensitive personal experiences with other people; i.e., “I enjoy intimate contacts with other people”

*Enduring relationships* – The capacity to love and to feel loved in order to form and maintain long-term, intimate relationships; also referred to as the capacity for ‘healthy attachment’; i.e., “I have been able to form lasting friendships”

*Feeling recognized* – The experience that others understand what you feel and believe; i.e., “My friends are really interested in my well-being”

*Responsible industry* – The capacity to set realistic goals and to achieve these through effective and responsible constructive actions; i.e., “Most of the time I try to perform tasks that are assigned to me conscientiously”

*Trustworthiness* – That one has internalized the values and norms of social collaboration and is usually able to behave in accordance to these; i.e., “When I have promised to do something I will always try to keep that promise”

The capacity to share sensitive personal experiences, to love, and to feel loved and recognized in order to maintain long-term, intimate relationships.

*Aggression regulation* – The ability to withhold aggressive impulses toward others; i.e., “It is hard for me to control my aggression towards others”

*Frustration tolerance* – The capacity to cope with disappointments and setbacks; i.e., “I can cope very well with disappointments”

*Cooperation* – The ability to work constructively with others, to be aware of the needs and ideas of others, and to establish mutual goals; i.e., “I like to create something together with other people”

*Respect* – The capacity to value others individual needs and personal identity; i.e., “I can easily accept people the way they are, even when they are different”

From Appendix A, pp. 710–711, in Pedersen et al. (2019). Reprinted with permission from the copyright holder (De Viersprong).