

## **Psychometric properties of the Spanish adaptation of the Externalizing Spectrum Inventory - Brief Form (ESI-BF)**

### **Abstract**

The Externalizing Spectrum Inventory - Brief Form (ESI-BF) measures tendencies towards disinhibition, lack of control, aggression and substance use. This study adapts the ESI-BF to the Spanish population and assesses its psychometric properties. The study included 742 community adults obtained by stratified random sampling with proportional allocation according to gender, age and geographical area of the Spanish territory and a clinical sample consisting of 333 patients. All participants completed the Personality Inventory for DSM-5 and the Alcohol Substance Dependence Severity Scale, in addition to the Spanish version of the ESI-BF. Reliability was quantified using McDonald's omega and Cronbach's alpha reliability coefficients. Validity evidences were studied applying confirmatory factor analysis and correlations. Results indicated adequate reliability of scores on the ESI-BF's general factors and most of its facets. Regarding internal structure, and in line with previous studies, both symmetric and S-1 hierarchical two-subfactor (bifactor) emerged as the best-fitting models. Considering both criticisms of symmetric models and parsimony, the S-1 bifactor model, which showed configural invariance across gender and samples, was retained. Validity evidence based on the relationship with other measures of personality and alcohol consumption show correlations values theoretically expected in both clinical and community samples. Findings suggest that the Spanish adaptation of the ESI-BF shows functional near-equivalence to the original version. Its effective psychometric properties make it useful instrument for further research related to the externalizing spectrum.

**Public significance statement:** The ESI-BF measure a broad set of disinhibitional, aggressive and substance use tendencies within a single instrument, reducing respondent burden. The present study extends the utility of ESI-BF, into Spanish-speaking population, providing evidence on the psychometric properties of a linguistic and cultural adaptation. Results show that scores provided by ESI-BF-S are consistent and correlate with other measures of personality and alcohol consumption, as theoretically expected.

**Keywords:** Externalizing; ESI-BF; psychometric properties; adaptation; Spanish.

### Background

In recent decades, the observed co-occurrence among psychiatric disorders, the heterogeneity of patients diagnosed using traditional diagnostic categories (Zimmerman et al., 2015), and the lack of progress in relating diagnostic categories with neurobiological variables (Allen et al., 2020) has led to the development of quantitative-hierarchical models that integrate groups of diagnostic conditions along dimensional lines (Hengartner & Lehmann, 2017; Kim & Eaton, 2015; Kotov et al., 2017). A notable example is the *externalizing spectrum* (ES) model, which encompasses substance use disorders and antisocial problems (e.g., conduct disorders, adult antisocial behavior) together with personality traits such as impulsivity and aggression (Kotov et al., 2017; Krueger et al., 2007). The ES model provides an alternative to the traditional view of disorders of these types as separate entities (Forbes et al., 2021; Pasion & Barbosa, 2019), characterizing them instead as alternative clinical expressions of broad dispositional factors. This model has received support from various lines of empirical evidence (see review by Krueger et al., 2021). Regarding the structure of externalizing problems, Soe-Agnie et al. (2018) conducted a systematic review of studies examining

the factor structure of different forms of externalizing psychopathology (i.e., substance use disorders and antisocial behaviors) in different samples. In spite of the different analytical approaches used in the 38 studies they reviewed, these authors found that 35 of them reported evidence for superordinate factors of externalizing disorders. In addition, some authors have presented evidence for genetic and neurobiological elements of externalizing disorders that distinguish them from other disorders (Patrick et al., 2013; Venables, 2017).

Externalizing spectrum disorders have traditionally been assessed using instruments that characterize specific disorders according to the ICD and DSM classification systems. Instruments of this sort yield categorical diagnoses of disorders by establishing their presence or absence based on symptom cutoffs. Another set of instruments evaluates specific traits or symptom facets linked to disorders included in the externalizing spectrum. Typically, such instruments allow for a dimensional assessment of these traits and symptom facets. However, these instruments differ in assessment time frames, response formats, and underlying conceptual models, which hinders the interpretability of scores for clinical and research purposes.

One instrument specifically designed to comprehensively assess the externalizing psychopathology domain is the Externalizing Spectrum Inventory (ESI; Krueger et al., 2007). This instrument aligns with the Hierarchical Taxonomy of Psychopathology (HiTOP; Kotov et al., 2017; Kotov et al., 2021) in that it assesses externalizing proneness in terms of narrow trait and symptom facets organizing around broad dimensions (factors). Scores on each of the ESI's 23 content scales allow for the evaluation of various facets of the externalizing spectrum using the same instrument and the same scoring system. Studies examining the psychometric properties of these content scales have revealed internal consistency reliabilities above .8 for all scales

across university student and prisoner samples (Krueger et al., 2007; Patrick et al., 2013). Regarding the factor structure of the ESI's 23 facet scales, prior research has reported the best fit for a hierarchical (bifactor) model specifying a broad externalizing factor (reflecting general disinhibitory propensities), together with two specific factors onto which scales pertaining to callous aggressiveness and substance use load (Krueger et al., 2007; Patrick et al., 2013). Recent simulation work, however, points out that comparative fit indices tend to favor bifactor models (Bonifay et al., 2017; Greene et al., 2019) compared to correlated factor models. Additionally, inconsistent results (e.g., irregular loadings or negative factor variances) have been reported for conventional (symmetric) bifactor models (Eid et al., 2017), precluding clear interpretation of the general externalizing factor. Several authors (Burke & Johnson, 2020; Eid et al., 2020; Junghänel et al., 2020) have suggested that S-1 bifactor models – with all indicators loading onto the general factor but only a subset of them loading on one or another specific factor – offer a preferred approach to bifactor modelling that avoids fit bias and allows for clearer interpretation of both the general and specific factors.

A limitation of the ESI for use in multi-measure research protocols and clinical settings is its length (i.e., 415 items). In order to reduce the time of administration, abridged versions of the ESI have been developed. Two of these, consisting of 100 items (Hall et al., 2007) and 159 items (Venables & Patrick, 2012), were developed to index the ESI's factors and have proven useful for clarifying their nomological networks (see, e.g., Bernat et al., 2011; Blonigen et al., 2011; Nelson et al., 2011; Venables & Patrick, 2012). However, a short version was needed to provide coverage of the ESI's specific facets along with its broad factors. Patrick et al. (2013) addressed this need by developing a 160-item brief form (ESI-BF) that allows for effective assessment at both the facet and factor level. This version has shown good internal consistency at

the facet scale level (Patrick et al., 2013; Soe-Agnie et al., 2016, 2021) and a bifactor structure that parallels the original version (Patrick et al., 2013).

A number of studies have employed the item-based factor scales of this brief-form version – labeled Disinhibition (ESI<sub>DIS</sub>), Callous Aggression (ESI<sub>AGG</sub>), and Substance Abuse (ESI<sub>SUB</sub>) – and provided validity evidence in terms of relations between scores on its scales and criterion measures of various types (Byrne et al., 2019; Byrne & Worthy, 2019; Rodriguez-Seijas et al., 2017; Venables et al., 2018). Some validity studies have examined relations with normative and pathological personality traits. For example, Poy et al. (2014) reported evidence for robust relations of ESI<sub>DIS</sub> and ESI<sub>AGG</sub> with traits of the Five Factor model of personality (see also Soe-Agnie et al., 2021), and Strickland et al. (2013) found strong associations between the ESI<sub>DIS</sub> and ESI<sub>AGG</sub> scales and the disinhibition and antagonism domains of the PID-5 (see also Wygant et al., 2016).

This body of evidence attests to the growing interest in measuring externalizing spectrum problems/traits, as well as the potential utility of the ESI brief form for pursuing research in this area. At present, the ESI-BF is only available in English (Patrick et al., 2013) and Dutch versions (Soe-Agnie et al., 2016), and studies using its factor scales have mostly used samples of specific population groups (e.g., university students), or community samples unrepresentative of the general population due to use of non-probabilistic sampling procedures. Therefore, the general objective of the present work was to provide psychometric evidence for the Spanish adaptation of the ESI-BF in a community sample selected to be representative of the general Spanish population, as well in a clinical sample.

The specific research aims of the current study were as follows: 1) to assess the reliability of scores for the specific scales (facets) of the Spanish-adapted ESI-BF and

its item-based broad factor scales; 2) to provide validity evidence based on the internal structure of the Spanish ESI-BF scores, through examination of the fit of alternative confirmatory factor models; and 3) to evaluate evidence based on relations to other variables through the associations of its scales and factors with pertinent criterion measures in domains of personality pathology and substance (alcohol) use, and its ability to discriminate between clinical and non-clinical (general community) samples and between genders.

Based on prior published work (Patrick et al., 2013; Soe-Agnie et al., 2016; 2021), internal consistency coefficient values of .70 or higher were expected for the inventory's 23 facet scales and its 3 broad factor scales. With regard to internal structure of the Spanish ESI-BF's facet scales, four scales that have shown robust loadings on the general factor but negligible loadings on either subfactor in prior research (Krueger et al., 2007; Patrick et al., 2013) – Problematic Impulsivity, Irresponsibility, Planful Control, Dependability – were selected to operate as “pure” reference indicators of the general factor in a S-1 bifactor model. A better fit was predicted for this S-1 bifactor model (Burke & Johnson, 2020) compared to one-factor and higher-order models, along with greater parsimony compared to previously reported symmetric bifactor factor models (Patrick et al., 2013; Soe-Agnie et al., 2016; 2021). Finally, scores on the ESI<sub>DIS</sub> and ESI<sub>AGG</sub> factor scales were expected to correlate positively with the disinhibition and antagonism domains of the PID-5 (Poy et al., 2014; Soe-Agnie et al., 2021; Strickland et al., 2013), whereas ESI<sub>SUB</sub> scale scores were expected to correlate positively with measures of alcohol use as well as with the disinhibition domain of the PID-5 (Rodriguez-Seijas, 2017; Soe-Agnie et al., 2021; Venables & Patrick, 2012).

## Method

### Participants

Two participant samples were employed in the current study, a non-clinical (community) sample and a clinical (mental health patient) sample. The sampling frame for community participants, which comprised 155,000 adults representative of the overall Spanish population, was recruited by an online access panel provider certified with the International Organization for Standardization (ISO) 26362 quality standards. The community sample was selected using a stratified random sampling approach with proportional allocation according to gender, age (range: 18 - 75 years), and geographical region of the Spanish territory. Only participants invited to participate who fully completed all measures were retained, with less than 3% of the invited participants providing incomplete answers. The final community sample consisted of 742 individuals (44.6% of them women), ranging in age from 18 to 75 years ( $M = 45.77$ ;  $SD = 14.83$ ). The age composition of study participants was: 13%, aged 18 to 24 years; 18%, aged 25-34; 21%, aged 35-44; 21%, aged 45-54; 27%, over 55 years old. Regarding representation of geographical regions of Spain, the sample included: 9% each from Northwest (Galicia, Asturias) and North center (Vasque country, Cantabria); 12% Northeast (Aragon, Catalonia, La Rioja); 10%, Center (Castilla La Mancha, Castilla León, Madrid, Extremadura); 15%, East (Valencia, Murcia, Balearic islands); 18%, South (Andalusia); 13%, Madrid metropolitan area; 10%, Barcelona metropolitan area; 4%, Canary islands of Ceuta and Melilla. This sample composition mirrored the geographic distribution of the Spanish population. At the time of the study, 61.3% of participants were gainfully employed, and 0.8% possessed no education, 2.6% had primary education, 9.2% had compulsory secondary education, 41.1% had post-compulsory secondary education, and 46.4% had completed a full bachelor's degree.

With regard to marital status, 27.2% were single, 62% were married or in a relationship, 9.6% were separated or divorced, and 1.2% were widowed.

The clinical sample for the current study consisted of 333 patients from public (65.8% of patients) and private (34.2%) mental health services in the Spanish province of Huelva, recruited via clinicians through use of a non-probabilistic procedure. During the data collection period, all patients initiating treatment in these participating mental health services were invited to participate. Those meeting either of the following criteria were excluded: failure to sign the informed consent form, or history of a diagnosis for a medical or psychological disorder that could invalidate their test results. The most frequent diagnostic categories among patients were depressive disorders (35.87%), anxiety disorders (31.31%), and trauma and stressor-related disorders (23.40%). Participants in this patient sample included 63.1% women and ranged in age from 18 to 80 years ( $M = 38.80$ ;  $SD = 14.38$ ); 45.3% were gainfully employed at the time of the study, and 3.6% had no education, 10.5% had primary education, 30.9% had compulsory secondary education, 37.2% had post-compulsory secondary education, and 17.7% had completed higher education. With regard to marital status, 48.6% were single, 33% were married or in a relationship, 14.7% were separated or divorced, and 3.3% were widowed.

## Measures

### *Externalizing Spectrum Inventory Brief Form – Spanish Version (ESI-BF-S).*

The ESI-BF is composed of 160 items with four response options, ranging from 0 ("disagree") to 3 ("strongly agree"); procedures for adapting this inventory into Spanish are described in the next section below. The items are grouped into 23 facet scales: Problematic Impulsivity, Irresponsibility, Theft, Fraud, Impatient Urgency, Planful Control, Dependability, Alienation, Boredom Proneness, Blame Externalization,

Honesty, Rebelliousness, Physical Aggression, Destructive Aggression, Relational Aggression, Empathy, Excitement Seeking, Marijuana Use, Marijuana Problems, Drug Use, Drug Problems, Drug Problems, Alcohol Use and Alcohol Problems. Sum scores on three item-based scales – labeled ESI Disinhibition (ESI<sub>DIS</sub>), ESI Callous Aggression (ESI<sub>AGG</sub>) and ESI Substance Abuse (ESI<sub>SUS</sub>; see Patrick et al., 2013) – were computed to index the three higher order factors of the ESI.

*Personality Inventory for DSM-5 Short Form (PID-5).* The Spanish version of the PID-5 (Diaz-Batanero et al., 2019) was used. It is composed of 100 items, answered using a Likert response format, with scores ranging between 0-3 (from “strongly disagree” to “strongly agree”). The items of this instrument are grouped into scales for assessing the 25 facet traits of the alternative model for personality disorders documented in Section III of the DSM-5, which are grouped in turn into the following five broad trait domains: negative affect, detachment, antagonism, disinhibition, and psychoticism. Higher scores on each PID-5 scale are indicative of a greater presence of the trait evaluated. In the present study, Cronbach's alpha internal consistency coefficients for the broad trait scales ranged from .87 (antagonism) to .90 (detachment and psychoticism).

*DSM-5 Substance Dependence Severity Scale (SDSS) – Alcohol Subscale* (Gonzalez-Saíz et al, 2014). The SDSS evaluates the severity of dependence on alcohol and other substances during the month prior to its administration using DSM-5 criteria (Dacosta-Sanchez et al., 2019). The subscale of the SDSS that assesses alcohol dependence severity was used in the current study. This scale is composed of 17 items with five response options, ranging from 1 ("never") to 5 ("almost always"). In the present study, assessment of internal consistency for the SDSS-Alcohol scale yielded a Cronbach's alpha value of .94.

**Procedure**

Before adapting the ESI-BF into the Spanish language, the research team requested authorization from the authors of the original version and made an assessment of the preconditions established by the International Test Commission (2018). The adaptation process was conducted with consideration given to the linguistic, psychological, and cultural aspects of both the culture of origin and the Spanish culture.

Initially, three members of the research team independently translated the items, discussing any discrepancies found. The consensus wordings for items were then translated back into English by a professional translation service, staffed by a native English speaker with a PhD in psychology. The original English-worded items and the proposed translations were compared by the research team and potential discrepancies in meaning and adequacy to Spanish context were discussed with the translation team. The final 160-item version was then revised for grammatical errors, typographical errors, or lexical inconsistencies.

The resultant Spanish-language inventory (ESI-BF-S) and the other two study questionnaires were administered online to the community sample. Clinic patients completed the tests in rooms set up in the centers where they were recruited. All participants gave their consent to participate in the study after being informed of its anonymous and voluntary nature. The study was approved by the Bioethics Committee of the Province of Huelva (Junta de Andalucía, Spain) (No. PI 040/18). Informed consent was obtained from all individual participants involved in the study.

**Statistical Analysis**

The data set (Diaz-Batanero, 2022) did not contain missing scores for the community sample, as the online survey company retained test data only for participants

who completed the questionnaire measures in full. Similar administration instructions for the clinical sample, who completed the questionnaires in person, resulted in missing data for less than 1% of scores. Descriptive statistics ( $M$  and  $SD$ ) for scores on the 23 scales of the ESI-BF and its three item-based factor scales were computed. Reliability coefficients consisting of McDonald's omega and Cronbach's alpha were also computed for the inventory's facet and item-based factor scales.

Confirmatory factor analyses (CFAs) were conducted to evaluate the fit of the following scale-level models specified for the original English-language version of the ESI-BF (Patrick et al., 2013): *Model 1* (one general factor), in which the 23 facet scales were specified as loading onto a single factor (general externalizing); *Model 2* (two higher order subfactors), in which two first-order factors were specified – callous-aggression  $ESI_{AGG}$  and substance abuse  $ESI_{SUS}$  – that loaded in turn onto a second-order factor; *Model 3* (hierarchical two-subfactor), in which a general factor was specified onto which all facet scales loaded (general externalizing), along with two specific factors of callous-aggression ( $ESI_{AGG}$ ) and substance abuse ( $ESI_{SUS}$ ) onto which certain specific facet scales loaded; and *Model 4* (S-1 bifactor), in which four of the ESI facet scales (Problematic Impulsivity, Irresponsibility, Planful Control, Dependability) were specified as loading only onto the general factor, with the remaining subscales loading onto both the general factor and their specific factor (either  $ESI_{AGG}$  or  $ESI_{SUS}$ ).

The CFA modeling analyses were conducted using Diagonal Weighted Squares (DWLS) estimation as implemented in the lavaan module of the R statistical package (Rosseel, 2012). The robust DWLS methods is recommended in situations where and data are measured in ordinal scale and multivariate normality cannot be assumed (Mindrila, 2010; Remtulla et al., 2012). To evaluate goodness of fit for each model, the following quantitative fit statistics were used: the Comparative Fit Index (CFI), the

Tucker-Lewis Index (TLI), the Root Mean Square Error Approximation (RMSEA), the Standardized Root Mean Square Residual (SRMR), and the Bayesian Information Criterion (BIC). Following Hu and Bentler (1999), values of CFI and TLI above .95, and values of RMSEA and SRMR below .06 and .08, respectively, were considered indicative of good fit. Models with lower BIC are generally preferred.

We tested for measurement invariance across samples (clinical vs non-clinical and men vs women) in four steps. First, configural invariance (M1) was evaluated using a multi-group CFA without equality constraints. Second, the multi-group model was tested for metric measurement invariance (M2) by restricting all factor loadings to be equal across the two samples. Third, the model was tested for scalar measurement invariance (M3) by restricting all intercepts to be equal across samples. Finally, strict invariance (M4) was tested for by constraining the residual variances for all indicators to be equal across samples. In order to test each level of invariance, the difference in fit of the more constrained model was compared with that of the next less constrained model. Changes in the CFI and RMSEA of less than .01 and .015, respectively, were interpreted as evidence for comparability of model fit (Chen, 2007; Cheung & Rensvold, 1999).

In order to compare the factor structure across samples, coefficients of congruence (Lorenzo-Seva & Ten Berge, 2006) were computed for the factor loadings obtained in pertinent subsamples (clinical vs community; men vs women). Two exploratory factor analysis were conducted: one (as to replicate the general factor) and two factor (to replicate each of the specific factors – Callous Aggression and Substance Use) solution. Congruence coefficients in the range of .85-.94 correspond to fair similarity, while a value greater than .95 implies that the factors compared can be considered equal (Lorenzo-Seva & Ten Berge, 2006).

T-tests were performed to assess for mean differences in facet and item-based factor scores across the clinical and non-clinical samples, and across men and women samples. Effect sizes were quantified using Cohen's *d*.

Finally, to evaluate the evidence based on relations to other variables (criterion-related evidence) of the ESI-BF-S scores, correlation coefficients were computed for its three factor scales (Disinhibition, Callous Aggression and Substance Abuse) and its 23 facets scales with the five domain scales of the PID-5 (negative affect, detachment, antagonism, disinhibition and psychoticism) and the Alcohol scale of the SDSS. In order to maximize the comparability with previous studies that report results obtained with either community or clinical samples, present research calculated the mentioned correlations on each sub-sample separately.

Confirmatory factor analysis and tests of measurement invariance were performed using the lavaan package of the R software program (Rosseel, 2012). Congruence factor coefficients were calculated using the 'psych' package in R (Revelle, 2022). The remaining analyses were conducted using SPSS version 25. This study was not preregistered. The data that support the findings is available as supplementary file and hosted at [https://osf.io/49rxc/?view\\_only=dc156a1fe45f47dd918c5c35bd7e1433](https://osf.io/49rxc/?view_only=dc156a1fe45f47dd918c5c35bd7e1433).

## Results

### **ESI-BF-S Facet and Factor Scales: Descriptive Statistics and Reliability Estimates**

Table 1 shows descriptive statistics and reliability coefficients (McDonald's omega, Cronbach's alpha) for each facet scale and item-based factor scale of the ESI-BF-S. McDonald's omega coefficients ranged from .72 to .96 for the facet scales, and from .82 to .92 for the item-based factor scales. The alpha internal consistency coefficient fell just below .70 for the Irresponsibility scale, but exceeded .70 for all

other facet scales, resulting in a median alpha value of .79 across the 23 facet scales.

The item-based factor scales exhibited alpha coefficients from .79 to .90 (median = .83).

INSERT TABLE 1 HERE

### **ESI-BF-S Facet Scales: Validity Evidence based on Internal Structure**

Table 2 shows the quantitative fit statistics for the four models tested. The fit indices revealed good fit for the hierarchical two-subfactor model (Model 3) and the S-1 bifactor model (Model 4) in the study sample as a whole (community + patient participants), with CFI and TLI values  $> .95$ , SRMRs  $< .08$  and RMSEAs  $< .06$ . By contrast, the fit of the other two alternative models (one general factor, higher order two-subfactor) was inadequate (CFI/TLIs  $< .92$ , SRMRs  $> .10$ , RMSEAs  $\geq .07$ ).

INSERT TABLE 2

Considering concerns that have been raised about symmetric bifactor modeling (e.g., Green et al., 2019), along with the criterion of model parsimony, Model 4 was selected as the model to be tested for measurement invariance. Fit indices indicated similarly good fit for the clinical and non-clinical samples when analyzed separately compared to when analyzed together (see Table 3), providing evidence for configural invariance. However, when imposing the restrictions for metric invariance, differences between the two samples were found in loadings of nine facet scales on the General externalizing factor (Problematic Impulsivity, Fraud, Impatient Urgency, Planful Control, Blame Externalization, Rebelliousness, Physical Aggression, Empathy, Drug Problems) and one facet scale (Marijuana Use) on the Substance Abuse factor. The analysis for scalar invariance revealed non-equivalence of intercepts across sample for one scale (Problematic Impulsivity), and the analysis for strict invariance revealed non-equivalence of residual variances for one other scale (Boredom Proneness). When

comparing factor loadings for the community and clinical samples, the congruence coefficient for the general factor (retaining one single factor, with all facets loading onto this general factor) was .99. Congruence coefficients for the factors of the two-factor model (Callous Aggression and Substance Use) were each .97.

Concerning measurement invariance across gender subsamples, fit indices were indicative of good fit for each when they were analyzed separately, or together in the same model, supporting configural invariance. Although results suggested scalar and strict invariance (see table 3), differences between the two genders were found in loadings of eight of the facet scales onto the General externalizing factor (Fraud, Theft, Drug Problems, Drug Use, Marijuana Problems, Problematic Impulsivity, Alcohol Problems and Irresponsibility). Concerning the comparison of factor loadings across the two genders, the congruent coefficients were .99 (general factor) .92 (Callous Aggression) and .81 (Substance use)

INSERT TABLE 3

### **Tests of Mean Differences Between Non-Clinical and Clinical Samples and Between genders**

Table 4 shows means for the clinical and non-clinical samples, along with means for each gender sample, and tests of differences in sample means for each scale. Statistically significant differences were evident between the clinical and non-clinical samples for 19 of the 23 facet scales and 2 of the 3 factor scales (ESI<sub>AGG</sub> and ESI<sub>SUB</sub>). The greatest differences at the facet level were found for the Problematic Impulsivity and Alienation scales, with the clinical sample showing higher scores than the non-clinical sample on each. General disinhibition was the factor scale on which the clinical and non-clinical samples differed most. With regard to gender, 13 scales showed

statistically significant differences between men and women, with the Lack of Empathy and Alienation facet scales showing the largest effect sizes.

#### INSERT TABLE 4

### **Validity Evidence Based on the Relationship with Other Variables**

Table 5 shows Pearson correlations ( $r_s$ ) for the 23 facet scales, the three CFA model factors (ESI<sub>GF</sub>, ESI<sub>SF1</sub> and ESI<sub>SF2</sub>), and the three item-based factor scales of the ESI-BF-S (ESI<sub>DIS</sub>, ESI<sub>AGG</sub> and ESI<sub>SUS</sub>) with the five PID-5 scales (Negative Affect, Detachment, Antagonism, Disinhibition and Psychoticism) and the SDSS Alcohol scale, in the non-clinical and clinical samples separately.

The ESI-BF-S's three item-based factor scales showed positive  $r_s$  (in most cases significant) with all but 3 of the criterion scale measures (PID-5 domains, SDSS Alcohol). In both samples, ESI-DIS was related most strongly to PID-5 Disinhibition ( $r_s$  for non-clinical and clinical participants = .68 and .78, respectively), whereas ESI-AGG was related most strongly to PID-5 Antagonism ( $r_s$  = .58 and .53, respectively) and ESI-SUB correlated most with the SDSS Alcohol scale ( $r_s$  = .47 and .47). ESI-DIS also correlated at appreciable levels ( $r_s > .4$ ) with all other PID-5 domains as well with as the SDSS Alcohol scale, as did ESI-AGG, except for relating only weakly to PID-5 Negative Affect ( $r_s$  = .18 and .07). In terms of relations with the PID-5 domains, ESI-SUB showed  $r_s$  of .3 or higher with Disinhibition and Antagonism in both samples, and with Psychoticism in the clinical sample only. Of the factors derived from the CFA model, the one that showed the highest  $r$  values with all external measures was the general factor ( $r$  = .33 – .72). In contrast, the two specific factors showed moderate to low correlations with both the PID-5-SF domains and the total SDSS score.

The facet scales of the ESI-BF-S that related most strongly and selectively to PID-5 Disinhibition ( $r$ s with this domain  $> |.5|$ , and exceeding other PID-5 domains) were Problematic Impulsivity and Irresponsibility in the non-clinical sample and Problematic Impulsivity, Impatient Urgency, Planful Control, Dependability, Boredom Proneness, and Rebelliousness in the clinical sample. The ESI-BF-S facet scales that related most strongly and selectively to PID-5 Antagonism ( $r$ s with this domain  $> |.5|$ , and exceeding others) were Fraud and Relational Aggression in the non-clinical sample and Relational Aggression in the clinical sample. The only facet scales that correlated above  $|.4|$  and (nominally) more strongly with a different PID-5 domain than these two were Alienation and Blame Externalization (with Psychoticism) in both samples, and Empathy (with Detachment) in the non-clinical sample and Irresponsibility (with Psychoticism) in the clinical sample. The facet scale that related most to the Alcohol scale of the SDSS in both samples ( $r \geq |.6|$  in each case) was, unsurprisingly, the ESI-BF-S Alcohol Problems scale.

#### INSERT TABLE 5

### **Discussion**

The present study describes the process of developing the Spanish-language version of the ESI-BF and provides evidence of its psychometric properties in a mixed non-clinical/clinical sample consisting of adults from the community and patients seeking mental health services. Overall, the results showed adequate reliability for the ESI-BF-S's item-based factor scales scores and most of its facet scales in the overall participant sample (community members + patients). Similar to what has been reported in previous studies (Patrick et al., 2013; Soe-Agnie et al., 2021; Cremeans, 2019), internal structural analyses of the facet scales of the ESI-BF-S revealed superior fit for

the two bifactor models (symmetric, S-1) relative to the one-factor or two-factor higher-order models. The S-1 bifactor model was retained, due to its greater parsimony. In addition, the ESI-BF-S factor and facet scales of the ESI-BF-S showed associations with measures of maladaptive personality and alcohol consumption consisted with those reported previously by other authors (Venables & Patrick, 2012; Soe-Agnie et al., 2016; 2021), as well as evidence of effectiveness in discriminating between the clinical and community samples.

More specifically, as expected, internal consistency coefficients (omegas, alphas) for the three item-based factor scales and all but one of the 23 facet scales of the Spanish-language ESI-BF (Irresponsibility) were acceptable and commensurate with those for the original and Dutch versions of the ESI-BF (Patrick et al., 2013; Soe-Agnie et al., 2016; 2021). Of note, an unpublished study by Cremeans (2019) also reported an alpha coefficient below .70 (i.e., .52) for the Irresponsibility scale of the full-form English-language ESI. An explanation for the discrepancy in results for this scale in the present study and that of Cremeans relative to other studies that have reported higher reliability values (Patrick et al., 2013; Soe-Agnie et al., 2016) may lie in differences in characteristics of the samples used. Nonetheless, our findings together with those of Cremeans suggest a need for a review of the items comprising this facet scale to determine whether there might be vagaries in the wording of some items that may contribute to decreased internal consistency in some samples.

Turning to findings from internal structural analyses, CFAs of the ESI-BF-S yielded results similar to those reported in previous studies (Patrick et al., 2013; Sellbom, 2016; Soe-Agnie et al., 2016; 2021). Alternative versions of a hierarchical two-subfactor (bifactor) model, specifying a general externalizing factor along with two specific factors of callous-aggression ( $ESI_{AGG}$ ) and substance abuse ( $ESI_{SUB}$ ), exhibited

the best fit in both the community and clinical samples. Consistent with results from both the current study and prior research using the ESI-BF (Soe-Agnie et al., 2016; 2021), studies conducted with other measurement instruments (Cremeans, 2019; Sellbom, 2016) have also found superior fit for the two-subfactor bifactor model over the unifactor model, indicating a general contribution of the broad factor and specific contributions of the substance use and aggression subfactors to differing expressions of externalizing.

Furthermore, vis-à-vis concerns raised by some authors regarding a fit bias for fully symmetric bifactor models (Greene et al., 2019; Heinrich et al., 2020), the present work provides evidence for improved fit as well, over one-factor and two-factor higher order models, for a nonsymmetric (S-1) bifactor model in which not all ESI-BF-S facet scales loaded onto one or the other specific factor. This structure is consistent with the HiTOP model, which proposes the division of the externalizing super spectra into two distinct subdimensions: disinhibited externalizing, encompassing tendencies to act on impulse and use substances, and antagonistic externalizing, encompassing tendencies toward interpersonal conflict and aggressiveness (Kotov et al., 2017; Krueger et al., 2021). Results from the current structural analyses also support the observation of traits entailing problematic impulsivity, irresponsibility, deficient planfulness, and lack of dependability as pure indicators of general externalizing proneness (Krueger et al., 2021). The S-1 bifactor structure of the ESI-BF-S was found to be invariant across our two study samples and across gender subsamples, although only in terms of partial metric and scalar invariance. This result is congruent with findings from a study by Sellbom (2016) that used a different set of externalizing indicator, and found evidence only for configural invariance across community, forensic, and correctional samples. Of note, although items for the ESI-BF were selected from the full ESI using item response

theory methods with the primary aim of measuring facets of externalizing in a maximally efficient manner, evidence from the current work along with prior research supports a highly comparable factor structure for the full- and brief-form ESI versions.

Finally, in line with theoretical expectations, evidence for the validity of the ESI-BF-S scores was found in relation to criterion measures of maladaptive personality and substance use. Consistent with prior research (Soe-Agnie et al., 2021; Strickland et al., 2013), positive correlations were found for the ESI scale scores with the trait domains of the PID-5 and alcohol consumption as indexed by the SDSS Alcohol scale. With regard to item-based factors, we observed relationships (with larger effect sizes) between the ESI<sub>DIS</sub> and ESI<sub>AGG</sub> scales and the PID-5 domains of Disinhibition and Antagonism. These results are in accord with those reported by Strickland et al. (2013). These authors likewise found correlations of the Disinhibition and Meanness subscales of the Triarchic Psychopathy Measure (TriPM; Patrick, 2010), corresponding to the ESI<sub>DIS</sub> and ESI<sub>AGG</sub> scales of the ESI, to be highest with these two domains of the PID-5. Relatedly, Soe-Agnie et al. (2021) reported evidence for preferential associations of the ESI<sub>AGG</sub> and ESI<sub>DIS</sub> scales with the Agreeableness and Conscientiousness factors, respectively, of the NEO-FFI (Costa & McCrae, 1992). Results consistent with expectation were also found at the ESI facet level, where subscales of Problematic Impulsivity, Irresponsibility, Impatient Urgency, Planful Control, Boredom Proneness, and Rebelliousness were related most to PID-5 Disinhibition, and scales of Fraud and Relational Aggression were related most to PID-5 Antagonism. Of note, the four scales used as pure indicators of the general factor in the S-1 bifactor model of the ESI-BF-S were among the six scales that related most strongly to the Disinhibition domain of the PID-5.

Turning to substance-related measures, the ESI<sub>SUB</sub> factor scale of the ESI-BF-S (ESI<sub>SUB</sub>) and its Alcohol Problems facet related most strongly to the Alcohol scale of the SDSS, providing convergent evidence. However, moderate correlations were evident as well for other ESI facet scales with the SDSS Alcohol scale. These correlations are explainable in terms of the well documented relationship between disinhibitory traits and problematic drinking (Creswell et al., 2016; Jahng et al., 2011). In particular, higher ESI<sub>DIS</sub> scores have been linked to sex-related alcohol and drug consumption (Rodriguez-Seijas et al., 2017). Similarly, ESI<sub>DIS</sub> scores — when measured using the older 159-item shortened version of the ESI (Venables & Patrick, 2012) — were also correlated with a diagnosis of alcohol and substance use disorder in a sample of prisoners. In general terms, this evidence provides support for the HiTOP conceptualization of a disinhibitory subdimension of externalizing psychopathology encompassing impulsive traits and substance-related problems (Krueger et al., 2021; Sellbom et al., 2020).

Additionally, the present study found moderate effect sizes for the ESI<sub>DIS</sub> scale and its constituent facets in differentiating clinical sample participants from community sample participants, with differentiation especially evident for facets of problematic impulsivity, boredom proneness, alienation, and blame externalization. Although this is the first piece of evidence supporting the effectiveness of the ESI-BF for distinguishing mental health patients from community participants, it converges with findings from previous work comparing forensic patients with community controls (Soe-Agnie et al., 2021). From a clinical point of view, these results support the use of this instrument in applied contexts with different clinical samples, including patients exhibiting affective disorders or personality disorders. In addition, the ESI-BF may be particularly helpful for identifying subtypes of high externalizing antisocial patients who, according to

recent research by Brennan et al., (2017), exhibit different patterns of substance misuse. Measurement at the lower-order facet level may contribute to characterizing patient subtypes more effectively and identifying better targets for intervention.

Some limitations of the current study warrant mention. One is that criterion measures were limited to self-report questionnaires. It will be useful to include measures from other modalities, such as interview-assessed symptom scores and pertinent lab performance measures, in future research. Finally, evaluation of substance use in the current study was limited to alcohol consumption, which is the most widely used substance in the general population. Although relationships found for this measure were congruent with our hypotheses, it will be useful to extend this analysis to other psychoactive substances.

Notwithstanding these limitations, results of the current work provide evidence that the Spanish adaptation of the ESI-BF shows psychometric properties similar to those observed for both the original and Dutch versions (Soe-Agnie et al., 2016). Additionally, in relation to prior studies, ours is the first to use a randomly selected community sample balanced with respect to age range and gender along with a general sample of mental health patients. Taken together, results from the current study provide evidence for the reliability and validity of the scores of the Spanish-language version of the ESI-BF, and further support the use of this measure in research related to the externalizing spectrum, and potentially in clinical assessment work.

## References

Allen, T. A., Schreiber, A. M., Hall, N. T., & Hallquist, M. N. (2020). From description to explanation: Integrating across multiple levels of analysis to inform neuroscientific accounts of dimensional personality pathology. *Journal of*

*personality disorders*, 34(5), 650-676.

<https://doi.org/10.1521/pedi.2020.34.5.650>

Bernat, E. M., Nelson, L. D., Steele, V. R., Gehring, W. J., & Patrick, C. J. (2011).

Externalizing psychopathology and gain–loss feedback in a simulated gambling task: Dissociable components of brain response revealed by time-frequency analysis. *Journal of abnormal psychology*, 120(2), 352-64.

<https://doi.org/10.1037/a0022124>

Blonigen, D. M., Patrick, C. J., Gasperi, M., Steffen, B., Ones, D. S., Arvey, R. D., de

Oliveira Baumgarti, V. & do Nascimento, E. (2011). Delineating the construct network of the Personnel Reaction Blank: Associations with externalizing tendencies and normal personality. *Psychological assessment*, 23(1), 18-30.

<https://doi.org/10.1037/a0021048>

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor

model as a structure of psychopathology. *Clinical Psychological Science*, 5, 184–186. <http://dx.doi.org/10.1177/2167702616657069>

Brennan, G. M., Stuppy-Sullivan, A. M., Brazil, I. A., & Baskin-Sommers, A. R.

(2017). Differentiating patterns of substance misuse by subtypes of antisocial traits in male offenders. *The Journal of Forensic Psychiatry &*

*Psychology*, 28(3), 341-356. <http://dx.doi.org/10.1080/14789949.2017.1280072>.

Burke, J.D., Johnston, O.G. (2020). The Bifactor S-1 Model: A Psychometrically

Sounder Alternative to Test the Structure of ADHD and ODD? *Journal of Abnormal Child Psychology*, 48(7), 911-915. <https://doi.org/10.1007/s10802-020-00645-4>. PMID: 32285353.

Byrne, K. A., Otto, A. R., Pang, B., Patrick, C. J., & Worthy, D. A. (2019). Substance

use is associated with reduced devaluation sensitivity. *Cognitive, Affective, &*

- Behavioral Neuroscience*, 19(1), 40-55. <https://doi.org/10.3758/s13415-018-0638-9>
- Byrne, K. A., & Worthy, D. A. (2019). Examining the link between reward and response inhibition in individuals with substance abuse tendencies. *Drug and alcohol dependence*, 194, 518-525  
<https://doi.org/10.1016/j.drugalcdep.2018.11.014>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464– 504.  
<https://doi.org/10.1080/10705510701301834>
- Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, 25(1), 1–27. <https://doi.org/10.1177/014920639902500101>
- Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological assessment*, 4(1), 5-13.  
<https://doi.org/10.1037/10403590.4.1.5>
- Cremeans, M. (2019). *An investigation of the structure of externalizing psychopathology*. Unpublished Master's Thesis. Balls State University.  
<https://cardinalscholar.bsu.edu/handle/123456789/202012>
- Creswell, K. G., Bachrach, R. L., Wright, A. G., Pinto, A., & Ansell, E. (2016). Predicting problematic alcohol use with the DSM–5 alternative model of personality pathology. *Personality Disorders: Theory, Research, and Treatment*, 7(1), 103-11. <https://doi.org/10.1037/per0000131>
- Dacosta-Sánchez, D., Fernández-Calderón, F., González-Ponce, B., Díaz-Batanero, C., & Lozano, Ó. M. (2019). Severity of substance use disorder: Utility as an outcome in clinical settings. *Alcoholism: Clinical and Experimental*

*Research*, 43(5), 869-876. <https://doi.org/10.1111/acer.14020>

Diaz-Batanero, C., (2022). ESI-BF Spanish adaptation database. Retrieved from [osf.io/49rxc](https://osf.io/49rxc).

Diaz-Batanero, C., Ramírez-López, J., Domínguez-Salas, S., Fernández-Calderón, F., & Lozano, Ó. M. (2019). Personality inventory for DSM-5–short form (PID-5-SF): Reliability, factorial structure, and relationship with functional impairment in dual diagnosis patients. *Assessment*, 26(5), 853-866. <https://doi.org/10.1177/1073191117739980>

Eid, M. (2020). Multi-faceted constructs in abnormal psychology: Implications of the bifactor S-1 model for individual clinical assessment. *Journal of abnormal child psychology*, 48(7), 895-900.

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods*, 22, 541–562. <https://doi.org/10.1037/met0000083>.

Forbes, M. K., Sunderland, M., Rapee, R. M., Batterham, P. J., Callear, A. L., Carragher, N., ... & Krueger, R. F. (2021). A detailed hierarchical model of psychopathology: From individual symptoms, up to the general factor of psychopathology. *Clinical Psychological Science*, 9(2), 139-168. <https://doi.org/10.1177/2167702620954799>

Gonzalez-Sainz, F., Lozano, O. M., Velez-Moreno, A., Ramirez, J. L. (2014). *Manual de la Escala de Gravedad de la Dependencia de Sustancias*. 1st ed. Servicio de Publicaciones de la Universidad de Huelva, Spain.

Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., ... & Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of abnormal psychology*, 128(7), 740.

<https://doi.org/10.1037/abn0000434>

Gregoire, J. (2018). ITC guidelines for translating and adapting tests. *International Journal of Testing*, 18(2), 101-134.

<https://doi.org/10.1080/15305058.2017.1398166>

Hall, J. R., Bernat, E. M., & Patrick, C. J. (2007). Externalizing psychopathology and the error-related negativity. *Psychological science*, 18(4), 326-333.

<https://doi.org/10.1111/j.1467-9280.2007.01899.x>

Heinrich, M., Zagorscak, P., Eid, M., & Knaevelsrud, C. (2020). Giving G a meaning:

An application of the bifactor-(S-1) approach to realize a more symptom-oriented modeling of the Beck depression inventory–II. *Assessment*, 27(7),

1429-1447. <https://doi.org/10.1177/1073191118803738>

Hengartner, M. P., & Lehmann, S. N. (2017). Why psychiatric research must abandon traditional diagnostic classification and adopt a fully dimensional scope: two solutions to a persistent problem. *Frontiers in Psychiatry*, 8, 101.

<https://doi.org/10.3389/fpsyt.2017.00101>

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55

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

Jahng, S., Trull, T. J., Wood, P. K., Tragesser, S. L., Tomko, R., Grant, J. D., Bucholz, K. K. & Sher, K. J. (2011). Distinguishing general and specific personality disorder features and implications for substance dependence comorbidity. *Journal of abnormal psychology*, 120(3), 656-69.

<https://doi.org/10.1037/a0023539>

Junghänel, M., Rodenacker, K., Dose, C. et al. Applying the Bifactor S-1 Model to

Ratings of ADHD/ODD Symptoms: A Commentary on Burns et al. (2019) and a Re-Analysis. *Journal of Abnormal Child Psychology*, 48, 905–910 (2020).

<https://doi.org/10.1007/s10802-020-00637-4>

Kim, H., & Eaton, N. R. (2015). The hierarchical structure of common mental disorders: Connecting multiple levels of comorbidity, bifactor models, and predictive validity. *Journal of abnormal psychology*, 124(4), 1064.

<https://doi.org/10.1037/abn0000113>

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Avshalom, C., Clark, A. L., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., ... Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of abnormal psychology*, 126(4), 454.

<https://doi.org/10.1037/abn0000258>

Kotov, R., Krueger, R. F., Watson, D., Cicero, D. C., Conway, C. C., DeYoung, C. G., Eaton, N. R., Forbes, M. K., Hallquist, M. N., Latzman, R. D., Mullins-Sweat, S. N., Ruggero, C. J., Simms, L. J., Waldman, I. D., Waszczuk, M. A., & Wright, A. G. (2021). The Hierarchical Taxonomy of Psychopathology (HiTOP): A quantitative nosology based on consensus of evidence. *Annual review of clinical psychology*, 17, 83-108. <https://doi.org/10.1146/annurev-clinpsy-081219-093304>

Krueger, R. F., Hobbs, K. A., Conway, C. C., Dick, D. M., Dretsch, M. N., Eaton, N. R., Forbes, M. K., Forbush, K. T., Keyes, K. M., Latzman, R. D., Michelini, G., Patrick, C. J., Sellbom, M., Slade, T., South, S. C., Sunderland, M., Tackett, J., Waldman, I., Waszczuk, M. A., ...HiTOP Utility Workgroup. (2021). Validity

- and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): II. Externalizing superspectrum. *World Psychiatry*, 20(2), 171-193.  
<https://doi.org/10.1002/wps.20844>
- Krueger, R. F., Markon, K. E., Patrick, C. J., Benning, S. D., & Kramer, M. D. (2007). Linking antisocial behavior, substance use, and personality: an integrative quantitative model of the adult externalizing spectrum. *Journal of abnormal psychology*, 116(4), 645. <https://doi.org/10.1037/0021-843X.116.4.645>
- Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: a dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of abnormal psychology*, 114(4), 537. <https://doi.org/10.1037/0021-843X.114.4.537>
- Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(2), 57–64. <https://doi.org/10.1027/1614-2241.2.2.57>
- Maples, J. L., Carter, N. T., Few, L. R., Crego, C., Gore, W. L., Samuel, D. B., Williamson, R. L., Lynam, D. R., Widiger, T. A., Markon, K. E., Krueger, R. F., & Miller, J. D. (2015). Testing whether the DSM-5 personality disorder trait model can be measured with a reduced set of items: An item response theory investigation of the Personality Inventory for DSM-5. *Psychological Assessment*, 27(4), 1195-1210. <https://doi.org/10.1037/pas0000120>  
<https://doi.org/10.1037/pas0000120>
- Mindrila, D. (2010). Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) Estimation Procedures: A Comparison of Estimation Bias with Ordinal and Multivariate Non-Normal Data. *International Journal of Digital*

*Society (IJDS)*, 1 (1), 6029 - 66

Nelson, L. D., Patrick, C. J., & Bernat, E. M. (2011). Operationalizing proneness to externalizing psychopathology as a multivariate psychophysiological phenotype. *Psychophysiology*, 48(1), 64-72. <https://doi.org/10.1111/j.1469-8986.2010.01047.x>

Nelson, L. D., Strickland, C., Krueger, R. F., Arbisi, P. A., & Patrick, C. J. (2016). Neurobehavioral traits as transdiagnostic predictors of clinical problems. *Assessment*, 23(1), 75-85. <https://doi.org/10.1177/1073191115570110>

Pasion, R., & Barbosa, F. (2019). ERN as a transdiagnostic marker of the internalizing-externalizing spectrum: A dissociable meta-analytic effect. *Neuroscience & Biobehavioral Reviews*, 103, 133-149. <https://doi.org/10.1016/j.neubiorev.2019.06.013>

Patrick, C. J. (2010). *Triarchic psychopathy measure (TriPM)*. Retrieved from <https://www.phenxtoolkit.org/index.php?pageLink=browse.protocoldetails&id=121601>.

Patrick, C. J., Kramer, M. D., Krueger, R. F., & Markon, K. E. (2013). Optimizing efficiency of psychopathology assessment through quantitative modeling: development of a brief form of the Externalizing Spectrum Inventory. *Psychological assessment*, 25(4), 1332-48. <https://doi.org/10.1037/a0034864>

Patrick, C. J., Venables, N. C., Yancey, J. R., Hicks, B. M., Nelson, L. D., & Kramer, M. D. (2013). A construct-network approach to bridging diagnostic and physiological domains: application to assessment of externalizing psychopathology. *Journal of Abnormal Psychology*, 122(3), 902. <https://doi.org/10.1037/a0032807>

- Poy, R., Segarra, P., Esteller, À., López, R., & Moltó, J. (2014). FFM description of the triarchic conceptualization of psychopathy in men and women. *Psychological Assessment, 26*(1), 69–76. <https://doi.org/10.1037/a0034642>
- Revelle, W. (2022). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.2.5, <https://CRAN.R-project.org/package=psych>.
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological methods, 17*, 354-373. <https://doi.org/10.1037/a0029315>.
- Rodriguez-Seijas, C., Arfer, K. B., Thompson Jr, R. G., Hasin, D. S., & Eaton, N. R. (2017). Sex-related substance use and the externalizing spectrum. *Drug and alcohol dependence, 174*, 39-46. <https://doi.org/10.1016/j.drugalcdep.2017.01.008>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Sellbom, M. (2016). Elucidating the validity of the externalizing spectrum of psychopathology in correctional, forensic, and community samples. *Journal of Abnormal Psychology, 125*(8), 1027-38. <https://doi.org/10.1037/abn0000171>
- Sellbom, M., Carragher, N., Sunderland, M., Calear, A. L., & Batterham, P. J. (2020). The role of maladaptive personality domains across multiple levels of the HiTOP structure. *Personality and mental health, 14*(1), 30-50. <https://doi.org/10.1002/pmh.1461>
- Soe-Agnie, S. E., Paap, M. C. S., VanDerNagel, J. E. L., Nijman, H. J. M., & De Jong, C. A. J. (2018). The generalizability of the structure of substance abuse and

- antisocial behavioral syndromes: A systematic review. *Psychiatry research*, 259, 412-421. <https://doi.org/10.1016/j.psychres.2017.11.007>
- Soe-Agnie, S. E., Paap, M. C., Nijman, H. L., & De Jong, C. A. (2021). Psychometric properties of the externalizing spectrum inventory: Replication and extension across clinical and non-clinical samples. *Journal of personality assessment*, 103(3), 332-341. <https://doi.org/10.1080/00223891.2020.1753752>
- Soe-Agnie, S. E., Patrick, C. J., Nijman, H. L., & De Jong, C. A. (2016). Validation of the full and brief Externalizing Spectrum Inventory in Dutch forensic inpatients. *The Journal of Forensic Psychiatry & Psychology*, 27(1), 77-91. <https://doi.org/10.1080/14789949.2015.1090621>
- Strickland, C. M., Drislane, L. E., Lucy, M., Krueger, R. F., & Patrick, C. J. (2013). Characterizing psychopathy using DSM-5 personality traits. *Assessment*, 20(3), 327-338. <https://doi.org/10.1177/1073191113486691>
- Venables, N. C. (2017). Reward mechanisms across aggressive and addictive forms of externalizing psychopathology. *Biological psychiatry*, 82(4), e25-e27. <https://doi.org/10.1016/j.biopsych.2017.06.028>
- Venables, N. C., Foell, J., Yancey, J. R., Kane, M. J., Engle, R. W., & Patrick, C. J. (2018). Quantifying inhibitory control as externalizing proneness: A cross-domain model. *Clinical Psychological Science*, 6(4), 561-580. <https://doi.org/10.1177/2167702618757690>
- Venables, N. C., & Patrick, C. J. (2012). Validity of the Externalizing Spectrum Inventory in a criminal offender sample: relations with disinhibitory psychopathology, personality, and psychopathic features. *Psychological assessment*, 24(1), 88-100. <https://doi.org/10.1037/a0024703>
- Wygant, D. B., Sellbom, M., Sleep, C. E., Wall, T. D., Applegate, K. C., Krueger, R. F.,

- & Patrick, C. J. (2016). Examining the DSM–5 alternative personality disorder model operationalization of antisocial personality disorder and psychopathy in a male correctional sample. *Personality Disorders: Theory, Research, and Treatment*, 7(3), 229-239. <https://doi.org/10.1037/per0000179>
- Yancey, J. R., Venables, N. C., Hicks, B. M., & Patrick, C. J. (2013). Evidence for a heritable brain basis to deviance-promoting deficits in self-control. *Journal of Criminal Justice*, 41(5), 309-317. <https://doi.org/10.1016/j.jcrimjus.2013.06.002>
- Zimmerman, M., Ellison, W., Young, D., Chelminski, I., & Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Comprehensive psychiatry*, 56, 29-34  
<https://doi.org/10.1016/j.comppsy.2014.09.007>

**Table 1.***Descriptive statistics and reliability estimation of the ESI-BF-S scales and factors*

	<i>M (SD)</i>	Total sample		N° items
		McDonald $\omega$	Cronbach's $\alpha$	
<i>Scales (facets)</i>				
Problematic impulsivity	2.91 (3.92)	.84	.79	7
Irresponsibility	3.12 (3.55)	.73	.69	10
Theft	1.14 (2.57)	.89	.82	8
Fraud	0.82 (1.65)	.81	.72	6
Impatient urgency	5.13 (3.79)	.84	.81	5
Lack of planful control	4.51 (3.71)	.82	.79	6
Lack of dependability	3.68 (3.14)	.80	.73	7
Alienation	3.70 (2.48)	.72	.71	3
Boredom proneness	4.13 (3.53)	.87	.85	4
Blame externalization	3.01 (3.47)	.88	.88	4
Lack of honesty	2.89 (2.87)	.79	.75	5
Rebelliousness	2.67 (3.34)	.84	.80	6
Physical aggression	2.18 (3.12)	.83	.75	8
Destructive aggression	0.87 (2.17)	.88	.84	7
Relational aggression	2.83 (3.04)	.84	.78	8
Lack of empathy	5.26 (4.21)	.79	.74	11
Excitement seeking	2.62 (3.03)	.82	.74	6
Marijuana use	3.81 (5.94)	.96	.93	7
Marijuana problems	1.12 (3.15)	.91	.86	7
Drug use	3.19 (3.97)	.85	.78	6
Drug problems	1.74 (4.65)	.93	.91	11
Alcohol use	10.05 (5.93)	.83	.79	9
Alcohol problems	2.35 (4.50)	.88	.85	9
<i>Item-based factors</i>				
General disinhibition (ESI <sub>DIS</sub> )	10.78 (7.90)	.85	.83	20
Callous aggression (ESI <sub>AGG</sub> )	7.76 (5.67)	.82	.79	19
Substance abuse (ESI <sub>SUS</sub> )	10.54 (9.86)	.92	.90	18

**Table 2***Fit Statistics for Confirmatory Factor Models for the subsamples and total sample*

Model	$\chi^2$	df	$\chi^2/df$	CFI	TLI	SRMR	RMSEA	90% RMSEA CI	BIC**
Model 1. One general factor	1529.25	230	6.65	.910	.901	.108	.073	.069/.076	122339.61
Model 2. Higher order two-subfactor	1412.98	229*	6.17	.918	.910	.101	.069	.066/.073	121704.96
Model 3. Hierarchical two-subfactor (bifactor)	689.59	207	3.33	.967	.960	.072	.046	.042/.050	120045.99
Model 4. Hierarchical two-subfactor (S-1 bifactor)	704.12	211	3.34	.966	.959	.073	.047	.043/0.50	120073.35

*Note.* Estimation Method: Diagonally Weighted Least Squares (DWLS); CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error Approximation; SRMR = Standardized Root Mean Square Residual. All chi-square values are statistically significant. \*= In second order CFA models, the higher order factor must have at least three lower order factors (Bollen, 1989). As there are only two lower order factors, the variances of the errors of the first order factors have been set equal (Rosson et al, 2008). BIC = Bayesian Information Criterion (\*\*Maximum Likelihood-Robust estimation)

**Table 3**

*S1-Bifactor model measurement invariance analysis across samples (non-clinical vs clinical sample, and men vs women)*

<i>Non-clinical vs clinical</i>											
	$\chi^2$	<i>df</i>	$\chi^2/df$	<i>CFI</i>	$\Delta CFI$	<i>TLI</i>	<i>SRMR</i>	<i>RMSEA</i>	$\Delta RMSEA$	<i>RMSEA 90% CI</i>	
Non-Clinical	421.06	211	1.99	.978	--	.974	.071	.037		.032/.042	
Clinical	303.63	211	1.44	.983	--	.979	.079	.036		.027/.045	
M1. Configural Invariance	724.69	422	1.72	.980	--	.976	.070	.037		.032/.041	
M2. Metric Invariance (M1 vs M2)	1663.02	464	3.58	.920	.060	.913	.100	.069	.032	.066/.073	
M2B. Partial Metric Invariance (M1 vs M2B) <sup>a</sup>	895.41	454	1.97	.971	.009	.967	.077	.043	.006	.038/.047	
M3. Scalar Invariance (M2B vs M3)	1090.88	474	2.30	.959	.012	.956	.082	.049	.006	.045/.053	
M3B. Partial Scalar Invariance (M2B vs M3B) <sup>b</sup>	1052.66	473	2.23	.962	.009	.959	.081	.048	.005	.044/.052	
M4. Strict Invariance (M3B vs M4)	1279.02	496	2.58	.948	.014	.947	.092	.054	.006	.051/.058	
M4B. Partial Strict Invariance (M3B vs M4B) <sup>c</sup>	1223.84	495	2.47	.952	.010	.951	.092	.052	.004	.049/.056	
<i>Men vs women</i>											
	$\chi^2$	<i>df</i>	$\chi^2/df$	<i>CFI</i>	$\Delta CFI$	<i>TLI</i>	<i>SRMR</i>	<i>RMSEA</i>	$\Delta RMSEA$	<i>RMSEA 90% CI</i>	
Men	370.23	211	1.75	.980	--	.976	.075	.038		.031/.044	
Women	474.97	211	2.25	.960	--	.952	.081	.048		.042/.054	
M1. Configural Invariance	845.21	422	2.00	.971	--	.965	.075	.043		.039/.048	

M2. Metric Invariance (M1 vs M2)	1532.74	464	3.30	.926	.045	.920	.110	.066	.023	.062/.069
M2B. Partial Metric Invariance (M1 vs M2B) <sup>d</sup>	1003.70	456	2.20	.962	.009	.958	.081	.048	.005	.044/.051
M3. Scalar Invariance (M2B vs M3)	1080.69	476	2.27	.958	.004	.956	.083	.049	.001	.045/.053
M4. Strict Invariance (M3 vs M4)	1138.48	499	2.28	.956	.002	.955	.087	.049	.000	.045/.053

---

Note: Parameters released in each partial invariance models are: <sup>a</sup> Loadings from General factor to Problematic Impulsivity, Fraud, Impatient Urgency, Planful Control, Blame Externalization, Rebelliousness, Physical Aggression, Empathy, Drug Problems; Loading from Substance Abuse factor to Marijuana Use facet. <sup>b</sup> Intercept of Problematic Impulsivity. <sup>c</sup> Residual variances for Boredom Proneness. <sup>d</sup> Loadings from General externalizing factor to Fraud, Theft, Drug Problems, Drug Use, Marijuana Problems, Problematic Impulsivity, Alcohol Problems and Irresponsibility facets.

**Table 4.** Mean differences between the clinical and non-clinical samples and gender of the ESI-BF-S scales and factors

	Clinical sample	Non-clinical sample	Comparison of mean		Men	Women	Comparison of mean	
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t</i>	<i>d</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>t</i>	<i>d</i>
<i>Scales (facets)</i>								
Problematic impulsivity	5.23 (5.08)	1.88 (2.68)	11.31**	0.82	2.97 (4.22)	2.86 (3.59)	0.49	0.02
Irresponsibility	4.12 (4.08)	2.68 (3.20)	5.72**	0.39	3.34 (3.75)	2.91 (3.33)	1.96*	0.12
Theft	1.11 (2.69)	1.16 (2.52)	0.32	0.02	1.34 (3.04)	0.94 (1.98)	2.59*	0.15
Fraud	0.65 (1.37)	0.89 (1.76)	2.44*	0.15	1.04 (1.94)	0.60 (1.28)	4.31**	0.26
Impatient urgency	6.09 (4.43)	4.70 (3.38)	5.12**	0.35	4.93 (3.70)	5.32 (3.87)	1.70	0.10
Lack of planful control	6.06 (4.38)	3.82 (3.13)	8.41**	0.59	4.15 (3.50)	4.86 (3.86)	3.14**	0.19
Lack of dependability	4.19 (3.27)	3.46 (3.06)	3.54**	0.23	3.60 (3.21)	3.76 (3.07)	0.87	0.05
Alienation	5.06 (2.60)	3.09 (2.17)	12.06**	0.82	3.32 (2.41)	4.07 (2.49)	5.07**	0.30
Boredom proneness	5.98 (3.92)	3.30 (3.00)	11.06**	0.77	3.67 (3.36)	4.57 (3.63)	4.25**	0.25
Blame externalization	4.79 (4.01)	2.22 (2.86)	10.53**	0.74	2.84 (3.29)	3.18 (3.61)	1.62	0.09
Lack of honesty	2.31 (2.76)	3.15 (2.88)	4.46**	0.30	3.13 (2.91)	2.65 (2.79)	2.76**	0.16
Rebelliousness	3.32 (4.17)	2.38 (2.85)	3.73**	0.26	2.78 (3.40)	2.56 (3.28)	1.08	0.07
Physical aggression	2.71 (3.98)	1.94 (2.61)	3.22*	0.23	2.48 (3.33)	1.88 (2.86)	3.15**	0.19
Destructive aggression	1.13 (2.37)	0.75 (2.07)	2.56*	0.17	0.91 (2.20)	0.82 (2.13)	0.71	0.04
Relational aggression	2.62 (3.14)	2.93 (2.99)	1.56	0.10	3.01 (3.22)	2.66 (2.84)	1.86	0.11
Lack of empathy	3.88 (4.21)	5.87 (4.07)	7.35**	0.48	6.21 (4.41)	4.31 (3.78)	7.57**	0.45
Excitement seeking	2.83 (3.55)	2.53 (2.77)	1.39	0.09	2.90 (3.13)	2.33 (2.91)	3.10**	0.18
Marijuana use	4.74 (6.87)	3.39 (5.43)	3.19*	0.22	3.99 (6.32)	3.62 (5.53)	1.00	0.06
Marijuana problems	1.59 (4.13)	0.91 (2.56)	2.77*	0.20	1.35 (3.51)	0.87 (2.71)	2.52*	0.15
Drug use	3.05 (4.25)	3.25 (3.83)	0.76	0.05	3.69 (4.25)	2.69 (3.60)	4.16**	0.25
Drug problems	2.76 (6.15)	1.28 (3.70)	4.06**	0.29	2.16 (5.43)	1.32 (3.65)	2.97**	0.18
Alcohol use	8.74 (6.55)	10.63 (5.54)	4.58**	0.31	10.91 (5.57)	9.19 (6.16)	4.80**	0.28
Alcohol problems	2.98 (5.95)	2.07 (3.64)	2.58*	0.18	2.67 (4.83)	2.03 (4.13)	2.34*	0.14
<i>Factor scores</i>								
General Factor (ESI <sub>GF</sub> )	0.27 (1.09)	-0.12 (0.88)	5.73**	0.40	0.05 (1.06)	-0.05 (0.87)	1.74	0.10
Specific Factor 1 (ESI <sub>SF1</sub> )	-0.56 (0.84)	0.24 (0.68)	15.21**	0.98	0.19 (0.74)	-0.19 (0.85)	8.01**	0.47
Specific Factor 2 (ESI <sub>SF2</sub> )	-0.05 (1.07)	0.03 (0.82)	1.25	0.09	0.08 (0.96)	-0.08 (0.84)	2.95*	0.18
<i>Item-based factors</i>								
General disinhibition (ESI <sub>DIS</sub> )	14.96 (9.12)	8.91 (6.47)	10.92**	0.75	10.47 (8.45)	11.08 (7.30)	1.24	0.07
Callous aggression (ESI <sub>AGG</sub> )	6.46 (5.62)	8.34 (5.59)	5.09**	0.34	8.84 (6.02)	6.69 (5.07)	6.32**	0.38
Substance abuse (ESI <sub>SUS</sub> )	10.92 (11.10)	10.37 (9.25)	0.784	0.05	11.50 (10.26)	9.59 (9.35)	3.18**	0.19

Note. *d* = Cohen's *d*, effect size. \**p* < .05, \*\**p* < .001

**Table 5.** Pearson correlation coefficients between scores on specific ESI-BF-S scales and general factors and PID-5 and SDSS in the non-clinical / clinical samples respectively

	PID-5 Negative Affect	PID-5 Detachment	PID-5 Antagonism	PID-5 Disinhibition	PID-5 Psychoticism	SDSS Alcohol
<i>Facet scales</i>						
Problematic Impulsivity	<b>.35/.38</b>	<b>.39/.34</b>	<b>.51/.39</b>	<b>.59/.71</b>	<b>.48/.53</b>	<b>.44/.34</b>
Irresponsibility	<b>.31/.25</b>	<b>.37/.25</b>	<b>.42/.30</b>	<b>.51/.43</b>	<b>.46/.48</b>	<b>.42/.29</b>
Theft	<b>.21/.15</b>	<b>.35/.07</b>	<b>.43/.40</b>	<b>.45/.30</b>	<b>.41/.27</b>	<b>.46/.25</b>
Fraud	<b>.26/.17</b>	<b>.37/.15</b>	<b>.52/.44</b>	<b>.46/.28</b>	<b>.46/.35</b>	<b>.44/.16</b>
Impatient Urgency	<b>.40/.33</b>	<b>.34/.12</b>	<b>.49/.48</b>	<b>.45/.52</b>	<b>.39/.36</b>	<b>.21/.19</b>
Lack of Planful Control	<b>.13/.31</b>	<b>.21/.31</b>	<b>.23/.21</b>	<b>.41/.66</b>	<b>.23/.37</b>	<b>.31/.29</b>
Lack of Dependability	<b>.20/.27</b>	<b>.36/.23</b>	<b>.30/.22</b>	<b>.45/.54</b>	<b>.33/.32</b>	<b>.36/.32</b>
Alienation	<b>.39/.45</b>	<b>.36/.45</b>	<b>.27/.16</b>	<b>.38/.35</b>	<b>.41/.50</b>	<b>.14/.15</b>
Boredom Proneness	<b>.39/.41</b>	<b>.45/.41</b>	<b>.37/.21</b>	<b>.49/.54</b>	<b>.39/.46</b>	<b>.23/.20</b>
Blame Externalization	<b>.38/.38</b>	<b>.40/.33</b>	<b>.41/.24</b>	<b>.45/.46</b>	<b>.49/.49</b>	<b>.28/.15</b>
Lack of Honesty	<b>.10/.13</b>	<b>.22/.05</b>	<b>.31/.37</b>	<b>.31/.33</b>	<b>.23/.19</b>	<b>.24/.22</b>
Rebelliousness	<b>.15/.21</b>	<b>.30/.26</b>	<b>.42/.44</b>	<b>.44/.54</b>	<b>.42/.47</b>	<b>.38/.34</b>
Physical Aggression	<b>.24/.14</b>	<b>.37/.20</b>	<b>.44/.40</b>	<b>.41/.41</b>	<b>.45/.33</b>	<b>.42/.33</b>
Destructive Aggression	<b>.23/.23</b>	<b>.41/.16</b>	<b>.43/.37</b>	<b>.39/.43</b>	<b>.41/.39</b>	<b>.48/.32</b>
Relational Aggression	<b>.31/.23</b>	<b>.44/.21</b>	<b>.62/.61</b>	<b>.52/.49</b>	<b>.46/.49</b>	<b>.40/.29</b>
Lack of Empathy	<b>.05 /-.06</b>	<b>.42/.11</b>	<b>.41/.31</b>	<b>.32/.22</b>	<b>.33/.20</b>	<b>.31/.13</b>
Excitement Seeking	<b>.16/.12</b>	<b>.28/.05</b>	<b>.44/.43</b>	<b>.39/.36</b>	<b>.41/.35</b>	<b>.36/.40</b>
Marijuana Use	<b>.13/.05</b>	<b>.23/-.01</b>	<b>.24/.28</b>	<b>.28/.16</b>	<b>.23/.27</b>	<b>.32/.23</b>
Marijuana Problems	<b>.14/.12</b>	<b>.32/.10</b>	<b>.31/.26</b>	<b>.33/.25</b>	<b>.31/.34</b>	<b>.46/.25</b>
Drug Use	<b>.07 /.03</b>	<b>.23/.01</b>	<b>.22/.32</b>	<b>.24/.15</b>	<b>.21/.30</b>	<b>.34/.28</b>
Drug Problems	<b>.18/.18</b>	<b>.32/.14</b>	<b>.33/.37</b>	<b>.37/.35</b>	<b>.37/.40</b>	<b>.49/.35</b>
Alcohol Use	<b>.06 /.06</b>	<b>.17/-.06</b>	<b>.26/.21</b>	<b>.25/.28</b>	<b>.17/.14</b>	<b>.42/.57</b>
Alcohol Problems	<b>.17/.21</b>	<b>.38/.14</b>	<b>.39/.30</b>	<b>.40/.38</b>	<b>.36/.33</b>	<b>.64/.66</b>
<i>Factor scores</i>						
General Factor (ESI <sub>GF</sub> )	<b>.37/.36</b>	<b>.53/.33</b>	<b>.61/.61</b>	<b>.64/.56</b>	<b>.66/.72</b>	<b>.57/.47</b>
Specific Factor 1 (ESI <sub>SF1</sub> )	<b>-.31/-.42</b>	<b>-.13/-.37</b>	<b>-.15/-.33</b>	<b>.02/.10</b>	<b>-.12/-.21</b>	<b>.16/.06</b>
Specific Factor 2 (ESI <sub>SF2</sub> )	<b>-.05/-.09</b>	<b>.04/-.14</b>	<b>-.02/.07</b>	<b>-.01/.12</b>	<b>.01/-.10</b>	<b>.20/.17</b>
<i>Item-based factor scales</i>						
ESI <sub>DIS</sub>	<b>.46/.49</b>	<b>.52/.38</b>	<b>.57/.49</b>	<b>.68/.78</b>	<b>.57/.63</b>	<b>.44/.38</b>
ESI <sub>AGG</sub>	<b>.18/.08</b>	<b>.49/.20</b>	<b>.58/.53</b>	<b>.48/.44</b>	<b>.46/.41</b>	<b>.43/.33</b>
ESI <sub>SUS</sub>	<b>.14/.11</b>	<b>.29/.02</b>	<b>.33/.34</b>	<b>.35/.30</b>	<b>.29/.34</b>	<b>.47/.47</b>

Note. ESI<sub>DIS</sub> = General Disinhibition; ESI<sub>AGG</sub> = Callous Aggression; ESI<sub>SUS</sub> = Substance Abuse; PID-5 = Personality Inventory for DSM-5 Short Form; SDSS = Substance Dependence Severity Scale. For non-clinical sample ( $n = 742$ ),  $r_s > .13$  are significant at  $p < .001$ ; for clinical sample ( $n = 333$ ),  $r_s \geq .18$  are significant at  $p < .001$ . Values of  $r \geq .20$  are bolded.