Original Article

Translation and Validation of the BFI-2 in a Croatian Sample

Marcela Hausding[1] and Kai T. Horstmann[2]

[1]Department of Psychology, Humboldt-Universität zu Berlin, Germany
[2]Department of Psychology, University of Siegen, Germany

Abstract: This paper describes the process of translation and validation of the adaptation of the Big Five Inventory-2 (BFI-2) in the Croatian language. The translation process from English to Croatian was conducted using a forward and backward translation process. The resulting adaptation was then assessed for structural and construct validity, as well as reliability on a convenience sample of 320 Croatian participants. The results showed good reliability estimates at the domain level and somewhat lower yet satisfying estimates at the facet level. Confirmatory factor analyses (CFAs) supported the proposed hierarchical structure. The BFI-2 domains and facets showed adequate construct validity, estimated via the within- and between-domain correlations of the facet scales. Overall, the preliminary results of the Croatian adaptation are satisfactory and support efforts for further improvement and assessment in a larger sample, adequate for research in personality psychology.

Keywords: BFI-2, translation, test development, Croatia, Big Five

The Big Five model of personality consists of five broad personality dimensions commonly referred to as Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience (John & Srivastava, 1999). These five dimensions (*Big Five*; Goldberg, 1992; John et al., 2008; McCrae & Costa, 1987) facilitate an efficient description of stable individual differences in thinking, feeling, and behaving and are widely used in psychological research and application. The dimensions are the result of a lexical rather than theoretical approach, thus deriving from an analysis of language terms people use to describe themselves (John & Srivastava, 1999). The word "big" in the name emphasizes the breadth of abstraction of these dimensions with each one encompassing many specific personality traits (Costa & McCrae, 1995; John, 1990). The model provides a descriptive taxonomy for personality research, thus allowing cooperative research and integration of findings across different instruments (McCrae & John, 1992), languages, and cultures (Carlo et al., 2014).

Consequently, extensive research has been conducted using many different instruments to assess the Big Five personality traits. One of the most frequently used measures is the Big Five Inventory (BFI; John et al., 1991), a 44-item questionnaire assessing the core features of the five domains in a general population with good psychometric properties. However, since its development, the understanding of the structure of personality has advanced to a hierarchical view; accordingly, this was reflected in the instruments. The revised version of the BFI, the BFI-2 (Soto & John, 2017), therefore, contains a hierarchical structure as well.

The BFI-2 improved the original BFI along three objectives (for a detailed description, see Soto & John, 2017). First, for every Big Five domain, three facets were defined based on previous research on the hierarchical structure of personality (e.g., DeYoung et al., 2007; Goldberg, 1999). Each facet is measured with four items (i.e., 12 items per domain) that were developed with a focus on clarity, balancing the questionnaire's complexity with its psychometric quality. Second, to control for acquiescence (i.e., the tendency of a participant to agree to an item, regardless of its content; Jackson & Messick, 1958), the items are content-balanced (i.e., there is an equal number of positively and negatively keyed items) for every facet and domain. Finally, new labels for two of the domains were adopted: Negative Emotionality instead of Neuroticism (due to its clinical connotation) and Open-Mindedness instead of Openness (to clarify its reference to mental rather than social experiences and thus its distinction to Extraversion).

Altogether, the BFI-2 consists of 60 items (30 negatively keyed and 30 positively keyed) in the form of short phrases combined with (a) a synonym (e.g., "Is dependable, steady"), (b) a definition (e.g., "Is reliable, can always be counted on"), or (c) context (e.g., "Thinks poetry and plays are boring") that

https://doi.org/10.1027/2698-1866/a000045

can be completed in 10 min. Furthermore, the BFI-2 is freely available for use in research, thereby facilitating personality research around the world. The psychometric evaluation of the BFI-2 revealed good reliability of the scores at both the domain and facet levels, as well as robust factor structure (Soto & John, 2017). Also, it supported the convergent and discriminant validity of the scores by associating them with other personality inventories such as NEO PI-R (McCrae & Costa, 2010), BFAS (DeYoung et al., 2007), and Big Five mini-markers (Saucier, 1994). Moreover, compared to its precursor, the BFI-2 showed incremental power in predicting various psychological and behavioral, as well as peer-reported, criteria.

Importantly, the BFI-2 has been developed for research purposes and has been used in cross-cultural research (e.g., Gardiner et al., 2019). Therefore, since its development, various adaptations to different languages (e.g., German – Danner et al., 2019; Dutch – Denissen et al., 2019; Slovakian – Halama et al., 2020; Norwegian – Føllesdal & Soto, 2022; Spanish – Gallardo-Pujol et al., 2022) have been developed and have provided further evidence for its good psychometric properties. Many other preliminary translations, including a Croatian version, were created and are available online as a part of the International Situations Project (ISP; Baranski et al., 2020). However, this version has not yet been validated or used with a sample representing the general Croatian population. Therefore, this study aimed to develop a Croatian adaptation of the BFI-2 and assess its psychometric properties. We expect to replicate the findings of the original study (Soto & John, 2017) by demonstrating a robust multidimensional structure at the domain and facet levels, as well as strong internal consistency and between-domain discrimination.

Method

Translation of the BFI-2

We translated the BFI-2 in several steps: First, the items were translated from English by the first author of this study who is a trilingual (Croatian, English, and German) native Croatian speaker and a former US resident, in cooperation with a Croatian translator. In some instances where the items included phrases rather specific to the English language, the German adaptation (Danner et al., 2019) was used as an aid. For example, the item "Tends to find fault with others" (bfi2_agre_3_rec) can be ambiguously translated, whereas the German version of the item "Ich neige dazu, andere zu kritisieren" ("I tend to criticize others") directly corresponds to the phrase commonly used in the Croatian language "Je sklon kritiziranju drugih," and was therefore adopted as such. Nonetheless, it is important to emphasize that the German version was only used as a support in such cases as described. The goal of the translation process was to find the most literal adaptation possible for individual items while reflecting the content of the domains and facets as accurately as possible.

The second step included the comparison of the preliminary translation with the one created for the ISP (Baranski et al., 2020), with consent from its author. This version was an ad hoc version used on a student sample and had not yet been validated. Given that the target population of the BFI-2 is the general population, and considering the standard practice in BFI-2 adaptations to employ at least two independent translators (see, e.g., Danner et al., 2019; Gallardo-Pujol et al., 2022; Halama et al., 2020), we opted to regard this translation as one of the two independent translations for the current study. The differences between the translations were identified and discussed, and for each item, the more suitable version was chosen or both versions were combined. While deciding on one version or the other, we considered the nuances and connotations of the Croatian language and the potential impact of the choices on psychometric properties. For instance, to avoid using the same Croatian translation for the word "compassion" in two items: "Is compassionate, has a soft heart" (bfi2_agre_4_rec) and "sympathy" in the item "Feels little sympathy for others" (bfi2_agre_1), we decided for the word "samilost" for "compassion" and "suosjećanje" for "sympathy," as using the same word might have resulted in correlated residuals between items.

Finally, the back-translation to English was done by a Croatian PhD student in entrepreneurial studies and compared to the original version (Soto & John, 2017). Although there were no major contextual discrepancies between the original English version and the back-translated English version of our adaptation, a few modifications were made for the items where a single word could have impacted its expected psychometric difficulty. For instance, the preliminary adaptation of the item "Stays optimistic after experiencing a setback" ("Nakon neuspjeha ostaje optimističan") resulted in a back-translation of the word "neuspjeh" as "failure," rather than "setback." A strong word such as "failure" could result in a more positively skewed distribution of the item responses than the one assumed and was therefore replaced with a milder alternative (i.e., "nazadovanje"/"setback"). For a detailed description of the reasoning behind this decision and to better understand our translation process, refer to Supplementary Material 4.

As mentioned above, the aim of the adaptation was to obtain optimal balance between the closeness of the translation to the original BFI-2 and the intelligibility of the items. Consequently, we wanted to minimize any deviation from the original item formulation. Therefore, we decided

to maintain the structure of the items as uncompleted sentences (e.g., "Je otvoren, društven" ["Is outgoing, sociable"], combined with the "Ja sam netko tko…" ["I am someone who…"]) appearing at the beginning of the questionnaire and referring to each item respectively. Accordingly, we decided against the structure of completed sentences, as applied in the German version (Danner et al., 2019). That is also why we decided against the often-used alternative phrase "Ja sam osoba koja…" ["I am a person who…"]. The assumption behind the usage of the word "osoba" [person] rather than "netko" [someone] could be that *osoba* may be more inclusive, since it is of the female gender, whereas *netko* is of the male gender. However, in everyday language and contemporary literature as well as research, both expressions are used interchangeably and refer to both genders, respectively. Furthermore, we applied the same five-point rating scale as in the original BFI-2, with the labels *uopće se ne slažem* [*totally disagree*], *djelomično se ne slažem* [*somewhat disagree*], *neutralno; nemam mišljenje* [*neutral/no opinion*], *djelomično se slažem* [*somewhat agree*], and *u potpunosti se slažem* [*totally agree*].

Participants and Procedure

A convenience sample of Croatian participants was recruited through the personal network of the first author and social media websites. The participants were provided with a link to the online survey. After completion of the survey, participants were asked to forward the link to other potential participants. The participants were informed about the purpose of the study, as well as the data privacy and anonymity policy, to which they had to consent prior to taking part in the study. Additionally, the participants had to be of legal age (i.e., at least 18 years old) and of Croatian nationality. The entries that did not include responses to entire BFI-2 items were excluded from further analysis. The final sample consisted of 320 adults, ranging in age from 18 to 77 years ($M = 39.86$, $SD = 13.62$), with 61% under 35 years and 30% over 50 years of age. Two hundred nineteen participants identified as female (68%) and 101 as male (32%). Thirty-four percent of the participants graduated high school, 17% had a bachelor's degree, 33% had a master's degree, and 12% had an MD or PhD.

Measures

All participants rated themselves on the Croatian version of the BFI-2 questionnaire consisting of the translated items. Furthermore, all participants reported their age, gender, and level of education. Since we were not able to provide our participants with other incentives than feedback on their personality trait scores, we chose to keep the survey as short as possible. Therefore, no further measures were assessed in the survey.

Analyses

Overall, we conducted all analyses reported by Soto and John (2017), unless we lacked the relevant data (e.g., peer reports). Data and materials to reproduce the results of this article are available at https://osf.io/ha268/. First, we examined the descriptive and distributional properties and bivariate correlations of the Croatian BFI-2 scores at the item, facet, and domain levels. Similarly, we assessed the mean-level gender differences and reliability estimates for facet and domain scales. For the latter, we used Cronbach's α – as it is the most widely used reliability estimate and for the comparison to the estimates for the original study – and McDonald's ω – as a better estimate for hierarchical scales (e.g., Goodboy & Martin, 2020; McNeish, 2018). Most importantly, we expected to find satisfactory reliability estimates consistent with those reported for the original study and evidence for discriminant validity based on the comparison of between- and within-domain correlations.

To assess the structural properties of the Croatian BFI-2 items, we conducted confirmatory factor analysis (CFA) and principal component analysis (PCA). Given the specific assumptions on the structure of the BFI-2, as postulated by the authors of the original version, we only examined theoretically relevant models either at the facet level or at the domain level, but then modeled the facets as well; we excluded, for example, the models that Soto and John (2017) had considered, but fitted the data poorly, such as a one-factor model. Consequently, we fit a series of three different CFAs. First, the facets were modeled in individual measurement models. Here, if the model fit was not satisfactory, we added correlated residuals between same-keyed items based on modification indices. We then examined two further models that were identical to those tested in the original study (Soto & John, 2017, p. 133): First, all facets were modeled in one structural model, and in the next step, an acquiescence factor was added. Similar to Soto and John, we report CFI, TLI, BIC, RMSEA, and SRMR. However, in line with Kenny et al. (2015) and Shi et al. (2022), RMSEA was not used to evaluate model fit for models with few (i.e., 1 or 2) degrees of freedom. We expected acceptable to good model fit for each model, as well as improvement in fit for models accounting for acquiescence, thereby providing evidence for structural validity, congruent with the findings for the original version.

Finally, we conducted a series of three PCAs to further examine the structural model and test its robustness to acquiescence. The analyses were performed on (1) the 15 facet scores, where we expected to find a clear five-factor

solution since averaging item responses cancels out the acquiescence-related variance (see Soto & John, 2017, for a general description); (2) the 60 raw item scores (i.e., without controlling for acquiescence); and (3) the same item scores after within-person centering (i.e., subtracting each person's average score across all 60 items from each of their individual item responses). By doing so, the common variance of all items (i.e., acquiescence) should be removed, resulting in five factors, whereas, for the PCA on the raw item scores, we expected six factors to emerge.

Results

Descriptive Statistics and Correlations at the Item Level

The analysis of item descriptive and visual properties revealed some items to be somewhat easy (i.e., most of the participants rated themselves high on them). Those items included bfi2_agre_2, bfi2_agre_7, bfi2_agre_11, and bfi2_cons_9. These items displayed a limited answer range of three scale points (i.e., 2–5), meaning that no participant responded with *totally disagree* on these items. However, these items displayed no issues in any of the further analyses. Furthermore, when compared to the results of the descriptive analysis of the original items (Soto, 2022), some items, especially agreeableness items, seem to display extreme scores, making this a common characteristic rather than an issue specific to our adaptation. Therefore, a revision of these items, solely on the basis of high values, is not indicated. Also, considering that these items assess respect and care for others, the aspect of social desirability could explain why precisely these items scored so high.

Furthermore, interitem correlations were substantial, supported by Bartlett's test of sphericity [$\chi^2(59) = 1{,}139.4$, $p < .001$] and Kaiser–Meyer–Olkin coefficient (KMO = .84; MSAs ≥ .65), with the exception of two items: bfi2_open_1_rec and bfi2_extr_2 (MSAs ≤ .52). The descriptive statistics of the items are available in Supplementary Material 1.

Descriptive Statistics at the Facet and Domain Levels

Next, we examined the intercorrelations and descriptive statistics of the BFI-2 scales at the facet and domain levels. The results are presented in Tables 1, 2, and 3.

Twelve of 15 facet scales had a relatively high average score (≥ 3.42). The highest mean was that of the Respectfulness facet ($M = 4.23$), since two of its items (bfi2_agre_2 and bfi2_agre_11), mentioned before, were one of the highest-scoring items. Accordingly, the domain mean scores were also high [Extraversion (3.49), Agreeableness (3.89), Conscientiousness (3.90), Negative Emotionality (2.81), Open-Mindedness (3.75)], and variances constrained (*SD*s ≤ 0.95), but comparable to the descriptive statistics of the original version (Soto, 2022; Soto & John, 2017, p. 128). Table 1 also presents the mean-level gender differences. Comparing differences across groups requires at least scalar invariance across groups (Sass, 2011; Vandenberg & Lance, 2000). We report measurement invariance tests in Supplementary Material 3. Eleven of 18 tested models showed scalar invariance. We, therefore, interpret only the results for which scalar measurement invariance held. The results suggest that women tend to characterize themselves as significantly more conscientious than men. At the facet level, women reported higher scores for Compassion, Respectfulness, Organization, Responsibility, Depression, and Emotional Volatility. For those facets for which scalar measurement invariance held, no effects could be observed for Assertiveness, Trust, and Creative Imagination. These results mostly converge with those reported for the original study (Soto & John, 2017, p. 128). However, due to the size and much higher prevalence of women in this sample, the results may not illustrate the gender differences quite accurately.

Correlations at the Facet and Domain Levels

Tables 2 and 3 present interscale correlations among the BFI-2 facets and domains, respectively.

The absolute within-domain facet correlations (e.g., between Sociability and Assertiveness) ranged from .36 to .66, averaging .49 *r*-to-*z* transformation before averaging was applied, whereas absolute between-domain correlations (e.g., between Sociability and Intellectual Curiosity) ranged from .01 to .51, averaging only .19 *r*-to-*z* transformation before averaging was applied. At the domain level, absolute correlations between domains (e.g., Extraversion and Agreeableness) ranged from .16 to .39, averaging .28. Although some of the facets displayed high interdomain correlations, these results indicate overall good discriminant validity between the BFI-2 domains, as well as an appropriate level among the facets. Moreover, the average correlations are similar to those reported for the original study (Soto & John, 2017, pp. 125–126).

Reliability Estimates

Tables 2 and 3 present reliability scores estimated using the *psych* package (Revelle, 2022) in *R* (R Core Team,

Table 1. Descriptive statistics for the Croatian BFI-2 domains and facets

Domain or facet	Men *M* (*SD*)	Women *M* (*SD*)	Combined *M* (*SD*)	Gender *d*[a]
Extraversion	3.47 (0.63)	3.50 (0.62)	3.49 (0.62)	0.04[c, l, i]
Sociability	3.36 (0.84)	3.49 (0.85)	3.45 (0.85)	0.15[c, l]
Assertiveness	3.37 (0.75)	3.27 (0.75)	3.30 (0.75)	−0.14[c, l, i]
Energy level	3.69 (0.72)	3.73 (0.82)	3.72 (0.79)	0.06[c, l]
Agreeableness	3.77 (0.50)	3.94 (0.50)	3.89 (0.51)	0.33
Compassion	3.77 (0.65)	4.06 (0.62)	3.97 (0.64)	0.47[c, l, i]
Respectfulness	4.09 (0.61)	4.29 (0.55)	4.23 (0.57)	0.35[c, l, i]
Trust	3.46 (0.71)	3.46 (0.75)	3.46 (0.74)	0.01[c, l, i]
Conscientiousness	3.68 (0.69)	4.00 (0.71)	3.90 (0.72)	0.46[c, l, i]
Organization	3.59 (0.99)	3.92 (0.92)	3.82 (0.95)	0.35[c, l, i]
Productiveness	3.65 (0.76)	4.00 (0.79)	3.89 (0.80)	0.44[c, l]
Responsibility	3.79 (0.70)	4.08 (0.70)	3.99 (0.71)	0.41[c, l, i]
Negative emotionality	2.58 (0.62)	2.92 (0.67)	2.81 (0.67)	0.51[c, l]
Anxiety	3.09 (0.73)	3.57 (0.74)	3.42 (0.77)	0.66[c, l]
Depression	2.16 (0.78)	2.43 (0.85)	2.34 (0.84)	0.32[c, l, i]
Emotional volatility	2.50 (0.78)	2.75 (0.83)	2.67 (0.83)	0.31[c, l, i]
Open-mindedness	3.74 (0.57)	3.76 (0.64)	3.75 (0.62)	0.02[c, l]
Aesthetic sensitivity	3.45 (0.78)	3.73 (0.79)	3.64 (0.80)	0.36[c, l]
Intellectual curiosity	3.86 (0.72)	3.72 (0.77)	3.77 (0.75)	−0.18
Creative imagination	3.93 (0.74)	3.82 (0.89)	3.85 (0.85)	−0.13[c, l, i]
Sample size	101	219	320	

Note. Gender *d* = Cohen's *d* for the mean-level difference between men and women, with positive values indicating higher scores for women. Negatively keyed items were reverse-keyed before the computation. All entries were rounded on two decimal points. Differences of 0.31 or larger are significant at $p < .01$. [a]We tested for measurement invariance across gender (Sass, 2011; Vandenberg & Lance, 2000). The results of the measurement invariance tests can be found in the corresponding Supplementary Material 3. Models marked with *c* passed configural invariance, marked with *l* passed metric invariance, and with *i* scalar invariance.

2022) for the BFI-2 facets and domains, respectively. As shown in Table 2, Cronbach's α estimates for the 15 facet scales ranged from .53 to .86, averaging .66, whereas ω ranged from .60 to .88, averaging .73. The α values suggest somewhat lower internal consistency on the facet level compared to the original study (Soto & John, 2017, p. 126), but similar to those of other BFI-2 adaptations (e.g., Danner et al., 2019; Gallardo-Pujol et al., 2022).

As shown in Table 3, the five domain scales had Cronbach's α scores ranging between .75 and .89, averaging .81, and ω scores ranging from .66 to .85, averaging .73. Both estimates indicate acceptable to good internal consistency on the domain level, with α scores similar to those reported for the original version (p. 125).

Confirmatory Factor Analyses

Next, we examined the structure of the BFI-2 by conducting a series of three different CFA models using the *lavaan* package (Rosseel, 2012) in *R* (R Core Team, 2022). The robust maximum likelihood estimator was used to estimate parameters that are robust to non-normality. Fit statistics for these models are provided in Table 4. To evaluate model fit, we focus on the comparative fit index (CFI) and standardized root-mean-square residual (SRMR) for models with small degrees of freedom and root-mean-square error of approximation (RMSEA) for models with many degrees of freedom.

The first *single facet* model included one factor representing a single facet scale within each of the Big Five domains. This model allowed four items to load on their corresponding facet factor, respectively, and no residual variances were allowed to correlate. In contrast to the original study, this model was added to explore item and facet properties and identify possible misfits before combining them in a larger model. The *single facet* model provided a good fit (CFIs ≥ .947, SRMRs ≤ .060) for Assertiveness, Energy Level, Compassion, Trust, Organization, Productiveness, Anxiety, Depression, Emotional Volatility, Aesthetic Sensitivity, and Intellectual Curiosity. Contrarily, it

Table 2. Reliability estimates and intercorrelations of the Croatian BFI-2 facets

	Reliability		Extraversion			Agreeableness			Conscientiousness			Negative emotionality			Open-mindedness	
Facet	Alpha	Omega	Soc.	Ass.	Ene.	Com.	Res.	Tru.	Org.	Pro.	Resy.	Anx.	Dep.	Emo.	Aes.	Int.
Sociability	.71	.73	—													
Assertiveness	.53	.60	**.43**	—												
Energy level	.69	.80	**.36**	**.46**	—											
Compassion	.54	.62	.15	.03	.29	—										
Respectfulness	.62	.63	−.09	−.13	.08	**.38**	—									
Trust	.58	.79	.11	.01	.34	**.45**	**.37**	—								
Organization	.86	.88	.05	.18	.28	.17	.22	.03	—							
Productiveness	.71	.73	.12	.22	.42	.22	.30	.20	**.65**	—						
Responsibility	.70	.73	.07	.17	.32	.26	.37	.06	**.66**	**.66**	—					
Anxiety	.66	.67	−.17	−.23	−.27	−.01	−.07	−.30	−.01	−.14	−.04	—				
Depression	.69	.83	−.26	−.38	−.51	−.17	−.20	−.38	−.17	−.35	−.24	**.56**	—			
Emotional volatility	.66	.70	−.01	−.18	−.26	−.09	−.46	−.31	−.26	−.39	−.33	**.50**	**.55**	—		
Aesthetic sensitivity	.58	.78	−.03	.10	.19	.21	.09	.11	.14	.14	.14	.01	−.06	−.07	—	
Intellectual curiosity	.54	.60	.10	.34	.31	.05	−.12	.09	.07	.09	.05	−.06	−.15	−.11	**.39**	—
Creative imagination	.78	.79	.05	.40	.42	.17	.14	.15	.25	.35	.26	−.24	−.34	−.31	**.37**	**.46**
Mean	.66	.73														

Note. N = 320. Alpha = alpha coefficient based upon correlations; Omega = based upon CFA model fit; Soc. = Sociability; Ass. = Assertiveness; Ene. = Energy Level; Com. = Compassion; Res. = Respectfulness; Tru. = Trust; Org. = Organization; Pro. = Productiveness; Resy. = Responsibility; Anx. = Anxiety; Dep. = Depression; Emo = Emotional Volatility; Aes. = Aesthetic Sensitivity; Int. = Intellectual Curiosity. Within-domain correlations are in bold. Absolute correlations of .12 or stronger are significant at $p < .05$.

did not provide an acceptable fit (i.e., CFIs ≤ .901) for Sociability, Respectfulness, Responsibility, and Creative Imagination.

In most cases where the model fit was not satisfactory, modification indices suggested a substantial improvement by allowing residual variances of two same-keyed items to correlate. In the case of Respectfulness, for instance, we, therefore, allowed residual correlation between the negatively keyed items bfi2_agre_5_rec and bfi2_agre_8_rec, which led to improvement of the model fit. Since the residual correlation was positive, it is also possible to model it as an acquiescence factor with loadings fixed to one and uncorrelated with the other latent variables. Thus, in line with the original interpretation (Soto & John, 2017), it is plausible to assume that the residual correlation reflected the acquiescence factor at the facet level. We performed this for all four facets described above and attained a good fit (CFIs ≥ .988; SRMRs ≤ .036).

The second, *three facets* model, just like in the original study, included three factors representing the three facet scales within a Big Five domain. Each item was only allowed to load on a single facet factor, and the three facet factors were allowed to intercorrelate. No residual variances were allowed to correlate. Compared to the *single facet* model, this model combines *single facet* scales of each Big Five domain, respectively. For instance, Respectfulness, which was described before, was

Table 3. Reliability estimates and intercorrelations of the Croatian BFI-2 domains

	Reliability		Intercorrelations			
Domain	Alpha	Omega	Extraversion	Agreeableness	Conscientiousness	Negative emotionality
Extraversion	.79	.69	—			
Agreeableness	.75	.66	.16	—		
Conscientiousness	.89	.85	.29	.28	—	
Negative emotionality	.83	.78	−.39	−.35	−.29	—
Open-mindedness	.78	.68	.34	.17	.25	−.23
Mean	.81	.73				

Note. N = 320. All absolute correlations are significant at $p < .01$.

Table 4. Fit statistics for CFAs of the Croatian BFI-2 items

Model	χ^2	*df*	CFI	TLI	BIC	RMSEA[b] [90% CI]	SRMR
Extraversion							
Single facet							
Sociability	35.32	2	.884	0.653	3,809.69	.211 [.153, .274]	**.059**
Sociability[a]	18.37	1	**.953**	0.715	3,797.65	.191 [.121, .272]	**.036**
Assertiveness	0.80	2	**1.000**	**1.042**	3,922.26	.000 [.000, .088]	**.013**
Energy level	1.24	2	**1.000**	**1.010**	3,655.16	.000 [.000, .100]	**.014**
Three facets	139.98	51	.880	0.845	11,246.61	**.078** [.062, .093]	.064
Three facets plus acquiescence	114.49	50	**.914**	0.886	11,223.34	**.066** [.050, .082]	.061
Agreeableness							
Single facet							
Compassion	0.90	2	**1.000**	**1.038**	3,442.28	.000 [.000, .087]	**.014**
Respectfulness	55.06	2	.853	0.560	2,899.64	.184 [.144, .228]	**.060**
Respectfulness[a]	2.85	1	**.988**	**0.929**	2,885.53	.074 [.000. .181]	**.019**
Trust	1.18	2	**1.000**	**1.019**	3,806.28	.000 [.000, .095]	**.016**
Three facets	184.26	51	.767	0.699	10,027.34	.094 [.080, .109]	.079
Three facets plus acquiescence	135.55	50	.850	0.803	9,980.25	**.076** [.061, .092]	.068
Conscientiousness							
Single facet							
Organization	8.21	2	**.984**	**0.951**	3,422.10	.122 [.044, .214]	**.023**
Productiveness	0.89	2	**1.000**	**1.016**	3,622.90	.000 [.000, .091]	**.011**
Responsibility	103.64	2	.767	0.301	3,312.41	.310 [.261, .363]	.101
Responsibility[a]	0.80	1	**1.000**	**1.004**	3,256.01	.000 [.000, .131]	**.007**
Three facets	179.44	51	.893	0.862	9,986.50	.101 [.085, .117]	.066
Three facets plus acquiescence	138.93	50	**.927**	**0.904**	9,937.36	.084 [.068, .101]	.059
Negative emotionality							
Single facet							
Anxiety	2.56	2	**.996**	**0.989**	3,704.67	.031 [.000, .124]	**.020**
Depression	2.25	2	**.999**	**0.997**	3,756.86	.022 [.000, .127]	**.016**
Emotional volatility	2.66	2	**.996**	**0.988**	3,902.80	.033 [.000, .124]	**.021**
Three facets	117.98	51	**.919**	0.895	11,115.57	**.068** [.052, .085]	.056
Three facets plus acquiescence	103.01	50	**.936**	**0.915**	11,104.49	**.062** [.045, .078]	.052
Open-mindedness							
Single facet							
Aesthetic sensitivity	1.71	2	**1.000**	**1.004**	3,914.02	.000 [.000, .107]	**.018**
Intellectual curiosity	6.64	2	**.947**	0.841	3,943.81	.087 [.019, .165]	**.035**
Creative imagination	53.05	2	**.901**	0.703	3,535.82	.236 [.184, .293]	**.060**
Creative imagination[a]	1.02	1	**1.000**	**1.000**	3,505.67	.009 [.000, .161]	**.009**
Three facets	119.77	51	**.909**	0.882	11,234.47	**.070** [.054, .086]	.054
Three facets plus acquiescence	106.60	50	**.925**	**0.901**	11,225.75	**.064** [.047, .081]	.053

Note. $N = 320$. *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; BIC = Bayesian information criterion; RMSEA [90% CI] = root-mean-square error of approximation with 90% confidence intervals; SRMR = standardized root-mean-square residual. Table entries correspond to the robust values of the MLR estimator output. CFI and TLI values ≥ .900, and RMSEA ≤ .080 and SRMR ≤ .060, are in bold for models with many or few degrees of freedom, respectively. [a]Model in which residual correlations between two same-keyed items were allowed to improve the model fit. [b]RMSEA is reported but not interpreted to evaluate models with few (i.e., 1 or 2) degrees of freedom (Kenny et al., 2015; Shi et al., 2022)

Table 5. Loadings from a principal components analysis of the Croatian BFI-2 facets

Facet	Extraversion	Agreeableness	Conscientiousness	Negative emotionality	Open-mindedness
Extraversion					
Sociability	**.63**	.08	.02	−.05	−.04
Assertiveness	**.61**	−.15	.14	−.20	.31
Energy level	**.54**	.24	.26	−.25	.31
Agreeableness					
Compassion	.12	**.72**	.17	.07	.12
Respectfulness	−.29	**.52**	.33	−.26	−.04
Trust	.09	**.64**	−.03	−.33	.08
Conscientiousness					
Organization	.07	.02	**.79**	−.02	.10
Productiveness	.15	.14	**.75**	−.20	.12
Responsibility	.05	.14	**.82**	−.07	.08
Negative emotionality					
Anxiety	−.20	−.04	.04	**.69**	−.03
Depression	−.37	−.19	−.16	**.68**	−.11
Emotional volatility	.11	−.13	−.31	**.77**	−.10
Open-mindedness					
Aesthetic sensitivity	−.05	.16	.09	.05	**.57**
Intellectual curiosity	.18	−.05	−.02	−.05	**.70**
Creative imagination	.14	.06	.23	−.26	**.64**

Note. N = 320. Facet loadings on varimax-rotated principal components. The strongest loading for each facet is in bold.

combined with Compassion and Trust from the *single facet* model to build the three-facet structure of the Agreeableness domain. This model provided an acceptable or near-acceptable overall fit for Extraversion (CFI = .880, RMSEA = .078), Negative Emotionality (CFI = .919, RMSEA = .068), and Open-Mindedness (CFI = .909, RMSEA = .070), while the fit for Agreeableness (CFI = .767, RMSEA = .094), and Conscientiousness (CFI = .893, RMSEA = .101) was unacceptable. Nevertheless, the results were similar to those of the original study (p. 133).

The final *three facets plus acquiescence* model added an acquiescence factor to the *three facets* model just described. Thus, each item was allowed to load on both its facet factor and an acquiescence method factor. All loadings on the acquiescence factor were constrained to one, and the acquiescence factor was not allowed to correlate with any of the facet factors. Such constraints ensure the distinction between acquiescence and meaningful personality content (Soto & John, 2017). According to our expectations, compared with the previous model, this model provided an improvement in fit for each Big Five domain (ΔCFIs $\geq$ .016, ΔRMSEAs $\geq$ .006), further supporting our assumption made about the influence of acquiescence on the facet level. Moreover, it provided an acceptable overall fit for Extraversion (CFI = .914, RMSEA = .066), Negative Emotionality (CFI = .936, RMSEA = .062), Open-Mindedness (CFI = .925, RMSEA = .064), and Conscientiousness (CFI = .927, RMSEA = .084), as well as a near-acceptable fit for Agreeableness (CFI = .850, RMSEA = .076). The reason for the unacceptable fit for Agreeableness seems to lie in the items bfi2_agre_3_rec and item bfi2_agre_6. Both items belong to the Trust facet but displayed a secondary loading on the Respectfulness facet. Nevertheless, compared with the results from the original study, the model fit for all domains was very similar, with small discrepancies depending on which fit indices one focuses on. Moreover, the same pattern of improvement from the model without the acquiescence factor emerged.

Principal Component Analyses

To examine the viability of using reverse-keyed items to deal with acquiescence, a series of three principal component analyses (PCA) was conducted. The first PCA was performed on the 15 facet scores (calculated by averaging raw item responses) of the Big Five domains.

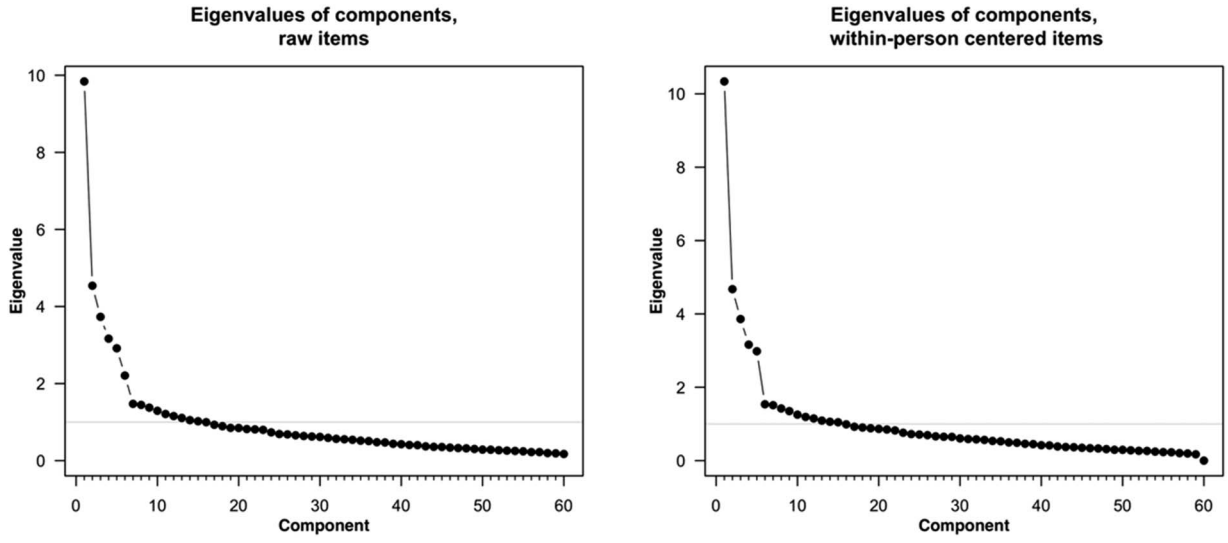

Figure 1. Eigenvalues from PCAs of the raw and within-person-centered Croatian BFI-2 items.

Conforming with the assumption that content-balanced facet scales control for acquiescence (Soto & John, 2017), the parallel analysis suggested five components. Therefore, we extracted five varimax-rotated components from the facet scales, which, as expected, yielded a clear five-factor structure. The loadings from this analysis are presented in Table 5. All facet scales had their primary loadings on the intended component. Also, just like reported for the original version, the highest secondary loadings were those of Depression on Extraversion and of Trust on Negative Emotionality, which hold conceptually meaningful associations. The extracted components could account for 57% of the observed variance.

To test for the effects of acquiescence at the item level, a PCA was conducted on the 60 raw items of our BFI-2 adaptation. The results were then compared to those of the second PCA, performed on the same items as the previous one, after within-person centering. As expected, the parallel analysis of the items without within-person centering suggested six components, indicating an additional acquiescence component. In contrast, the same analysis on the centered items proposed five components, thereby implying an effective elimination of the acquiescence component (see Figure 1).

We then extracted five varimax-rotated components from both the raw and centered items. The loadings of the PCA on within-centered items are presented in Table 6.

Both analyses yielded a clear Big Five structure with most of the items (55 in the PCA of the raw and 54 of the centered items) having their primary loadings on the intended component (the factor loadings of the uncentered items are available in the Online Supplement). The polarity of all the loadings was analogous to that reported for the original version (pp. 130–131), as well as the item-keying (i.e., given the positive loadings of the positive-keyed items, the negative-keyed items had a negative loading and vice versa). Items bfi2_open_1_rec and bfi2_extr_2 had very low loadings altogether ($\lambda \leq .18$), which was not surprising given their insufficient MSA coefficients. For the remaining items that did not display the expected pattern of loading, the difference between the primary and secondary loadings (i.e., the loadings on the intended component) was small [Δs (.01–.08)]. The extracted components could account for 42% of the item variance. Taken together, the results support the underlying assumption (Soto & John, 2017) that the within-centering of the items eliminates some of the acquiescence variance.

Discussion

The present study provides an adaptation of the BFI-2 in the Croatian language and assesses its psychometric properties. The translation process consisted of three separate steps: (1) primary translation of the original BFI-2 (Soto & John, 2017), using the German adaptation (Danner et al., 2019) as an aid; (2) comparison and modification in relation to the Croatian translation created for ISP (Baranski et al., 2020); and (3) back-translation and further modification of the items. Every single item of our adaptation was extensively discussed and deliberated, with valuable input from a professional

Table 6. Loadings from a principal components analysis of the Croatian BFI-2 items

Item, domain, facet	Item text	E	A	C	N	O
Extraversion						
Sociability						
bfi2_extr_4_rec	Tends to be quiet	−**.72**	.13	.03	.12	.07
bfi2_extr_10	Is talkative	**.64**	.16	−.03	.00	−.06
bfi2_extr_7_rec	Is sometimes shy, introverted	−**.62**	.04	−.09	−.32	−.04
bfi2_extr_1	Is outgoing, sociable	**.53**	.46	.10	.19	−.13
Assertiveness						
bfi2_extr_5	Is dominant, acts as a leader	**.63**	−.28	.08	.16	.14
bfi2_extr_8_rec	Finds it hard to influence people	−**.41**	−.09	−.14	−.13	−.34
bfi2_extr_11_rec	Prefers to have others take charge	−.37	.16	−.19	−.28	−**.38**
bfi2_extr_2	Has an assertive personality	.13	−.11	.05	**.15**	.11
Energy level						
bfi2_extr_12	Shows a lot of enthusiasm	**.42**	.30	.22	.37	.31
bfi2_extr_3_rec	Rarely feels excited or eager	−**.40**	−.20	−.03	−.01	−.20
bfi2_extr_9	Is full of energy	.39	.16	.33	**.44**	.22
bfi2_extr_6_rec	Is less active than other people	−.29	−.09	−.30	−.11	−**.35**
Agreeableness						
Compassion						
bfi2_agre_1	Is compassionate, has a soft heart	.00	**.65**	.12	−.12	−.01
bfi2_agre_10_rec	Can be cold and uncaring	−.21	−**.59**	−.03	−.02	.07
bfi2_agre_7	Is helpful and unselfish with others	.11	**.48**	.37	.06	.13
bfi2_agre_4_rec	Feels little sympathy for others	.01	−**.41**	−.08	.09	−.28
Respectfulness						
bfi2_agre_2	Is respectful, treats others with respect	−.08	**.47**	.34	.06	.05
bfi2_agre_8_rec	Is sometimes rude to others	.32	−**.47**	−.15	−.20	.13
bfi2_agre_11	Is polite, courteous to others	−.18	**.39**	**.39**	.04	.08
bfi2_agre_5_rec	Starts arguments with others	**.40**	−.35	−.22	−.25	.03
Trust						
bfi2_agre_12	Assumes the best about people	.20	**.61**	−.04	.28	.08
bfi2_agre_6	Has a forgiving nature	−.10	**.55**	−.05	.08	.05
bfi2_agre_3_rec	Tends to find fault with others	.10	−**.45**	−.08	−.26	−.09
bfi2_agre_9_rec	Is suspicious of others' intentions	−.10	−.36	.11	−**.44**	−.02
Conscientiousness						
Organization						
bfi2_cons_7	Keeps things neat and tidy	.06	.04	**.78**	−.05	.03
bfi2_cons_1_rec	Tends to be disorganized	−.02	.12	−**.76**	−.11	−.11
bfi2_cons_4	Is systematic, likes to keep things in order	.05	−.06	**.74**	−.07	.07
bfi2_cons_10_rec	Leaves a mess, doesn't clean up	−.01	−.09	−**.69**	.07	−.05
Productiveness						
bfi2_cons_2_rec	Tends to be lazy	−.04	−.07	−**.69**	−.10	−.04
bfi2_cons_11	Is persistent, works until the task is finished	−.04	.09	**.64**	.19	.09
bfi2_cons_5_rec	Has difficulty getting started on tasks	−.09	−.17	−**.54**	−.14	−.06
bfi2_cons_8	Is efficient, gets things done	.14	.02	**.54**	.22	.22

(Continued on next page)

Table 6. (Continued)

Item, domain, facet	Item text	E	A	C	N	O
Responsibility						
bfi2_cons_6_rec	Can be somewhat careless	−.05	.06	**−.74**	.03	.00
bfi2_cons_12_rec	Sometimes behaves irresponsibly	.01	−.08	**−.64**	.00	−.01
bfi2_cons_3	Is dependable, steady	−.02	.23	**.58**	.20	.12
bfi2_cons_9	Is reliable, can always be counted on	.04	.29	**.57**	.05	.12
Negative emotionality						
Anxiety						
bfi2_neur_1_rec	Is relaxed, handles stress well	.01	−.01	.07	**.67**	.12
bfi2_neur_10_rec	Rarely feels anxious or afraid	.16	−.01	.11	**.59**	−.02
bfi2_neur_7	Worries a lot	−.05	.14	.23	**−.59**	−.07
bfi2_neur_4	Can be tense	.02	−.15	.14	**−.54**	.06
Depression						
bfi2_neur_8	Often feels sad	−.25	−.12	−.12	**−.67**	−.05
bfi2_neur_11	Tends to feel depressed, blue	−.25	−.17	−.19	**−.65**	−.07
bfi2_neur_2_rec	Stays optimistic after experiencing a setback	.17	.27	−.05	**.51**	.16
bfi2_neur_5_rec	Feels secure, comfortable with self	.04	−.12	.30	**.41**	.14
Emotional volatility						
bfi2_neur_6_rec	Is emotionally stable, not easily upset	−.05	.09	.13	**.72**	.14
bfi2_neur_9_rec	Keeps their emotions under control	−.17	.03	.39	**.49**	.10
bfi2_neur_12	Is temperamental, gets emotional easily	.35	−.06	−.11	**−.48**	−.01
bfi2_neur_3	Is moody, has up and down mood swings	.05	−.13	−.34	**−.47**	−.07
Open-mindedness						
Aesthetic sensitivity						
bfi2_open_4	Is fascinated by art, music, or literature	−.04	.07	.00	−.04	**.68**
bfi2_open_7	Values art and beauty	−.01	.27	.13	−.13	**.64**
bfi2_open_10_rec	Thinks poetry and plays are boring	−.04	−.26	−.12	.06	**−.38**
bfi2_open_1_rec	Has few artistic interests	**.18**	−.03	.05	.15	**−.18**
Intellectual curiosity						
bfi2_open_11_rec	Has little interest in abstract ideas	−.12	−.05	.05	−.13	**−.58**
bfi2_open_2	Is curious about many different things	.01	−.03	.09	.17	**.55**
bfi2_open_5_rec	Avoids intellectual, philosophical discussions	−.30	.10	.12	−.04	**−.45**
bfi2_open_8	Is complex, a deep thinker	.00	−.11	.03	−.30	**.43**
Creative imagination						
bfi2_open_6_rec	Has little creativity	.09	.03	−.12	−.13	**−.67**
bfi2_open_9_rec	Has difficulty imagining things	−.03	−.10	−.16	−.18	**−.66**
bfi2_open_12	Is original, comes up with new ideas	.12	−.02	.22	.30	**.62**
bfi2_open_3	Is inventive, finds clever ways to do things	.00	−.01	.29	.28	**.49**

Note. N = 320. Loadings of within-person centered items on varimax-rotated principal components. E = Extraversion; A = Agreeableness; C = Conscientiousness; N = Negative Emotionality; O = Open-Mindedness. The strongest loading for each item is in bold.

translator. The resulting adaptation of the BFI-2 was then tested on 320 adult Croatian participants. The collected data were subsequently used to estimate the psychometric quality through different analyses that were for the most part identical to those conducted for the original English version.

Item Characteristics

The descriptive analyses of items yielded no noticeable discrepancies that were relevant for the subsequent analyses. The examination of the item correlations indicated a substantive level of interitem correlation

(KMO = .84; MSAs ≥ .65) for all but two items (bfi2_open_1_rec and bfi2_extr_2), which consequently yielded low PCA loadings (λs ≤ .18). The cause of the poor properties of these items could be contextual. Our effort to deliver the most literate translation possible could have come at the expense of the appropriate item wording for the general population and thus unambiguous interpretation across individuals. This may especially apply to the item bfi2_extr_2 "Ima asertivnu osobnost" ["Has an assertive personality"] given it consists of two words (i.e., assertive and personality) that are essentially technical terms rather than words commonly used. It is therefore possible that this item was incomprehensible to some of our participants. If so, the variance of this item would not capture the individual differences in assertiveness but rather something else that is not substantial and could henceforth not covariate with the rest of the items. We, therefore, propose an initial, yet unvalidated, alternative translation that addresses these issues. Since there is no single Croatian word for "assertiveness," other than the anglicism (i.e., "asertivnost") already used in this item, we propose an item consisting of two adjectives that combined capture the essence of this trait: "Je samosvjestan I samouvjeren," which translates to "Is self-conscious and self-confident."

On the other hand, the issue with the item bfi2_open_1_rec "Ima malo umjetničkih interesa" ["Has few artistic interests"] is somewhat less apparent. Once again, the focus on the fidelity to the original item may have resulted in a slightly uncommon phrasing of the item and thus possible confusion. To preserve the term "artistic interests" rather than "interest in art," we were bound to use the word "malo" ["little"] as the only possible translation of the word "few," thereby possibly failing to convey its intended meaning ("some but not many") and thus created additional interpretation related variance. We, therefore, propose an alternative, yet unvalidated, translation that may be more appropriate: "Se ne zanima pretjerano za umjetnost" ["Is not overly interested in art"]. It is important to emphasize that regardless of the results that suggest poor properties of these items and alternative translations provided, it does not mean that these results are not sample-specific and that these items are fundamentally inadequate. For instance, it is likely that in a sample of psychology students, the item bfi2_extr_2 would perform better than the alternative, given that such a sample would be familiar with the terminology (i.e., assertive personality). We, therefore, recommend using both versions for future data collection and comparing the results for a more definite appraisal.

We then examined the descriptive statistics and gender differences of the Croatian BFI-2 facet and domain scores. The results for the complete sample showed that most facets and domains had relatively high scores paired with somewhat limited variance. We tested measurement invariance across gender to examine mean differences in the facet and domain scores. Comparing means across groups was possible for 11 of 20 tested models, as for these models, scalar invariance held (Sass, 2011; Vandenberg & Lance, 2000). Overall, we observed gender differences that are in line with previous findings reported for the original BFI-2 (Soto & John, 2017, p. 128).

Reliability

The reliability estimates of the scales were, as already reported, satisfactory. Although some facets had comparatively low internal consistency estimates, the reliability of the domains was good. Note, however, that the internal consistency estimates are based on four items only and that the test–retest correlations may be higher. Moreover, a similar pattern could be observed for the German (Danner et al., 2019), Norwegian (Føllesdal & Soto, 2022), and Slovak (Halama et al., 2020) and Spanish (Gallardo-Pujol et al., 2022) adaptations, making it a common difficulty of cross-cultural adaptation of such questionnaires, rather than an issue specific to this adaptation. The scales may therefore be used for research purposes and the examination and comparison of groups, although they are most likely insufficiently reliable for individual-level analyses (Emons et al., 2007).

Structural Validity

The results of the series of CFAs provided evidence for the goodness of the Big Five model as the underlying theoretical structure of the data. This means that the translated version of the BFI-2 assesses five factors with three facets each. Although some fit indices were lower compared to the traditional standards by Hu and Bentler (1999), they were nevertheless comparable to the results provided for the original version (Soto & John, 2017, p. 133) and in line with expectations when fitting complex models to personality self-report data (Hopwood & Donnellan, 2010). Two items, bfi2_agre_3_rec "Je sklon kritiziranju drugih" ["Tends to find fault with others"] and bfi2_agre_6 "Je pomirljive naravi, lako oprašta" ["Has a forgiving nature"], had cross-loadings on other factors: Both items belong to the Trust facet but displayed a secondary loading on the Respectfulness facet.

Although post hoc explanations are possible, it is essentially unclear why these cross-loadings emerged. Yet, we believe this neither prevents one from using the facet scores (given the good model fit of both facets, Trust and Respectfulness), nor the domain score of Agreeableness, to which both items belong. However, future applications of this adaption may inspect if this also occurs in other samples, which in the long run, might call for item revision.

Acquiescence

The results of the principal component analyses provide further evidence for the robust five-factor structure of the Croatian BFI-2 items and the different ways of dealing with acquiescence. For item-level analyses, the within-centering of the items seems adequate. For facet-level analyses, the items can be averaged to form a composite reflecting the facet scores. Finally, the CFAs also show that including the acquiescence factor increases model fit (as do correlated residuals of the same-keyed items in measurement models). Overall, this series of analyses show that acquiescence occurs in the data but can effectively be dealt with using the approaches implemented by Soto and John (2017).

Limitations and Future Directions

Importantly, this study does not provide any convergent or discriminant validity evidence with respect to other (personality) self-reports or other modes of assessment (e.g., informant reports). However, the facets of the same domain (i.e., within-domain) correlated, on average, higher among each other than they did with facets of other domains (i.e., between-domain), which provides some evidence of discriminant validity. Furthermore, the factor structure emerged as expected, providing evidence for structural validity. Finally, the gender differences, as far as they could be interpreted, were also in line with previous research, which further corroborates our conclusion that the adaptation was successful. Therefore, it can safely be concluded that this translated version captures the Big Five personality traits reasonably well. Nevertheless, the expansion of the nomological net of these traits by adding criterion measures (e.g., important life outcomes, such as income, well-being, health) or further convergent and discriminant measures (e.g., other personality inventories in Croatian, such as IPIP Big Five markers [Mlačić & Goldberg, 2007] and peer reports) is desirable in the future research. This study allows for this by providing a well-translated and validated version of the BFI-2, including open data and open materials for further analysis. As the scale is increasingly used and applied, the evidence regarding its usefulness and practicality will continue to accumulate. We anticipate that this open access version will expedite this process and facilitate wider adoption.

Conclusion

The present study reports the process of translation of a Croatian adaptation of the BFI-2 and the evaluation of its psychometric properties. Based on the results of descriptive and correlational analysis, principal component, and CFA, as well as reliability estimates, this study provides evidence for the structural, construct validity, and internal consistency of the scores consistent with those reported for the original English version and the Big Five factor model. Thus, the current adaptation of the BFI-2 along with its corresponding open access materials and data could be useful and suitable for research in the Croatian population and international, cross-cultural research.

References

Baranski, E., Funder, D. C., Gardiner, G., Sauerberger, K. S., & Lee, D. I. (2020). *The International Situations Project*. https://osf.io/yv2nq/

Carlo, G., Knight, G. P., Roesch, S. C., Opal, D., & Davis, A. (2014). Personality across cultures: A critical analysis of Big Five research and current directions. In F. T. L. Leong, L. Comas-Díaz, G. C. Nagayama Hall, V. C. McLoyd, & J. E. Trimble (Eds.), *APA handbook of multicultural psychology, Vol. 1: Theory and research* (pp. 285–298). American Psychological Association. https://doi.org/10.1037/14189-015

Costa, P. T., Jr., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, *64*(1), 21–50. https://doi.org/10.1207/s15327752jpa6401_2

Danner, D., Rammstedt, B., Bluemke, M., Lechner, C., Berres, S., Knopf, T., Soto, C. J., & John, O. P. (2019). Das Big Five Inventar 2 [The Big Five Inventory-2]. *Diagnostica*, *65*(3), 121–132. https://doi.org/10.1026/0012-1924/a000218

Denissen, J. J. A., Geenen, R., Soto, C. J., John, O. P., & van Aken, M. A. G. (2019). The Big Five Inventory-2: Replication of psychometric properties in a Dutch adaptation and first evidence for the discriminant predictive validity of the facet scales. *Journal of Personality Assessment*, *102*(3), 309–324. https://doi.org/10.1080/00223891.2018.1539004

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880–896. https://doi.org/10.1037/0022-3514.93.5.880

Emons, W. H. M., Sijtsma, K., & Meijer, R. R. (2007). On the consistency of individual classification using short scales. *Psychological Methods*, *12*(1), 105–120. https://doi.org/10.1037/1082-989x.12.1.105

Føllesdal, H., & Soto, C. J. (2022). The Norwegian adaptation of the Big Five Inventory-2. *Frontiers in Psychology*, *13*, Article 858920. https://doi.org/10.3389/fpsyg.2022.858920

Gallardo-Pujol, D., Rouco, V., Cortijos-Bernabeu, A., Oceja, L., Soto, C. J., & John, O. P. (2022). Factor structure, gender invariance, measurement properties, and short forms of the Spanish adaptation of the Big Five Inventory-2. *Psychological Test Adaptation and Development*, *3*(1), 44–69. https://doi.org/10.1027/2698-1866/a000020

Gardiner, G., Sauerberger, K., Funder, D., Rizwan, M., & International Situations Project. (2019). Towards meaningful comparisons of personality in large-scale cross-cultural studies. In A. Realo (Ed.), *In praise of an inquisitive mind: A Festschrift in honor of Jüri Allik on the occasion of his 70th birthday* (pp. 123–139). University of Estonia Press.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. *7*, pp. 7–28). Tilburg University Press.

Goodboy, A. K., & Martin, M. M. (2020). Omega over alpha for reliability estimation of unidimensional communication measures. *Annals of the International Communication Association*, *44*(4), 422–439. https://doi.org/10.1080/23808985.2020.1846135

Halama, P., Kohút, M., Soto, C. J., & John, O. P. (2020). Slovak adaptation of the Big Five Inventory (BFI-2): Psychometric properties and initial validation. *Studia Psychologica*, *62*(1), 74–87. https://doi.org/10.31577/sp.2020.01.792

Hausding, M., & Horstmann, K. T. (2023). *Translation and validation of the BFI-2 in a Croatian sample* [Data set]. https://osf.io/ha268/

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, *14*(3), 332–346. https://doi.org/10.1177/1088868310361240

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, *55*(4), 243–252. https://doi.org/10.1037/h0045996

John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). Guilford Press.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory–Versions 4a and 54*. University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big-Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). Guilford Press.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). Guilford Press.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90. https://doi.org/10.1037/0022-3514.52.1.81

McCrae, R. R., & Costa, P. T. (2010). *NEO inventories professional manual*. Psychological Assessment Resources.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Mlačić, B., & Goldberg, L. R. (2007). An analysis of a cross-cultural personality inventory: The IPIP Big-Five factor markers in Croatia. *Journal of Personality Assessment*, *88*(2), 168–177. https://doi.org/10.1080/00223890701267993

R Core Team. (2022). *R: A language and environment for statistical computing [Computer program]*. R Foundation for Statistical Computing.

Revelle, W. (2022). *psych: Procedures for psychological, psychometric, and personality research (R package version 2.2.9)* [Computer software]. Northwestern University. https://CRAN.R-project.org/package=psych

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

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, *29*(4), 347–363. https://doi.org/10.1177/0734282911406661

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar big-five markers. *Journal of Personality Assessment*, *63*(3), 506–516. https://doi.org/10.1207/s15327752jpa6303_8

Shi, D., DiStefano, C., Maydeu-Olivares, A., & Lee, T. (2022). Evaluating SEM model fit with small degrees of freedom. *Multivariate Behavioral Research*, *57*(2–3), 179–207. https://doi.org/10.1080/00273171.2020.1868965

Soto, C. J. (2022). *Descriptive statistics for the BFI-2*. http://www.colby.edu/psych/personality-lab/

Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology*, *113*(1), 117–143. https://doi.org/10.1037/pspp0000096

Vandenberg, R. J., & Lance, C. E. (2000). A Review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

History
Received November 25, 2022
Revision received May 17, 2023
Accepted May 24, 2023
Published online July 14, 2023

Section: Personality

Acknowledgments
We thank Christopher J. Soto (Colby College) for the helpful advice regarding original item content and target interpretation in the translation process as well as the choice of analyses. We thank Prof. Željko Jerneić (University of Zagreb) for the permission to use his translated version of the BFI-2 for our research. We thank Borna Buljan (University of Zadar) for providing us with a back-translation of the Croatian BFI-2 adaptation into English. We also

thank Dubravka Marić for her professional input and support in the translation process. Any errors that remain are solely the authors' responsibility.

Open Science

Open Data: The information needed to reproduce all of the reported results is available at https://osf.io/ha268/ (Hausding & Horstmann, 2023).

Open Materials: The information needed to reproduce all of the reported methodology is available at https://osf.io/ha268/ (Hausding & Horstmann, 2023).

The online supplementary materials are available at https://osf.io/ha268/ (Hausding & Horstmann, 2023).

Preregistration and Analysis Plan: This study was not pre-registered.

ORCID

Marcela Hausding
https://orcid.org/0000-0001-9539-4337
Kai T. Horstmann
https://orcid.org/0000-0003-3224-1880

Kai T. Horstmann

Department of Psychology
Faculty V: School of Life Sciences
University of Siegen
Adolf-Reichwein-Str. 2a
57076 Siegen
Germany
kai.horstmann@uni-siegen.de