Psychometric Properties of the Severity Indices of Personality Problems (SIPP) in a Norwegian Community Sample and in Clinical Samples of Patients with and without Personality Disorders

Pedersen, G.^{1,2}, Arnevik E. A.^{3,4}, Hummelen B.⁵, Walderhaug E.⁶, & Wilberg T.^{5,7}

From: 1) Department of Personality Psychiatry, Division of Mental Health and Addiction, Oslo University Hospital, Norway. 2) NORMENT, Institute of Clinical Medicine, University of Oslo, Norway, 3) National Advisory Unit on SUD Treatment, Oslo University Hospital, Norway, 4) Institute of Psychology, University of Oslo, Norway. 5) Department of Research and Development, Division of Mental Health and Addiction, Oslo University Hospital, Norway. 6) Department of addiction treatment, Oslo University Hospital, Norway. 7) Institute of Clinical Medicine, University of Oslo, Norway.

Correspondence: Geir Pedersen, Department of Personality Psychiatry, Division of Mental Health and Addiction. Oslo University Hospital, Ullevaal, PO Box 4956 Nydalen, 0424 Oslo, Norway. E-mail: geir.pedersen@medisin.uio.no

None of the authors have any financial disclosures or other conflicts of interest related to this manuscript.

Abstract

This study investigated psychometric properties of a dimensional measure of maladaptive core pathology for personality disorders, the Severity Indices of Personality Problems (SIPP). The study analyzed data from 941 respondents in a community sample, 949 psychiatric patients with Personality disorders (PD), and 413 psychiatric patients without PD in Norway. The facets of SIPP had acceptable internal consistency, but the construct validity of some facets had potentials for improvement. The original SIPP domains (the factor structures of the facets) were not supported, and factor analysis gave different results across the three current study samples. All facets of SIPP have good discriminative properties with respect to differentiating between a nonclinical sample, a clinical sample without PD, and a clinical PD sample. Further research and improvements on SIPP are suggested.

Keywords

Normative estimates; personality disorders; severity, SIPP; validity.

Introduction

Nearly a decade has passed since the first technical report was published (Andrea et al., 2007) describing the rationale, purpose, and development of the Severity Indices of Personality Problems (SIPP) – a self-report measure developed by Dutch and English researchers. This questionnaire has now formally been translated from Dutch into five languages (English, Norwegian, Spanish, Danish and Italian), and is increasingly reported in clinical research.

The SIPP was based on the theory that the development of personality is an interplay between biologically based, more stable personality traits (e.g. extraversion or openness to experience) and more environmentally influenced, adaptive personality capacities (e.g. the capacity of having control over impulses and emotions or keeping stable, close relationships). It is assumed that high adaptive capacity relates to healthy functioning and that personality disorder is characterized by deficient levels of adaptive capacities (Andrea et al., 2007). Variations in the adaptive capacities measured by SIPP are assumed to be strongly related to PD severity, irrespective of the specific type of PD (Verheul et al., 2008).

Several publications have supported the clinical relevance and utility of the SIPP (Arnevik et al., 2009; Bales et al., 2012; Bales et al., 2014; Feenstra et al., 2011; Feenstra, Hutsebaut, Verheul, & van Limbeek, 2014; Gudjonsson, Sigurdsson, Guðmundsdóttir, Sigurjónsdóttir, & Smari, 2010; Lien & Arnevik, 2016), its underlying assumption of representing changeable core aspects of PDs (Bastiansen et al., 2013; Berghuis, Kamphuis, & Verheul, 2012; Berghuis, Kamphuis, Verheul, Larstone, & Livesley, 2013; Feenstra et al., 2014; Verheul et al., 2008), and its ability to capture levels of PD severity (Berghuis, Kamphuis, & Verheul, 2012; Morey et al., 2011). There is also support for cross-national equality of mean scores on the SIPP in The Netherlands and Norway with respect to PDs

(Arnevik et al., 2009). However, until now, estimates of normative references have been based on a non-clinical Dutch sample, and there is a current need for subsequent community studies and normative references from other countries. As higher SIPP scores indicate better functioning, it should be expected that psychiatric outpatients without PD revealed higher scores than patients with PD, and that a community sample should reveal higher scores than both.

The structure of SIPP includes 16 core components (termed facets). The selection of facets was originally founded on recognized concepts in clinical literature, actual data documentation, and a criterion of interpretability (Andrea et al., 2007). The identification of five higher order domains (self-control, social concordance, identity integration, relational functioning and responsibility, Appendix A), was based on a subsequent exploratory factor analysis.

The psychometric properties of the SIPP have been investigated in several studies and the internal consistency of the 16 theoretically derived facets have been found acceptable. Cronbach's alpha coefficients (Cronbach, 1951) ranged from .59–.89 for the different facets (Arnevik, Wilberg, Andrea, Monsen, & Karterud 2009; Feenstra, Hutsebaut, Verheul, & Busschbach, 2011; Verheul et al., 2008). However, the application of Cronbach's alpha is not always an appropriate method for assessing internal consistency or reliability of a measurement scale (Graham, 2006; Tavakol & Dennick, 2011). Further analyses on the facet level of SIPP are warranted in order to support generalizability of psychometric properties across samples.

The five higher order SIPP domains did not originally have a clear and empirically supported consensus, but were the result of exploratory factor analyses and thus data-driven. The computation was based on primary and secondary factor loadings as regression weights

for the facets combined with clinical judgment (Verheul et al., 2008). However, this strategy represents a considerable conceptual and statistical challenge. The domains are confounded by facet overlap and factor loadings do not necessarily replicate across samples (Beavers, Lounsbury, Richards, Huck, Skolits, et al., 2013; Floyd, & Widaman, 1995; Gregorich, 2006; Hirschfeld, von Brachel, & Thielsch, 2014). In order to support the higher order construction of five domains, the partly data-driven model (Andrea et al., 2007, Appendix A) needs to be supported by subsequent studies of different study samples. Currently, four different factor models have been proposed; three somewhat different five-factor models (Andrea et al., 2007; Verheul et al., 2008; Feenstra et al., 2011), and one four-factor model (Bastiaansen et al., 2013). With four deviating proposals for factor composition, further studies should assess the generalizability of these compositions in data including patients with PD (the target population of SIPP) and also other psychiatric patients without PD. If the underlying factor structure in former findings is not replicated, subsequent exploratory factor analysis should be conducted to assess the diversions.

The present study aims to investigate the validity and psychometric properties of the SIPP in three different samples; a community sample, a clinical sample of psychiatric patients without PD, and a clinical sample of psychiatric patients with PD. Starting with the core ingredients of SIPP, the first step will be a confirmatory factor analysis (CFA) conducted on the item level of each of the 16 SIPP facets. The second step will be to conduct a CFA of the four different factor solutions previously proposed. Anticipating the possibility that none of the four factor models may result in good fit, the third step involves examining the factor structure of the 16 SIPP facets by way of an exploratory factor analyses (EFA). All analyses will be conducted separately for each of the three current samples. After having disclosed the psychometric properties of SIPP, further analysis will investigate mean facet scores and comparisons between the three Norwegian samples (community and psychiatric patients with

and without PD). Lastly, differences between the Norwegian and Dutch community samples will be described.

Method

Participants

The SIPP questionnaire was mailed to a random sample of 5000 inhabitants in all 19 counties of Norway. They were asked to complete the questionnaire and to return it anonymously to a professional data-collecting company (ViaScan Ltd.) The selection procedure was as follows: 1) 875 persons were randomly selected in each of four age groups: 18–30, 31–40, 41–50, and 51–65 years (a total of 3500 addresses). 2) To adjust for an anticipated skewed response rate for subpopulations, an additional 500 men, 500 young people (age 18–25), and 500 people in the capital of Norway (Oslo), were selected (a total of extra 1500 addresses). Of the 5000 mailed questionnaires, 247 were returned due to ‘wrong recipient’, leaving a total of 4753 questionnaire invitations. After one reminder, 941 (19.8%) recipients returned the questionnaire, 54% of whom were women. One respondent had a missing value for gender. Mean age in the community sample was 44 years ($SD = 14$), and most of the respondents had high school or college/university education (Table 1). Based on the sample strategy an equal amount of respondents should be expected in each age intervals. From this, participants younger than the age of 30 might be underrepresented, whereas participants older than the age of 50 might be overrepresented in the community sample. According to Norwegian Statistics (www.ssb.no) around 31% of the population have an advanced educational level (college/university). In the current sample this proportion was 49%, indicating a possible bias.

- Insert Table 1 approx. here -

The total clinical sample included data from 1362 psychiatric outpatients who were consecutively admitted to 13 treatment units within the Norwegian Network of Personality-Focused Treatment Programs (Karterud et al., 1998) between January 2009 and March 2015. According to the DSM-IV (APA, 1994), 70% of the patients had one or more PDs, and 92% had one or more symptom disorders; of the latter, 48% had a depressive disorder, 52% had an anxiety disorder, and 6% had a substance-related disorder. Further details regarding the sociodemographic and diagnostic characteristics have been reported previously by Pedersen and Karterud (2010).

Among the PD patients ($n = 949$), mean age was 32 years ($SD = 9$) and the majority were females (73%). Among patients without PD ($n = 413$), mean age was 36 years ($SD = 10$) and the majority (72%) were also females (Table 2). There are no exact records of the educational level within the clinical sample. Our only estimate is based on a single item in the initial questionnaire: "How many years of education have you completed after junior high school?" (i.e. from the age of 16). The mean number of self-reported years was four ($SD = 3$) among those with PD (PD sample) and five ($SD = 3$) among those without PD (No-PD sample), indicating that the majority had an educational level of high school or higher. This might seem higher than would be expected from such a clinical sample. However, education is generally free in Norway, and mandatory from first to tenth grade. Considerable governmental efforts are given to help adolescents get through high school, and education grants are easy to obtain at university level.

All patients provided written informed consent for participation in the study. The individual treatment units collected patient data, and the data were registered in an anonymous central database administered by the Section of Personality Psychiatry, Oslo University Hospital. The State Data Inspectorate and the Regional Committee for Medical Research and Ethics approved the study procedures.

Instruments

The SIPP is a self-report questionnaire comprising 118 items rated on a four-point Likert scale that ranges from ‘fully disagree’ (1) to ‘fully agree’ (4). The SIPP facets each comprise 7–8 items that measure individual differences on the 16 facets and their five higher order domains. SIPP score interpretation is reference-based using norm-referenced standardized deviation scores (T-scores). Higher scores on the SIPP facets or domains are presumed to indicate more adaptive and thus less pathological capacities.

The SIPP was administered upon admission to treatment. All patients were diagnosed according to the DSM-IV using the Mini International Neuropsychiatric Interview (M.I.N.I.; Sheehan & Lecrubier, 1994) for symptom disorders and the Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II; First, Spitzer, Gibbon, Williams, 1995; First, Gibbon, Spitzer, Williams, & Benjamin, 1997) for PDs. The reliability of the diagnostic assessments was not investigated, but clinical staff in these units are thoroughly trained in diagnostic procedures and diagnoses were based on referral letters, self-reported complaints, patient history, and the two semi-structured interviews, SCID-II and M.I.N.I. (Pedersen, Karterud, Hummelen, & Wilberg, 2013). Table 2 shows the distribution of PDs in the patient sample.

- Insert Table 2 approx. here -

Statistical analysis

There were some missing items on the SIPP in both the community and patient samples. In total, 71% had no missing items, 15% had one missing item, and 98% had less than six missing items on the questionnaire. 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 from the study.

Using IBM SPSS Statistics for Windows, Version 19.0 (IBM, 2010), group differences were analyzed by the independent samples t-test (two-sided), internal consistency was estimated by Cronbach's alpha (Cronbach, 1951), exploratory factor analysis (EFA) of the facets of SIPP was conducted by Principal axis factoring with Promax rotation. Effect sizes of group differences were estimated by unbiased Hedges' g (Hedges, 1981). The magnitude of Hedges' g may be interpreted using Cohen's (1988) convention as small (0.2), medium (0.5), and large (0.8). Standard error of measurement of facet scores is computed by the formula: $s\sqrt{(1-r_{xx})}$, where s is the standard deviation of sample scores and r_{xx} is the reliability of the score, estimated by Cronbach's alpha. Confirmatory factor analysis (CFA) was conducted using Mplus 7.11 (Muthén & Muthén, 2012). The CFA was based on the robust maximum likelihood (MLR) function (Joop, Hox, Maas, & Brinkhuis, 2010; Yuan, & Bentler, 2000). By default Mplus uses FIML estimation with missing data (Enders, 2010).

To evaluate the CFA models, goodness of fit was estimated by the root mean square error of approximation (RMSEA; Steiger, 1990), by the comparative fit index (CFI; Bentler, 1990), by the Tucker-Lewis index (TLI; Tucker & Lewis, 1973), and by the standardized root mean square residual (SRMR; Hu & Bentler, 1999).

A RMSEA value of 0.05 or below indicates a good model fit, values between 0.05 and 0.08 indicate a reasonable fit, values between 0.08 and 0.10 a mediocre fit, and values greater than 0.10 an unacceptable fit (MacCallum, Browne, & Sugawara, 1996). The general consensus is to use a cut-off 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. The SRMR is the mean absolute value of the covariance residuals, and it ranges from 0 to 1.0. Well-fitting models should have values less than 0.05 (Byrne, 1998; Diamantopoulos & Siguaw, 2000), but values up to 0.08 are acceptable (Hu & Bentler, 1999).

Results

Confirmatory factor analysis and internal consistency of facets

CFA's were conducted on item levels of the facets in the three samples. In the community sample all indices of model fit were within the acceptable range, although the facet Purposefulness revealed marginal fit according to conventional rules (Table 3). In both of the patient samples, indices of fit were poor for the five facets Purposefulness, Stable self-image, Responsible industry, Intimacy and Enduring relationships (Tables 3 and 4). Additionally, Feeling recognized revealed poor indices of model fit in the data from the PD sample (Table 4). Poor model fit was mainly due to weak factor loadings, residual covariance, and multidimensionality.

A noteworthy finding in all three samples, was that three of the facets each seemed to represent two different constructs: Conceptually, the facet Purposefulness could be meaningfully divided into "appreciating life" and "interests and goals in life". Likewise, the facet Responsible industry encompassed both "responsibility" and "resilience", and the facet Frustration tolerance addressed both "good spirit" and "frustration". Although the facet Respect revealed the lowest internal consistency, as estimated by Cronbach's alpha, it revealed acceptable model fit in all three samples.

Measured by Cronbach's alpha, the internal consistency was in the acceptable range within all three samples, except for Respect within the community and No-PD sample (Tables 3 and 4).

- Insert Table 3 approx. here -

- Insert Table 4 approx. here -

Confirmatory factor analysis based on former proposed models

Altogether, four different factor models were tested by CFA; three five-factor models, as proposed by Andrea et al. (2007), Verheul et al. (2008) and Feenstra et al. (2011), and one four-factor model as proposed by Bastiaansen et al. (2013). CFA of our data revealed poor fit for all these models, not only for the community sample but also the patient samples (Table 5). Although the model proposed by Verheul et al. (2008) appeared to fit slightly better than the other models, differences were small. Poor fitness of models was mainly due to cross loadings of facets (facets loading on two or more factors) and residual covariance (systematic error indicating multidimensionality or sample specific bias).

– Insert Table 5 approx. here –

Exploratory factor analysis

An EFA of the facets was performed on all three samples due to the misfit of former proposed factor structures. The results are shown in Table 6. Within the community sample, two factors emerged with eigenvalues above 1.0, accounting for 66.7% of total variance. The first factor was very large, accounting for 58% of the variance. This factor represents positive attitudes towards self (self-respect), others (e.g., enduring relationships), and life in general (e.g., enjoyment). Therefore, this factor may be labeled "positive life attitude". The second factor is much smaller and appears to reflect a general regulatory factor, including facets like,

Emotion regulation, Effortful control, Aggression regulation, Frustration tolerance. Within the No-PD sample four factors emerged, accounting for 70.9% of total variance. The first two factors are highly similar to the factors from the community sample, with the most notable differences for Responsible industry and Trustworthiness, as well as Cooperation and Respect, which loaded on a separate factor. EFA of the PD sample revealed three factors, accounting for 63.0% of the variance. The first two factors are very similar to the first factors of the community sample; they had just “changed place”, i.e., the first factor became the second one and vice versa. As found in the No-PD sample Cooperation and Respect loaded on a separate factor.

– Insert Table 6 approx. here –

Norwegian community sample versus clinical samples

The comparison of mean facet scores of the two clinical samples (Table 7) and mean facet scores of the Norwegian community sample (Table 8) showed that the community sample had higher mean scores than the clinical samples on all facets ($p < .001$, Figure 1)., The psychiatric No-PD sample had higher mean scores than the PD sample on all facets ($p < .001$). Effect sizes for differences between the community sample and the NoPD sample ranged from .22 (Respect) to 1.37 (Self-respect), and effect sizes (Hedge’s g) for differences between the psychiatric samples were mostly within the moderate range (0.5 - 0.8). Effect sizes for differences between the community sample and the PD sample were much larger, ranging from .76 to 2.30.

– Insert Table 7 approx. here –

– Insert Figure 1 approx. here –

Facet scores across the Dutch and Norwegian community samples

Table 8 shows mean facet scores and standard deviations of for the Norwegian and Dutch community samples, as well as gender differences within the Norwegian sample. Scores from the Norwegian sample were somewhat higher than scores from the Dutch sample with effect sizes ranging from small (Responsible industry) to medium (Cooperation) and an average effect size of 0.30 (see also Figure 1). Mean levels of “Responsible industry” and “Aggression regulation” were equal in the two samples.

There were several statistically significant differences in facet scores across gender. For most facets, female participants had higher scores than male participants, with the greatest difference for Intimacy, although males had higher scores than females on the facets Self-respect and Frustration tolerance. It should be noted, however, that these effect sizes were small.

– Insert Table 8 approx. here –

Discussion

Since the initial development of SIPP, the present study is the first to analyze psychometric properties of SIPP in a community sample and two clinical samples, and compare mean score levels for SIPP facets. Based on the current study of SIPP facets, the questionnaire seems a promising instrument. In short, four main findings emerged: Firstly, the psychometric properties of the 16 SIPP facets were acceptable, although some have a potential for improvement. Secondly, the former proposed factor structures for the facets were not replicated. Thirdly, different factor structures of the facets were found across the three study samples, and some resemblance was found between them. Lastly, all facets of SIPP revealed good discriminative properties with respect to differentiating between the Norwegian community sample, the clinical sample without PD, and the clinical PD sample.

Properties of the facets

With some exceptions, the current study found acceptable psychometric properties of the SIPP facets. For all practical means, the validity of inferences drawn from the facet scores are most important. However, clinical usefulness of a facet can be good despite mediocre or even poor CFA model fit indices (according to conventional rules). It is beyond the scope of this study to fully explore this, but a refinement of the three facets Purposefulness, Responsible industry and Frustration tolerance should be given attention, since they do not seem to represent homogeneous constructs. In the present study and in all former studies in which Cronbach's alpha has been estimated, the lowest coefficients, (below .70) have been reported for the facet Respect. These results indicate that inferences from scores on this facet should be drawn with care, especially at an individual level.

Confirmatory factor analysis based on former proposed models

Our study neither supported a five-dimensional structure of the 16 SIPP facets, nor a four-factor solution as proposed by Bastiaansen et al. (2013). Initially, the item compositions of the facets are keyed in a way that addresses adaptive capacity. Respondents from non-clinical samples generally present themselves as having adaptive capacity. This makes the distribution of facet scores considerably skewed. In clinical samples the lack of adaptive capacities is more variable. This results in a more normal sample distribution of mean facet scores. Thus, in the framework of confirmatory factor analysis, a lack of metric invariance between clinical and non-clinical samples is probably introduced at the very beginning by the nature of the constructs.

Factor analysis of the same set of observed variables across different samples can reveal different patterns between the variables as well as different estimations of loadings associated with latent constructs (Hirschfeld, von Brachel & Thielsch, 2014). This means that

the regression weights for the computation of domain scores might not be generalizable across different normative or clinical samples or from normative to clinical samples. Accordingly, different combinations of facets have been suggested for operationalizing the SIPP domains (Bastiaansen et al. 2013; Feenstra et al., 2011; Verheul et al., 2008). The samples differed in these latter three studies: there were 159 patients in the study by Bastiaansen et al. (2013); Feenstra et al. (2010) had a clinical sample of 263 adolescents; and Verheul et al. (2008) based their analysis of internal consistency and factor structure on a clinical sample of 1208 PD patients. The proposed five higher order domains of SIPP are based on the algorithm of Andrea and co-workers (2007, pp. 37) which uses primary and secondary factor loadings as regression weights. As noted above, exact factor loadings tend to be sample-specific and difficult to replicate. Based on our current knowledge of the diversity of SIPP factor structure across different samples, we would strongly advise against using regression weights to force the facets into some theoretical model. There are no reasons to expect such a model to generalize. Moreover, the development of SIPP was initially focused on defining core components of maladaptive functioning, which resulted in the operationalization of 16 latent constructs. It seems that the relations between these 16 constructs reveal diverse patterns across different samples. However, this does not limit the clinical potentials of the facets or the instrument in general. A theoretically coherent model of higher order domains of the PD core components might still be confirmed by a subset of items from SIPP, i.e., operationalized by a short version of the questionnaire (Rossi, Debast, & van Alphen, 2016).

Exploratory factor analysis across the three samples

There were different numbers of factors across the current three samples. This is an indication of a lack of dimensional invariance of SIPP, which in turn implies a lack of configural invariance (that the same facets are not loading on the same dimensions across

samples). However, from the pattern matrix in Table 6, there are also indications of some resemblance across all three samples. Although varying in how much they account for the total variance, some facets are closely related regardless of the other facets. Those are: Emotion regulation and Effortful control; Self-respect, Enjoyment and Purposefulness; Responsible industry and Trustworthiness; Intimacy, Enduring relationships and Feeling recognized, and; Aggression regulation and Frustration tolerance. From this, the current analyses could support the existence of two main factors; a “positive attitude towards life” factor and a “general regulation” factor.

In the two clinical samples, Cooperation and Respect, and Stable self-image and Self-reflexive functioning seem to ‘agree’ on the choice of factors to regress on. Put another way, five different sets of facets are somewhat ‘unstable’ within the SIPP, as their relations to each other varies across different levels of severity. From the two clinical samples two additional facet-sets emerge. The clinical implications of this should be further investigated, but it might turn out to be unreasonable to claim factorial invariance among some clinical constructs at all levels of severity.

Facets mean scores across the three samples

All facets of SIPP differentiated, in a statistically significant way and with substantial effect sizes, between the three groups of patient with PD, patients without PD and the community sample. One exception might be the facet Respect. Although the No-PD sample had lower mean scores than the community sample on the facet Respect, the magnitude of the difference (effect size of 0.22) was so small that it is unclear to what extent it is sensitive to mental distress. Furthermore, the distribution of facet scores shows estimates of skewness in accordance with the nature of the clinical constructs. This means that clinical samples representing substantially decreased personality functioning reveal facet scores that are

normally distributed. In contrast, non-clinical samples reveal scores with negatively skewed distribution, and occasionally higher kurtosis. As a consequence, inferences made by individual deviation scores above the reference levels might be less valid than inferences from standardized scores below the normative reference.

Dutch and Norwegian community samples

The Dutch normative sample comprised 478 participants from the general population; 68% were women, and the mean age was 36 years ($SD = 12$). The educational level was categorized as low in 19.3%, intermediate in 49.1%, and high in 31.6% (Verheul et al., 2008). In the Norwegian community sample, gender proportions was somewhat more balanced (54%), mean age was somewhat higher than in the Dutch sample (44 years; $SD = 14$), and most participants had higher education levels (Table 1). We also have reasons to believe that participants younger than the age of 30 might be underrepresented and that participants older than the age of 50 might be overrepresented in the Norwegian community sample. Moreover, around 31% of the Norwegian population has a higher educational level, whereas in the current sample this proportion was 49%. This might indicate a bias in the Norwegian sample toward an older and more educated part of the population. On the other hand, such a reservation would only be valid if a relation between age and or educational level and scores had been shown on the SIPP facets. This should be further investigated, as it might have impact on the validity when making inferences from individual scores.

Group comparisons of SIPP facet scores should reflect true group differences, not contaminated by group-specific attributes that are unrelated to the constructs. Furthermore, testing theories concerning the nature of the maladaptive core components of SIPP relies on the level of factor invariance between the target cultures. As to the Dutch and Norwegian community samples, the adaptive capacities assessed by SIPP may mean something different

in the two countries, statements and questions may have different semantic meaning, and even the translation of the SIPP from one language to another, or from one culture to another may cause culturally biased effects on the equivalence of measurements. Therefore, such cross-cultural comparisons has several challenges (Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014; Harkness, 2006; Harzing, 2006).

Since only estimates of mean scores and standard deviations of the Dutch normative sample were available, the cross-national validity of the SIPP could not be investigated properly. Even if the two normative samples seemed quite similar, the differences between mean scores on the facets Stable self-image, Purposefulness, Enduring relationships and Respect would probably be statistically significant. Using the Dutch sample as a norm reference could result in higher individual deviation scores when used on Norwegian non-clinical groups, as well as indicating less individual deviations when used within clinical groups, especially when interpreting scores on Trustworthiness, Cooperation and Respect from more sub-clinical groups. This means that scores from these two populations are not fully interchangeable and that inferences with respect to increased or decreased scores cannot be fully generalized. Thus, specific norm scores for Norway should be established. On the other hand, such a conclusion would only be valid if the two normative samples are representative of their respective populations, and if cross-national validity is supported properly by adequate methods.

Limitations

To our knowledge, the three samples in this study are currently among the largest that have been used to analyze the psychometric properties of the SIPP. However, this study has some limitations that should be considered regarding the generalizability of the findings. First, the response rate of the Norwegian community sample was relatively low (ca. 20%).

Second, the mean age and educational level in the community sample was somewhat high compared to the Dutch normative sample (Verheul et al., 2008), and the level of education was higher than in the general Norwegian population (Statistics Norway, 2014). In addition, the patient sample comprised 73% women. Third, depression and anxiety disorders were highly represented in the patient samples, and borderline and avoidant PD were the most frequent PD diagnoses in the PD sample. Finally, inferences of clinical significance based on the observed differences on facet scores across the community and two clinical samples must be applied with some caution. The validity of such inferences might be influenced by sample specific bias and multidimensionality within the operationalization of some of the constructs (facets). Furthermore, measurement invariance of the facets across non-clinical and clinical samples should also be established.

Conclusion

In conclusion, SIPP seems to be a highly promising instrument. Due to some multidimensionality in the item compositions, the construct validity has a potential for improvement on some facets. Based on the current operationalizations of the 16 facets, there seems to be no consistent factor structure that is generalizable across different samples. Since the facets of SIPP revealed good discriminative properties with respect to differentiating between the current three samples, we suggest that the current Norwegian estimates can be used in Norway as reference scores for evaluations of severity at the facet level of the SIPP.

We suggest that future studies focus on the clinical utility and changeability of the facet scores in different groups of patients, measurement invariance of the facets, and possible interaction effects from age and educational level when estimating reference scores. As to the higher order domains of maladaptive personality functioning, we anticipate further research on a short version of SIPP.

Supplemental materials

SM1: Goodness-of-fit statistics from CFA, Internal consistency and distribution characteristics of each SIPP facet – Community sample.

SM2: Goodness-of-fit statistics from CFA, Internal consistency and distribution characteristics of each SIPP facet – Clinical PD sample.

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.) Washington DC.
- Andrea, H., Verheul, R., Berghout, C., Dolan, C., Vanderkroft, P., Busschbach, J., Bateman, A., & Fonagy, P. (2007). *Measuring the Core Components of Maladaptive Personality: Severity Indices of Personality Problems (SIPP-118) - The first technical report*. Viersprong Institute for Studies on Personality Disorders (VISPD) Medical Psychology and Psychotherapy, Netherlands.
- Arnevik, E., Wilberg, T., Andrea, H., Monsen, J. T., & Karterud, S. (2009). A cross-national validity study of the Severity Indices of Personality Problems (SIPP-118). *Personality and Mental Health*, 3, 41–55.
- Bales, D., van Beek, N., Smits, M., Willemsen, S., Busschbach, J. J. V., Verheul, R., & Andrea, H. (2012). Treatment outcome of 18-month, day hospital Mentalization-Based Treatment (MBT) in patients with severe borderline personality disorder in the Netherlands. *Journal of Personality Disorders*, 26, 568–582.
- Bales, D. L., Timman, R., Andrea, H., Busschbach, J. J. V., Verheul, R., & Kamphuis, J. H. (2014). Effectiveness of Day Hospital Mentalization-Based Treatment for Patients with Severe Borderline Personality Disorder: A Matched Control Study. *Clinical Psychology and Psychotherapy*, DOI: 10.1002/cpp.1914.
- Bastiaansen, L., De Fruyt, F., Rossi, G., Schotte, C., & Hofmans, J. (2013). Personality disorder dysfunction versus traits: structural and conceptual issues. *Personality Disorders: Theory, Research, and Treatment*, 4, 293–303.

- Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical Considerations for Using Exploratory Factor Analysis in Educational Research. *Practical Assessment, Research & Evaluation*, 18, 1–13.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238–246.
- Berghuis, H., Kamphuis, J. H., Verheul, R., Larstone, R., & Livesley, J. (2013). The General Assessment of Personality Disorder (GAPD) as an instrument for assessing the core features of personality disorders. *Clinical Psychology and Psychotherapy*, 20, 544–557.
- Berghuis, H., Kamphuis, J. H., & Verheul, R. (2014). Specific personality traits and general personality dysfunction as predictors of the presence and severity of personality disorders in a clinical sample. *Journal of Personality Assessment*, 96, 410–416.
- Berghuis, H., Kamphuis, J. H., & Verheul, R. (2012). Core features of personality disorder: Differentiating general personality dysfunctioning from personality traits. *Journal of Personality Disorders*, 26, 704–716.
- Byrne, B. M. (1998), *Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.

- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement Equivalence in Cross-National Research. *Annual Review of Sociology*, 40, 55–75.
- Diamantopoulos, A., & Siguaw, J. A. (2000). *Introducing LISREL*. London: Sage Publications.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Press.
- 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, 646–655.
- Feenstra, D. J., Hutsebaut, J., Verheul, R., & van Limbeek, J. (2014). Changes in the identity integration of adolescents in treatment for personality disorders. *Journal of Personality Disorders*, 28, 101–112.
- First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *The Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*. Washington, DC, American Psychiatric Press.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (1995). The Structured Clinical Interview for DSM-III-R Personality Disorders (SCID-II): Part I. Description. *Journal of Personality disorders*, 9, 83–91.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment*, 7, 286–299.
- Graham, J. M. (2006). Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability - What They Are and How to Use Them. *Educational and Psychological Measurement*, 66, 930–944.

- Gregorich, S. E. (2006). Do Self-Report Instruments Allow Meaningful Comparisons Across Diverse Population Groups? Testing Measurement Invariance Using the Confirmatory Factor Analysis Framework. *Medical Care*, 44, 78–94.
- Gudjonsson, G. H., Sigurdsson, J. F., Guðmundsdóttir, H. B., Sigurjónsdóttir, S., & Smari, J. (2010). The relationship between ADHD symptoms in college students and core components of maladaptive personality. *Personality and Individual Differences*, 48, 601–606.
- Harkness, J. A., (2006). ZUMA-Nachrichten Spezial Volume 12: *Conducting Cross-National and Cross-Cultural Surveys*. Mannheim, Germany: ZUMA.
- Harzing, A. W. (2006). Response Styles in Cross-national Survey Research. A 26-country Study. *International Journal of Cross Cultural Management*, 6, 243–266.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6, 107–128.
- Hirschfeld, G., von Brachel, R., & Thielsch, M. T. (2014). Selecting items for Big Five questionnaires: At what sample size do factor loadings stabilize? *Journal of Research in Personality*, 53, 54–63.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- IBM (2010). *IBM SPSS Statistics for Windows*, Version 19.0. Armonk, NY: IBM Corp.

- Joop, J. Hox, J. J., Maas, J. M. M., & Brinkhuis, M. J. S. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, 64, 157–170.
- Karterud, S., Pedersen, G., Friis, S., Urnes, Ø., Irion, T., Brabrand, J., Leirvåg, H., & Falkum, L. R. (1998). The Norwegian network of psychotherapeutic day hospitals. *Therapeutic Communities*, 1, 15–28.
- Lien, I. E., & Arnevik, E. A. (2016). Assessment of personality problems among patients with substance use disorders. *Nordic Studies on Alcohol and Drugs*, 33, 399–414.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods*, 1, 130–149.
- Morey, L. C., Berghuis, H., Bender, D. S., Verheul, R., Krueger, R. F., & Skodol, A. E. (2011). Toward a model for assessing level of personality functioning in DSM-5, part II: Empirical articulation of a core dimension of personality pathology. *Journal of Personality Assessment*, 93, 347–353.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén, 1998–2012.
- Pedersen, G., & Karterud, S. (2010). Using measures from the SCL-90-R to screen for personality disorders. *Personality and Mental Health*, 4, 121–132.
- 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, 277–287.

- Rossi, G., Debast, I., & van Alphen, S. P. J. (2016). Measuring personality functioning in older adults: construct validity of the Severity Indices of Personality Functioning – Short Form (SIPP-SF). *Aging and Mental Health*. DOI: 10.1080/13607863.2016.1154012.
- Statistics Norway (2014). *Population's level of education, 1 October 2014*.
<http://www.ssb.no/en/utdanning/statistikker/utniv/aar>
- Sheehan, D. V., & Lecrubier, Y. (1994). *Mini International Neuropsychiatric Interview (M.I.N.I.)*. Tampa, Florida / Paris, France: University of South Florida Institute for Research in Psychiatry / INSERM-Hôpital de la Salpêtrière.
- Steiger, J. H. (1990), Structural model evaluation and modification. *Multivariate Behavioral Research*, 25, 214–212.
- Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modelling. *Personality and Individual Differences*, 42, 893–98.
- Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55.
- Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1–10.
- 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, 23–34.

Yuan, K.H & Bentler, P.M. (2000). Three Likelihood-Based Methods For Mean and Covariance Structure Analysis With Nonnormal Missing Data, *Sociological Methodology*, 30, 165–200.

Appendix A. Descriptions of the SIPP-118 domains and facets

Domain: Self-control

The capacity to tolerate and control emotions, to focus concentration and to direct impulses.

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”

Domain: Identity integration

Mature identity: the capacity to experience self-worth, purposefulness, enjoyment, a sense of self-continuity, and an understanding of internal and external events.

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-reflexive 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”

Domain: Responsibility

The ability to set realistic goals and to achieve these goals in line with the values and norms of social collaboration.

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”

Domain: Relational capacities

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

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”

Domain: Social concordance

The ability to value someone’s individual needs and personal identity, to withhold aggressive impulses towards others, to cope with disappointments, and to work constructively together with others.

Aggression regulation – The ability to withhold aggressive impulses towards 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 someone’s individual needs and personal identity. I.e. “I strongly believe that everybody is entitled of his own opinion”

From Appendix 3, pp. 33-34, in Andrea et al., (2007). Used with permission. Example items added.

Table 1. <i>Gender, Age, and Education – Community sample</i>			
	Total (<i>n</i> = 940) <i>n</i> (%)	Male (<i>n</i> = 432) <i>n</i> (%)	Female (<i>n</i> = 508) <i>n</i> (%)
Age 18–25 years	103 (11)	34 (8)	69 (14)
Age 26–30 years	109 (12)	43 (10)	66 (13)
Age 31–40 years	171 (18)	75 (17)	96 (19)
Age 41–50 years	224 (24)	101 (23)	123 (24)
Age > 50 years	333 (35)	179 (41)	154 (30)
Junior high school ^a	78 (8)	34 (8)	44 (9)
High school ^b	393 (42)	202 (47)	191 (38)
College / university ^c	464 (49)	194 (45)	270 (53)
<i>Note.</i> The three educational levels are comparable to the Dutch educational levels as described by Verheul et al. (2007): ^a primary school/lower vocational education; ^b secondary school/intermediate vocational education); ^c upper vocational education/university. Percentages are rounded.			

Table 2. <i>Gender, age, and prevalence of PD – Clinical sample</i>			
No PD Sample:	Total (n = 413) n (%)	Male (n = 114) n (%)	Female (n = 299) n (%)
<i>Age</i>			
Age 18–25 years	73 (18)	13 (11)	60 (20)
Age 26–30 years	66 (16)	21 (18)	45 (15)
Age 31–40 years	137 (33)	41 (36)	96 (32)
Age 41–50 years	110 (27)	34 (30)	76 (25)
Age > 50 years	27 (7)	5 (4)	22 (7)
PD Sample:	Total (n = 949) n (%)	Male (n = 252) n (%)	Female (n = 697) n (%)
<i>Age:</i>			
Age 18–25 years	280 (30)	40 (16)	240 (34)
Age 26–30 years	230 (24)	57 (23)	173 (25)
Age 31–40 years	261 (28)	91 (36)	170 (24)
Age 41–50 years	134 (14)	46 (18)	88 (13)
Age > 50 years	44 (5)	18 (7)	26 (4)
<i>Frequency of PD's:</i>			
Schizoid	6 (0)	4 (1)	2 (0)
Schizotypal	9 (1)	2 (1)	7 (1)
Paranoid	109 (8)	36 (10)	73 (7)
Borderline	342 (25)	48 (13)	294 (30)
Antisocial	16 (1)	7 (2)	9 (1)
Narcissistic	15 (1)	9 (3)	6 (1)
Histrionic	3 (0)	1 (0)	2 (0)
Avoidant	371 (27)	120 (33)	251 (25)
Dependent	50 (4)	9 (3)	41 (4)
Obsessive-compulsive	69 (5)	25 (7)	44 (4)
PD NOS	236 (17)	65 (18)	171 (17)
No PD	413 (30)	114 (31)	299 (30)
<i>Note.</i> PD NOS = Personality disorder not otherwise specified. Presentences are rounded.			

Table 3.
Goodness-of-fit statistics from CFA and Internal consistency of each SIPP facet

	Community sample				Clinical PD sample		
	χ^2 (df)	RMSEA (90% C.I.)	Cronbach's alpha (95% CI; MIIC)		χ^2 (df)	RMSEA (90% C.I.)	Cronbach's alpha (95% CI; MIIC)
<i>Facets</i>							
Emotion regulation	89.86 (14)	0.076 (0.061–0.091)	.82 (.80–.84; .40)		79.49 (14)	0.070 (0.056–0.086)	.81 (.79-.83; .38)
Effortful control	30.93 (14)	0.036 (0.019–0.053)	.78 (.76–.81; .35)		99.84 (14)	0.080 (0.066–0.096)	.81 (.79-.83; .37)
Self-respect	101.37 (20)	0.066 (0.053–0.079)	.84 (.82–.86; .41)		130.61 (20)	0.076 (0.064–0.089)	.79 (.77-.81; .33)
Stable self-image	66.53 (14)	0.063 (0.048–0.079)	.79 (.77–.81; .35)		107.32 (14)	0.084 (0.069–0.099)	.75 (.72-.77; .30)
Self-reflexive functioning	41.28 (14)	0.046 (0.030–0.062)	.79 (.76–.81; .35)		96.26 (14)	0.079 (0.064–0.094)	.73 (.70-.76; .28)
Enjoyment	46.20 (14)	0.049 (0.034–0.066)	.80 (.78–.82; .37)		73.05 (14)	0.067 (0.052–0.082)	.74 (.71-.77; .30)
Purposefulness	102.96 (14)	0.082 (0.068–0.097)	.75 (.73–.77; .33)		151.02 (14)	0.102 (0.087–0.116)	.72 (.69-.74; .27)
Responsible industry	81.14(14)	0.071 (0.057–0.087)	.68 (.65–.71; .25)		129.44 (14)	0.093 (0.079–0.108)	.73 (.70-.76; .28)
Trustworthiness	46.70 (20)	0.038 (0.024–0.052)	.69 (.66–.72; .22)		47.82 (20)	0.038 (0.024–0.052)	.75 (.73-.78; .28)
Intimacy	69.07 (14)	0.065 (0.050–0.080)	.80 (.78–.82; .37)		207.11 (14)	0.121 (0.106–0.135)	.79 (.77-.81; .35)
Enduring relationships	62.87 (14)	0.061 (0.046–0.077)	.80 (.78–.82; .36)		200.26 (14)	0.118 (0.104–0.133)	.70 (.67-.73; .25)
Feeling recognized	87.07 (20)	0.060 (0.047–0.073)	.81 (.80–.83; .36)		259.72 (14)	0.112 (0.100–0.125)	.77 (.74-.79; .29)
Aggression regulation	58.45 (20)	0.045 (0.032–0.059)	.80 (.78–.82; .35)		107.41 (20)	0.068 (0.056–0.081)	.85 (.84-.87; .42)
Frustration tolerance	94.05 (20)	0.063 (0.050–0.076)	.78 (.76–.88; .30)		129.49 (20)	0.076 (0.064–0.089)	.72 (.69-.74; .25)
Cooperation	98.34 (20)	0.065 (0.052–0.078)	.76 (.75–.80; .32)		192.97 (20)	0.095 (0.083–0.108)	.76 (.73-.78; .28)
Respect	40.15 (14)	0.045 (0.029–0.061)	.67 (.63–.70; .22)		28.85 (14)	0.033 (0.016–0.051)	.70 (.67-.73; .25)

Note. RMSEA, root mean square error of approximation; CI, confidence interval. Chi-square statistics: p values < .0001. For full Goodness-of-fit statistics from CFA, Internal consistency and distribution characteristics (as reported in Table 4), see supplementary material SM1 and SM2.

Goodness-of-fit statistics from CFA, Internal consistency and distribution characteristics of each SIPP facet – Clinical No PD sample

Note. RMSEA, root mean square error of approximation; CI, confidence interval; CFI, comparative fit index; TLI, Tucker-Lewis index; SRMR, standardized root mean square residual. Chi-square statistics: p values < .0001; MIIC = Mean inter-item correlation within scale. SE Skewness = 0.120; SE Kurtosis = 0.240.

Table 5. <i>Goodness-of-fit statistics from CFA of four suggested factor models</i>					
Community sample					
	X^2 (df)	RMSEA (90% C.I.)	CFI	TLI	SRMR
<i>Models</i>					
Andrea	1034.34 (94)	0.103 (0.097–0.109)	0.893	0.864	0.052
Verheul	924.26 (94)	0.097 (0.091–0.103)	0.906	0.880	0.052
Feenstra	978.94 (94)	0.100 (0.094–0.106)	0.899	0.872	0.053
Bastiaansen	1153.70 (98)	0.107 (0.101–0.113)	0.880	0.853	0.059
Clinical No PD sample					
Andrea	598.18 (94)	0.114 (0.105–0.123)	0.852	0.811	0.070
Verheul	570.47 (94)	0.111 (0.102–0.120)	0.860	0.821	0.070
Feenstra	598.69 (94)	0.114 (0.105–0.123)	0.852	0.811	0.072
Bastiaansen	626.47 (98)	0.114 (0.106–0.123)	0.845	0.810	0.079
Clinical PD sample					
Andrea	1589.35 (94)	0.129 (0.124–0.135)	0.795	0.738	0.097
Verheul	1486.91 (94)	0.125 (0.01190.131)	0.809	0.756	0.096
Feenstra	1591.24 (94)	0.130 (0.124–0.135)	0.794	0.738	0.101
Bastiaansen	1644.76 (98)	0.129 (0.124–0.134)	0.788	0.740	0.102
<i>Note.</i> RMSEA, root mean square error of approximation; CI, confidence interval; CFI, comparative fit index; TLI, Tucker-Lewis index; SRMR, standardized root mean square residual. Chi-square statistics: p values < .0001. Five-factor models: Andrea et al., (2007); Verheul et al., (2008); Feenstra et al., (2011). Four-factor model: Bastiaansen et al., (2013).					

Pattern Matrix from Exploratory factor analysis (PAF) of the SIPP facets

Note. Rotated loadings (regression weights) at or above 0.40 in bold.

Table 7.
Mean scores on the SIPP facets – Clinical samples

[illegible]

Table 8.
Mean scores and gender differences on the SIPP facets – Community sample

	Dutch sample ^a		Norwegian sample			Hedges' g^b	Norwegian sample by gender					
	(n = 478)		(n = 941)				Males (n = 432)		Females (n = 508)		Gender diff. ^c	Hedges' g^d
	Mean	SD	Mean	SD	S _E ^a		Mean	SD	Mean	SD		
Facets												
Emotion regulation	3.30	0.61	3.46	0.54	0.23	0.28	3.46	0.50	3.46	0.57	<i>ns</i>	0.00
Effortful control	3.16	0.56	3.31	0.58	0.27	0.26	3.32	0.57	3.30	0.58	<i>ns</i>	0.04
Self-respect	3.30	0.59	3.46	0.55	0.22	0.28	3.52	0.49	3.41	0.58	<i>p</i> < .01	0.20
Stable self-image	3.24	0.67	3.48	0.52	0.24	0.41	3.47	0.52	3.48	0.51	<i>ns</i>	0.02
Self-reflexive functioning	3.20	0.56	3.40	0.54	0.25	0.37	3.39	0.51	3.41	0.56	<i>ns</i>	0.04
Enjoyment	3.34	0.62	3.44	0.54	0.24	0.18	3.42	0.55	3.45	0.53	<i>ns</i>	0.06
Purposefulness	3.34	0.49	3.53	0.47	0.24	0.40	3.55	0.46	3.52	0.47	<i>ns</i>	0.06
Responsible industry	3.44	0.50	3.42	0.48	0.27	0.04	3.36	0.49	3.48	0.46	<i>p</i> < .001	0.25
Trustworthiness	3.49	0.42	3.60	0.39	0.22	0.27	3.53	0.40	3.66	0.37	<i>p</i> < .001	0.34
Intimacy	3.17	0.60	3.30	0.58	0.26	0.22	3.17	0.58	3.40	0.56	<i>p</i> < .001	0.40
Enduring relationships	3.31	0.58	3.52	0.56	0.25	0.37	3.49	0.57	3.55	0.54	<i>ns</i>	0.11
Feeling recognized	3.23	0.56	3.39	0.56	0.24	0.29	3.36	0.56	3.43	0.55	<i>ns</i>	0.13
Aggression regulation	3.66	0.45	3.63	0.45	0.20	0.07	3.60	0.48	3.66	0.43	<i>p</i> < .05	0.13
Frustration tolerance	2.96	0.56	3.13	0.52	0.24	0.32	3.17	0.51	3.10	0.52	<i>p</i> < .05	0.14
Cooperation	3.28	0.51	3.54	0.44	0.22	0.56	3.47	0.47	3.60	0.40	<i>p</i> < .001	0.30
Respect	3.34	0.45	3.53	0.41	0.24	0.45	3.45	0.45	3.59	0.37	<i>p</i> < .001	0.34

Note. *) From Table 3, pp. 26, in Andrea et al., (2007), used with permission a) Standard error of measurement. b) Hedges' g for sample differences, Dutch and Norwegian. c) Independent samples t-test (two-sided), and d) Hedges' g for gender differences, Norwegian sample. – Also note that higher mean scores indicate better functioning.

Figure 1.

Mean facet levels across the Dutch and Norwegian community sample and the two clinical samples