Multistudy Report

Validation of the Short and Extra-Short Forms of the Big Five Inventory-2 (BFI-2) and Their German Adaptations

Beatrice Rammstedt,[1] Daniel Danner,[2] Christopher J. Soto,[3] and Oliver P. John[4]

[1]GESIS – Leibniz Institute for the Social Sciences, Mannheim, Germany
[2]University of Applied Labor Studies, Mannheim, Germany
[3]Colby College, Waterville, MN, USA
[4]University of California, Berkeley, CA, USA

Abstract: The present study investigates the validity and utility of the German adaptations of the two short forms of the Big Five Inventory-2 (BFI-2), the 30-item BFI-2-S, and the 15-item BFI-2-XS, developed by Soto and John (2017b). Both scales assess the Big Five domains. The BFI-2-S allows, in addition, the brief assessment of three facets per domain. Based on a large and heterogeneous sample, we show that the psychometric properties of these adapted short scales are consistent with those of the Anglo-American source versions, and we demonstrate substantial convergence between the adaptations and the source versions. Extending the original scale development study, we demonstrate high retest stability of the scales and their facets. Our results clearly indicate the construct and criterion validity of the two scales: Both show substantial convergence with the NEO-PI-R domain scales. Moreover, the distinctive correlation pattern found between the facets of the BFI-2 and the NEO-PI-R could be replicated for the facets of the BFI-2-S. Furthermore, we show that the domain scales of both instruments are associated in the hypothesized directions with important life outcomes, such as life satisfaction and intelligence, and that the facets of the BFI-2-S have incremental validity for predicting these outcomes.

Keywords: Big Five Inventory-2, Big Five, short forms, validation, personality measurement

In the course of the last half a century, consensus has grown among personality researchers that a person's personality can be described on the most global level in terms of the "Big Five" or "Five-Factor Model" framework (Goldberg, 1981; John, Naumann, & Soto, 2008).[1] According to this framework, personality can be summarized along five independent and bipolar dimensions – namely, Extraversion, Agreeableness, Conscientiousness, Negative Emotionality (or Neuroticism), and Open-Mindedness (or Openness to Experience). Although different researchers have sometimes interpreted these trait dimensions in somewhat different ways – such as defining the fifth factor as Openness within the questionnaire-based Five-Factor Model tradition (Costa and McCrae, 1992), versus defining it as Culture or Intellect within the lexical Big Five tradition (Goldberg, 1990) – the general agreement within the field of personality psychology on a common framework to measure personality has made personality assessment more attractive for a wide range of researchers and contexts outside the narrow field of personality research.

Due to this agreement on a common framework, measures of personality are increasingly considered to be useful tools in applied settings, including education, health, industrial psychology, and economics. For example, influential national surveys, such as the German Socio-Economic Panel (SOEP), the German National Educational Panel Study (NEPS), the Household, Income and Labour Dynamics in Australia (HILDA) Survey, the GESIS Panel, and the UK Household Longitudinal Study (UKHLS); and major international surveys, such as the World Values Survey

---

[1] Historically, the term "Big Five" has been associated with psycholexical research examining personality-descriptive terms in natural language, whereas the term "Five-Factor Model" has been associated with research examining the content and structure of traditional personality questionnaires (for a review, see John et al., 2008). In the interests of simplicity and readability, we generally refer to the "Big Five" throughout this paper, while acknowledging that the connotations of the "Big Five" and the "Five-Factor Model" differ somewhat.

(WVS) and the International Social Survey Programme (ISSP), have included measures of personality in their core questionnaires. Moreover, all of these surveys follow the Big Five approach when assessing personality.

The growing spread of the Big Five into diverse contexts, and especially the widespread assessment of the Big Five dimensions in large-scale surveys, has increased the need for efficient instruments for their measurement. To meet this need, several short-scale measures have been developed over the last decade. For example, Rammstedt and John developed the BFI-10 (Rammstedt & John, 2007) and the BFI-K (Rammstedt & John, 2005), 10-item and 21-item abbreviated versions, respectively, of the established Big Five Inventory (BFI; John, Donahue, & Kentle, 1991). In the US context, Gosling and colleagues developed a 5-item and a 10-item personality inventory (TIPI; Gosling, Rentfrow, & Swann, 2003; for a critical review of the German adaptation see Herzberg & Brähler, 2006), and Donnellan and colleagues developed a 20-item measure from the International Personality Item Pool (Mini-IPIP; Donnellan, Oswald, Baird, & Lucas, 2006). All of these ultrashort measures are widely used in a variety of contexts that usually suffer from strict time limitations not allowing for a full-scale assessment of personality. However, none of them have directly addressed a key issue: how to ensure that a very brief scale will adequately sample the heterogeneous content of a construct as broad as one of the Big Five trait domains. In fact, some brief Big Five measures have been developed using criteria, such as maximizing the size and simplicity of factor loadings, that promote the construction of scales with a narrow rather than broad range of item content (Smith, McCarthy, & Anderson, 2000).

In recent years, researchers have placed increasing emphasis on examining not only the global Big Five dimensions but also the more specific facets of these broad domains. As shown by Paunonen and Ashton (2001), taking facets into account – in addition to the global Big Five domains – can incrementally predict important personal, academic, social, and health outcomes, such as physical activity or the grade point average in college. Moreover, studies have shown that the facets of a domain can have differential predictive power for different outcome variables. For example, Roberts, Chernyshenko, Stark, and Goldberg (2005) showed that, compared to the Conscientiousness domain scale, specific facets of this domain substantially increase the predictive validity for self-reported drug consumption and health prevention.

Only very few, and comparatively lengthy, instruments for the assessment of these more fine-grained facets of the Big Five domains have existed to date. The most established and widely used instrument for this purpose – developed in the questionnaire tradition of the Five-Factor Model – is the Revised NEO Personality Inventory

(NEO-PI-R; Costa & McCrae, 1992). To meet the need for a more efficient Big Five measure that also allows the facet structure of the Big Five domains to be assessed, the well-established Big Five Inventory (BFI) was recently revised (Soto & John, 2017a; for a German adaptation, see Danner et al., 2016). The original BFI was developed to combine the strengths of the lexical Big Five and questionnaire-based Five-Factor Model traditions by using short phrases to assess the prototypical content of each Big Five dimension (John et al., 2008). The new BFI-2 incorporates several key advances, such as a robust facet-level structure, and balanced keying to minimize the influence of acquiescent responding, while still maintaining the brevity and accessibility of the original BFI (Soto & John, 2017a). The resulting 60-item BFI-2 allows the assessment of both the domain level of the Big Five and the three most prototypical facets of each of these domains.

Based on this 60-item BFI-2, Soto and John (2017b) recently developed and validated two short-form measures for the Anglo-American context. The first measure, the BFI-2-S, consists of 30 items and also allows the 15 facets of the BFI-2 to be investigated. The second measure, the even shorter BFI-2-XS, consists of only 15 items; by including one item from each of the three facets defining each Big Five domain, these 3-item domain scales reflect the full breadth of the Big Five dimensions as defined on the original BFI-2. Due to its brevity, however, the BFI-2-XS can be used only to assess the Big Five domains. These BFI-2 short forms were developed for use in specific research contexts that impose severe constraints on the amount of time that can be devoted to personality assessment, such as large-scale surveys or laboratory studies.

Soto and John (2017b) validated the two short forms of the BFI-2 using a university student sample and a more extensive and heterogeneous Internet sample. Their results indicate that, at the level of the Big Five domains, the BFI-2-S and the BFI-2-XS capture approximately 91% and 80%, respectively, of the total variance in the full BFI-2 domain scales. Both short-form instruments were shown to have a clear factorial structure, thus indicating appropriate factorial validity. With regard to the facet level, the authors could show that BFI-2-S facets retained approximately 89% of the predictive power of the full BFI-2.

Given the promising findings of this initial validation study for the BFI-2-S and BFI-2-XS, and in view of the need for short scales that also incorporate a more fine-grained facet structure of the Big Five, we adapted the BFI-2-S and the BFI-2-XS to the German context. When doing so, we used the item translations carried out when adapting the full BFI-2 to German. These translations were based on the state of the art method, the so-called TRAPD (Translation, Review, Adjudication, Pretesting, and Documentation) approach (Harkness, 2003; see also Harkness, Villar,

& Edwards, 2010): Two translations into German were carried out independently by professional translators experienced in translating questionnaires. In a reconciling step, these two translations were then compared and aggregated to one single optimal solution.

The present study aims, first, to investigate the psychometric properties of these German adaptations of the two BFI-2 short forms. Second, and more importantly, we aim to shed further light on the reliability and validity of the two short forms of the BFI-2. Therefore, we will investigate both the stability and the construct validity of the facets by comparing them to the corresponding scales of the NEO-PI-R. In addition, by investigating the criterion validity and incremental validity of the BFI-2-S facets for predicting central life outcomes that have been previously shown to relate with the Big Five (educational attainment, crystallized intelligence, life satisfaction, health, and income; see Rammstedt, Danner, & Lechner, 2017), we will examine the utility of the scales and the merits of assessing the facets in addition to the domain scales.

Method

Samples and Procedure

Data were collected through an online survey. Respondents were part of a regularly recruited online panel survey. The survey was conducted by a commercial online research organization in Germany (Respondi). A monetary incentive was paid to respondents upon participation. The survey included a 1-item attention check ("This is a functional check of the survey. Please choose the response category *disagree a little* here."). Participants who failed this check were excluded from the survey.

$N = 1,338$ respondents (50% female) were included in our analyses. The sample was heterogeneous with regard to age ($M = 42.77$, $SD = 13.94$) and education (36% lower secondary, 33% intermediate secondary, 16% higher secondary or general higher education entrance qualification, 15% university degree). A subset of the sample ($N = 406$, 50% female) participated in a retest after an interval of 6 weeks (for a more detailed description of the sample, see Danner et al., 2016).

To investigate the measurement invariance of the German adaptations of the BFI-2-S and BFI-2-XS compared to the original Anglo-American version, we reanalyzed the Internet sample described and analyzed by Soto and John (2017b, Study 2) in their validation study of the two short-form versions. The 2,000 members of this sample (50% female, $M_{age} = 28.85$, $SD = 11.82$) volunteered to complete an online version of the BFI-2 in exchange for automatically generated feedback about their personalities.

Measures

All participants in our sample completed the German adaptation of the full BFI-2 (Danner et al., 2016). The BFI-2 consists of 60 short-phrase items, with responses made on a 5-point rating scale ranging from *strongly disagree* (1) to *strongly agree* (5). Based on these responses, the scale scores for the BFI-2-S and BFI-2-XS were computed. The items of the two short scales are displayed in the Electronic Supplementary Material, ESM 1.

In addition, all respondents reported (a) their health status based on the single item "How would you describe your health status in general?", rated on a scale from *poor* (1) to *excellent* (5), and (b) their satisfaction with life measured by the well-established single item (see Beierlein, Kovaleva, László, Kemper, & Rammstedt, 2014) "How satisfied are you with your life in general?", rated on a scale from *not satisfied at all* (0) to *completely satisfied* (11). In addition, we assessed the following sociodemographic variables: age (in years); educational attainment (six categories from 1 = no formal education to 6 = university degree); labor force status (working vs. not working); and income based on 17 categories ranging from less than € 300 per month (1) to more than € 10,000 per month (17).

A subset of participants ($N = 411$) additionally completed the BEFKI-GC-K, a 12-item scale for the measurement of crystallized intelligence (Schipolowski et al., 2014).

After a retest interval of 6 weeks, another subset of respondents ($N = 406$) completed the BFI-2 items again, and part of this subsample ($N = 204$) also completed the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992; German adaptation: Ostendorf & Angleitner, 2003). The NEO-PI-R comprises 240 items and captures the Big Five personality domains and six facets of each domain (see Table 7).

Other subsamples completed additional personality scales that were not analyzed for the present study.

Results and Discussion

Psychometric Properties of the German Adaptations of the BFI-2-S and BFI-2-XS

The most central questions are (a) to what extent the German adaptations of the BFI-2-S and BFI-2-XS scales demonstrate adequate psychometric properties and (b)

whether these properties are comparable to those of the Anglo-American source versions.[2] Table 1 displays the means, standard deviations, and reliability estimates (retest correlations and Cronbach's alpha) for the BFI-2-S and BFI-2-XS domain scales and the BFI-2-S facet scales (for interscale correlations, see Table 2). Both the BFI-2-S and BFI-2-XS assess the Big Five by means of extremely short scales consisting of six and three items per dimension, respectively, and these items were selected to cover a maximum bandwidth – namely, three different facets of each dimension. Therefore, the widely used Cronbach's alpha coefficient will tend to underestimate the reliability of the scales (see Rammstedt & Beierlein, 2014), and retest coefficients will provide more appropriate reliability estimates. As shown in Table 1, 6-week retest reliabilities of the domain scales averaged .83 for the 6-item scales on the BFI-2-S, and still .77 for the 3-item scales on super-brief BFI-2-XS. For the 2-item facet scale scored from the BFI-2-S, retest reliability was still .72. The sizes of the reliability coefficients are in line with, or even exceed, those of similarly brief Big Five measures (see Rammstedt & John, 2005, 2007; Gosling et al., 2003).

For completeness, and to compare reliability indicators for the German adaptations with those of the Anglo-American source versions, Table 1 also includes Cronbach's alpha coefficients for the domain scales and facet scales. Not surprisingly, given the brevity of the short measures, the internal consistency coefficients are quite low compared to what is expected from full-length measures. The 6-item domain scales of the BFI-2-S showed an average internal consistency of .73; for the 3-item BFI-2-XS scales they averaged .53. As shown in Table 1, these coefficients are similar to those of the original US versions, both in terms of their sizes (although these are, in most cases, somewhat smaller) and their rank order.

Finally, Table 1 shows the part-whole correlations of both short-form measures with the full German BFI-2. Although the number of items is half that of the BFI-2, the BFI-2-S with its 30 items still explains approximately 90% of the variance in the full scale at the domain level and 78% at the facet level. And although the number of items in the BFI-2-XS is just one-quarter of that of the BFI-2, the 15 items of the extra-short form still explain three-quarters of the variance in the BFI-2 domain scales.

We additionally investigated the between-domain discrimination of the BFI-2S and the BFI-2-XS by means of their scale intercorrelations (see Table 2). As expected and in line with the literature (Rammstedt & John, 2007; Soto & John, 2017a), there were only small to medium correlations between the domain scores. On average, the intercorrelation (absolute values, Fisher-transformed, averaged,

back-transformed) was .26 for the BFI-2-S and .15 for the BFI-2-XS. The correlations between the BFI-2-S facet scores are shown in ESM 2, Table E1. Also for the facets, there were – as expected – small to medium correlations between facets of different domains (.17 on average) but medium to large correlations between facets of the same domain (.44 on average).

Factorial Structure

To examine the factorial structure of the BFI-2-S and BFI-2-XS, we conducted exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009) with orthogonal target rotation. In the case of the BFI-2-XS, we modeled the Big Five based on its 15 items. Because the BFI-XS incorporates positively and negatively keyed items – thus allowing the measurement of acquiescence – and because acquiescence can bias both the factorial structure and the model fit, we included in the model an acquiescence factor capturing the tendency to agree regardless of item content (Aichholzer, 2014; Soto & John, 2017b). In the case of the BFI-2-S, the model is based on the 15 facet scores. We did not include an acquiescence factor for the BFI-2-S model because the facet scores are the arithmetic mean of one positively and one negatively keyed item, and they have thus already been corrected for acquiescence (Danner & Rammstedt, 2016; Rammstedt & Danner, 2017). The standardized factor loadings and the model fit are reported in Tables 3 and 4, respectively. For both instruments, all indicators (facets and items, respectively) loaded most strongly on their corresponding factors. The model fit and the pattern of loadings in both cases suggests a good fit (e.g., CFI ≥ .979, RMSEA ≤ .033) which suggests that – like the US versions of the scales – the German-language adaptations of both the BFI-2-S and the BFI-2-XS clearly reflect the intended five-dimensional structure.

In addition, we analyzed the BFI-2-S facet scores and the BFI-2-XS items (centered items, see Soto & John, 2017b) with principal component analyses (five factors, varimax rotated). The respective loadings are also shown in Tables 3 and Table 4. As can be seen, the patterns of loadings were highly similar to the ESEM results.

Cross-Cultural Equivalence

To investigate the comparability of the German adaptations of the scales with their Anglo-American source versions, we used the Internet sample from the US validation study and directly examined the measurement invariance for the scales of the BFI-2-S and BFI-2-XS across the two languages and cultures.

---

[2] The original inputs and outputs of the present and all following analyses are available on request.

Table 1. Means, standard deviations (SD), test-retest correlations, and Cronbach's α for the BFI-2-S and BFI-2-XS, and their correlations with the full BFI-2 domain and facet scales

| | M | | SD | | Retest | | Cronbach's α | | | | BFI-2 Domain/Facet | |
| | | | | | | | S | | XS | | | |
	S	XS	S	XS	S	XS	G	US	G	US	S	XS
Extraversion	3.13	2.92	0.65	0.71	.88	.82	.71	.77	.47	.63	.94	.84
Sociability	3.01		0.83		.77		.47	.70			.90	
Assertiveness	3.07		0.90		.82		.66	.72			.90	
Energy	3.32		0.82		.71		.57	.60			.86	
Agreeableness	3.70	3.64	0.57	0.68	.79	.72	.65	.75	.45	.55	.94	.85
Compassion	3.94		0.74		.69		.51	.48			.90	
Respectfulness	3.96		0.73		.63		.39	.48			.89	
Trust	3.20		0.74		.66		.21	.53			.87	
Conscientiousness	3.70	3.64	0.63	0.73	.84	.78	.75	.78	.53	.61	.95	.88
Organization	3.67		0.88		.79		.65	.79			.93	
Productiveness	3.56		0.80		.76		.55	.58			.90	
Responsibility	3.86		0.68		.65		.33	.47			.85	
Negative Emotionality	2.74	2.84	0.74	0.86	.85	.84	.80	.84	.67	.73	.96	.89
Anxiety	3.01		0.81		.75		.41	.65			.88	
Depression	2.61		0.97		.79		.69	.67			.95	
Volatility	2.60		0.89		.76		.67	.75			.92	
Open-Mindedness	3.27	3.20	0.70	0.78	.79	.66	.72	.74	.53	.57	.95	.88
Aesthetic Sensitivity	2.94		1.10		.64		.69	.54			.93	
Intellectual Curiosity	3.36		0.85		.65		.57	.42			.87	
Creative Imagination	3.50		0.83		.69		.64	.64			.93	
Mean (Domain)	3.31	3.25	0.66	0.75	.83	.77	.73	.78	.53	.62	.95	.87
Mean (Facet)	3.31		0.84		.72		.53	.60			.90	

Notes. Retest interval was 6 weeks. Coefficients for the US data were taken from Soto and John (2017b); G = German; XS = BFI-2-XS; S = BFI-2-S; Averages of all the correlational indices were computed using Fisher Z transformation.

Table 2. Correlations between manifest domains scores for the BFI-2-S/BFI-2-XS

	Extraversion	Agreeableness	Conscientiousness	Negative Emotionality
Agreeableness	.19/.04			
Conscientiousness	.25/.11	.31/.22		
Negative Emotionality	−.34/−.20	−.31/−.18	−.30/−.22	
Open-Mindedness	.35/.30	.18/.12	.14/.00	−.21/−.13

Note. N = 1,338, all r ≥ .11 are p < .001.

We investigated exact measurement invariance using confirmatory factor analysis (CFA; see Chen, 2007) and approximate measurement invariance using Bayes structural equation modeling (see Cieciuch, Davidov, Schmidt, Algesheimer, & Schwartz, 2014). Whereas exact measurement invariance tests whether factor loadings and item intercepts are exactly identical across groups, approximate invariance allows some a small amount of variation across the two solutions (technically, variance is 0.01 in the present study) for these parameters, and it tests whether they differ *meaningfully* between groups. We evaluated measurement invariance separately for each of the five domains measured by each of the two instruments. This made the analyses more sensitive to non-invariance of single items, and it also avoided the complication of cross-loadings of single items on different domains (see Tables 3 and 4). Once again, the 15 facet scores were used as manifest indicators for the BFI-2-S, and the 15 items were used as manifest indicators for the BFI-2-XS. For both the exact and the approximate approach, we tested three levels of invariance: configural invariance (same factorial structure); metric invariance (same factor loadings); and scalar invariance

Table 3. Standardized Factor Loadings for the 2-item BFI-2S scale scores

BFI-2S facets	ESEM					PCA				
	E	A	C	N	O	E	A	C	N	O
Sociability	**.57**	−.12	.18	−.11	−.02	**.87**	.11	−.02	.07	−.02
Assertiveness	**.38**	−.16	.10	−.18	.35	**.62**	−.23	.07	−.22	.32
Energy	**.38**	.09	.27	−.28	.18	**.62**	.24	.24	−.29	.16
Compassion	.30	**.47**	.16	−.02	−.05	.27	**.76**	.18	.11	.17
Respectfulness	.07	**.48**	.19	−.16	−.01	.00	**.72**	.31	−.19	.09
Trust	.03	**.32**	.03	−.17	−.10	−.06	**.72**	−.07	−.20	−.06
Organization	−.09	.13	**.51**	.04	.10	−.03	.06	**.80**	.09	−.01
Productiveness	−.07	.10	**.53**	−.16	.25	.11	.04	**.77**	−.26	.07
Responsibility	.01	.17	**.45**	−.09	.15	.15	.21	**.76**	−.14	.07
Anxiety	.03	−.01	.00	**.61**	−.19	−.09	−.01	.00	**.87**	−.05
Depression	−.19	−.02	−.21	**.66**	−.23	−.37	−.07	−.19	**.72**	−.06
Volatility	.16	−.39	−.06	**.61**	−.19	.12	−.36	−.18	**.76**	−.13
Aesthetic Sensitivity	.23	.26	−.15	.08	**.41**	−.02	.17	−.06	.07	**.75**
Intellectual Curiosity	.18	.18	−.09	.00	**.57**	.07	−.01	.07	−.15	**.80**
Creative Imagination	.30	.08	.06	−.07	**.44**	.36	−.01	.15	−.12	**.64**

Notes. E = Extraversion, A = Agreeableness, C = Conscientiousness, N = Negative Emotionality, O = Open-Mindedness, highest loadings are bolded, ESEM: RMSEA = .029, CFI = .990, SRMR = .012, *N* = 1,338.

Table 4. Standardized factor loadings of the BFI-2-XS items

Item	ESEM						PCA (centered items)				
	E	A	C	N	O	ARS	E	A	C	N	O
Ich bin eher ruhig. (R)	**−.55**	.13	.08	−.18	.03	.16	**.83**	−.03	−.11	−.16	.02
Ich neige dazu, die Führung zu übernehmen.	**.43**	−.10	.04	−.12	.30	.15	**−.61**	−.16	−.13	−.25	.26
Ich bin voller Energie und Tatendrang.	**.51**	.23	.23	−.33	.20	.19	**−.46**	.26	−.31	−.45	.20
Ich bin einfühlsam, warmherzig.	.05	**.58**	.16	.06	.14	.20	.01	**.67**	−.28	.10	.17
Ich bin manchmal unhöflich und schroff. (R)	.10	**−.48**	−.28	.12	−.02	.16	−.27	**−.61**	.30	.18	−.08
Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	.05	**.38**	−.09	−.08	.01	.16	−.10	**.72**	.21	−.13	−.08
Ich bin eher unordentlich. (R)	.01	−.01	**−.63**	.01	.05	.15	−.06	.03	**.76**	−.04	.05
Ich neige dazu, Aufgaben vor mir herzuschieben. (R)	−.11	−.04	**−.61**	.19	.05	.16	.06	−.01	**.73**	.21	.05
Ich bin verlässlich, auf mich kann man zählen.	.10	.30	**.34**	−.06	.06	.22	−.05	.31	**−.57**	−.05	.05
Ich mache mir oft Sorgen.	−.15	.12	.01	**.65**	.02	.16	.07	.09	−.09	**.81**	.00
Ich bin oft deprimiert, niedergeschlagen.	−.31	−.13	−.16	**.70**	−.01	.14	.16	−.14	.18	**.78**	−.04
Ich bin ausgeglichen, nicht leicht aus der Ruhe zu bringen. (R)	−.19	.21	.11	**−.62**	.16	.16	.34	.14	−.13	**−.71**	.11
Ich kann mich für Kunst, Musik und Literatur begeistern.	.05	.14	−.07	.00	**.43**	.14	.05	.18	.10	.02	**.69**
Mich interessieren abstrakte Überlegungen wenig. (R)	−.13	−.02	.03	.03	**−.54**	.16	.05	.04	−.03	.02	**−.78**
Ich bin originell, entwickle neue Ideen	.30	.01	.07	−.12	**.55**	.17	−.32	−.00	−.13	−.20	**.62**

Note. E = Extraversion, A = Agreeableness, C = Conscientiousness, N = Negative Emotionality, O = Open-Mindedness, ARS = acquiescence; (R) = item to be revised; highest loadings are bolded, ESEM: RMSEA = .033, CFI = .979, SRMR = .016, *N* = 1,338.

(same factor loadings and same item intercepts). Exact measurement invariance was evaluated on the basis of the criteria suggested by Chen (2007), where a difference of CFI < .01 suggests a stricter level of invariance. Approximate invariance was evaluated on the basis of criteria suggested by Cieciuch et al. (2014), where a posterior predictive probability (*ppp*) greater than zero suggests that a specified level of invariance can be accepted. The results are shown in Table 5. For the BFI-2-S, results suggest exact scalar invariance for Negative Emotionality, metric invariance for Extraversion and Conscientiousness, and only configural invariance for Agreeableness and Open-Mindedness. However, because the *ppp* was greater than zero for all scalar models, results suggest approximate scalar invariance for all domains. For the BFI-2-XS, results suggest exact metric invariance for Open-Mindedness and only

configural invariance for Extraversion, Agreeableness, Conscientiousness, and Negative Emotionality. However, because the *ppp* was greater than zero for all scalar models, results again suggest approximate scalar invariance for all domains.

Convergent Validity

As an extension to the initial development and validation study for the BFI-2-S and BFI-2-XS conducted by Soto and John (2017b), we investigated the construct validity of the scales in more detail. First, we compared the BFI-2-XS and BFI-2-S domain scores and the facet scores of the BFI-2-S with those of the NEO-PI-R, as one of the most established full-length measures assessing both the Big Five domains and their facets. Then, we investigated the criterion validity of the domain scales and the differential validity of the facet scales for important life outcomes.

The NEO-PI-R was assessed only on the second measurement occasion. Thus, the correlations between the BFI-2-S, BFI-2-XS, and the NEO-PI-R scores were estimated on the basis of the data from the second assessment. Table 6 displays the correlations of the BFI-2-XS and BFI-2-S domain scales with the corresponding NEO-PI-R domains. As the BFI/BFI-2 and the NEO-PI-R differ somewhat in the construct definitions of the Big Five used (see John et al., 2008; Rammstedt & John, 2007), these differences necessarily limit the absolute size of the correlations. Of more importance here is the extent to which the short-form versions of the BFI-2 can reflect the convergence of the full-length BFI-2 with the NEO-PI-R. Therefore, for comparison purposes, the correlations of the NEO-PI-R domain scales with the German-language full BFI-2 are also reported in Table 6. Not surprisingly, the size of the correlations decreased slightly with scale length. Whereas the convergent correlations averaged .79 for the full BFI-2, they averaged .76 for the BFI-2-S, and of .69 for the BFI-2-XS. That is, the full BFI-2 and the NEO-PI-R domain scales share, on average, 62% common variance; this value drops only to 58% for the BFI-2-S, and then more substantially to 48% for the BFI-2-XS. However, given that the number of items was halved from version to version (60 compared to 30 and 15, respectively), the decrease in convergence of 3% points for the BFI-2-S can be regarded as minor. For the BFI-2-XS, shared variance with the NEO-PI-R decreases by 9% compared to the full BFI-2. This is due primarily to Extraversion and Open-Mindedness. Thus, it seems that, in these domains, the 3-item scales of the BFI-2-XS cover markedly less of the content contained in the 48-item NEO-PI-R scales.

To investigate in more detail the convergence between the BFI-2-S and the NEO-PI-R, we computed correlations of the BFI-2-S facets with the NEO-PI-R domains and the six NEO-PI-R facets of the corresponding Big Five domains (see Table 7). Overall, each BFI-2-S facet correlated substantially with the relevant NEO-PI-R domain scales, with the average correlation being .60.

With regard to the correlations of the BFI-2-S facets with the most relevant NEO-PI-R facets, it must be kept in mind that the BFI-2 facets were selected to reflect the common core across different facet-structural approaches to the Big Five – of which the NEO-PI-R is just one example – and that the specific facets of the BFI-2 do not always match a single facet in the NEO-PI-R. For example, the NEO-PI-R Openness facets do not include a unique Creative Imagination facet. However, in the cases in which a direct conceptual match can be made, these facets also show the highest correlations with this corresponding facet (e.g., .79 for the Depression facets on BFI-2 and the NEO-PI-R). Soto and John (2017a) report for the full BFI-2 the correlations of the BFI-2 facets with the NEO-PI-R facets of the corresponding domains, and identified 21 such correlations as distinctive because a NEO-PI-R facet correlated more strongly with a particular BFI-2 facet than with the two other, same-domain BFI-2 facets. Thus, for comparison purposes, distinctive correlations between the facets reported for the Anglo-American full-length version are bolded in Table 7. This table shows that, of the 21 NEO-PI-R facets that Soto and John (2017a) identified as showing a distinctive correlation with a particular BFI-2 facet, 19 also showed their strongest correlation with the same facet in the German adaptation of the BFI-2-S, and all were significant at the .001-level. The lone exceptions were that NEO-PI-R Gregariousness correlated more strongly with German BFI-2-S Energy Level than with Sociability and that NEO-PI-R Fantasy correlated slightly more strongly with German BFI-2-S Aesthetic Sensitivity than with Creative Imagination. The former exception may be explained by the inclusion of emotional content (e.g., enjoyment vs. boredom) in some NEO-PI-R Gregariousness items, whereas the latter exception may be due to the exclusion of the BFI-2 item "Has trouble imagining things" from the BFI-2-S, as well as the focus of NEO-PI-R Fantasy on idle daydreaming rather than creativity and originality (cf. Soto & John, 2017a). A complete comparison of the correlation coefficients of the German BFI-2-S facets with the NEO-PI-R facets of the corresponding domains to those of the Anglo-American BFI-2 is visualized in ESM 3.

To more formally compare the similarity of the correlations found in the present study for the German adaptation of the BFI-2-S with those found for the Anglo-American version of the BFI-2 reported by Soto & John (2017a), we computed a column-vector correlation for each German BFI-2-S facet comparing its set of 30 correlations with the German NEO-PI-R facets to the corresponding set of correlations between the Anglo-American BFI-2 facets and the

Table 5. Exact and approximate measurement invariance for the BFI-2-S and the BFI-2-XS

Domain	Invariance level	BFI-2-S Exact invariance (CFI)	BFI-2-S Approximate invariance (*ppp*)	BFI-2-XS Exact invariance (CFI)	BFI-2-XS Approximate invariance (*ppp*)
Extraversion	Configural	1.000	.450	1.000	.483
	Metric	1.000	.468	0.953	.390
	Scalar	0.984	.516	0.926	.315
Agreeableness	Configural	1.000	.482	1.000	.515
	Metric	0.981	.485	0.976	.404
	Scalar	0.967	.475	0.546	.355
Conscientiousness	Configural	1.000	.485	1.000	.505
	Metric	0.991	.478	0.989	.515
	Scalar	0.973	.237	0.955	.406
Negative Emotionality	Configural	1.000	.479	1.000	.513
	Metric	0.995	.472	0.981	.462
	Scalar	0.990	.278	0.927	.140
Open-Mindedness	Configural	1.000	.500	1.000	.489
	Metric	0.985	.466	1.000	.457
	Scalar	0.869	.092	0.986	.040

Note. N = 3,338, CFI = comparative fit index, *ppp* = posterior predicted *p* value.

Table 6. Convergent validity correlations with the NEO-PI-R domain scales for the BFI-2-S and the BFI-2-XS domain scales as compared to the full BFI-2

NEO-PI-R	BFI-2	BFI-2-S	BFI-2-XS
Extraversion	.80	.75	.64
Agreeableness	.78	.75	.71
Conscientiousness	.81	.79	.72
Neuroticism/Negative Emotionality	.87	.84	.81
Open-Mindedness	.66	.65	.52
Mean	.79	.76	.69

Notes. N = 204; all *p*s < .001. Mean correlations were computed using Fisher's *r* to *Z* transformation.

Anglo-American NEO-PI-R facets. Across the 15 BFI-2-S facets, these column-vector correlations ranged from .57 to .97, and averaged .83 (computed via Fisher's *r* to *Z* transformation), thus suggesting a highly similar correlation pattern for the German BFI-2-S compared to the full-length Anglo-American version.

Criterion Validity

To investigate the criterion validity of the two short scale measures, we first computed the correlations of the domain scores with the important life outcomes assessed in the first measurement occasion, namely, educational attainment, intelligence, life satisfaction, self-reported health, and income. Numerous previous studies, some of which were based on the original full BFI or its short forms, have demonstrated reliable associations between the Big Five and these outcomes. For example, educational attainment

is known to relate positively with Open-Mindedness (e.g., Caspi, Roberts, & Shiner, 2005; George, Helson, & John, 2011; Rammstedt, 2007a). For crystallized intelligence (G_c), a moderate positive correlation with Open-Mindedness and a small negative correlation with Negative Emotionality are typically found (Rammstedt, Danner, & Martin, 2016; Von Stumm & Ackerman, 2013). Life satisfaction and self-perceived health have been found to be negatively associated with Neuroticism or Negative Emotionality and positively associated with Extraversion (DeNeve & Cooper, 1998; Gutierrez, Jimenez, Hernández, & Puente, 2005; Rammstedt, 2007b; Rammstedt, Kemper, Klein, Beierlein & Kovaleva, 2013). Moreover, some studies have also found a positive association between health and Conscientiousness (e.g., Bogg & Roberts, 2004). And finally, emotionally stable individuals have been found to report higher incomes (e.g., Judge, Higgins, Thoresen, & Barrick, 1999).

Table 7. Correlations of the BFI-2-S facet scales with the corresponding facet scales of the NEO-PI-R

BFI-2-S Facets	Domain	NEO-F1	NEO-F2	NEO-F3	NEO-F4	NEO-F5	NEO-F6
Extraversion		Warmth	Gregariousness	Assertiveness	Activity	Excitement-Seeking	Positive Emotionality
F1 Sociability	.55	.32	**.43**	.43	.49	.23	.39
F2 Assertiveness	.49	.13	.28	**.74**	.43	.25	.17
F3 Energy	.64	.44	.53	.35	**.57**	.22	**.56**
Agreeableness		Trust	Straightforwardness	Altruism	Compliance	Modesty	Tender-Mindedness
F1 Compassion	.60	.46	.34	**.63**	.30	.15	**.57**
F2 Respectfulness	.59	.48	**.38**	.55	**.41**	.22	.34
F3 Trust	.61	**.61**	.30	.44	.40	.28	.37
Conscientiousness		Competence	Order	Dutifulness	Achievement Striving	Self-Discipline	Deliberation
F1 Organization	.60	.38	**.66**	.51	.41	.53	.42
F2 Productiveness	.71	.47	.55	.54	**.52**	**.78**	.51
F3 Responsibility	.66	.46	.49	**.56**	.48	.64	**.53**
Negative Emotionality		Anxiety	Angry Hostility	Depression	Self-Consciousness	Impulsiveness	Vulnerability
F1 Anxiety	.69	**.69**	.61	.63	.43	.39	.60
F2 Depression	.73	.64	.56	**.79**	.56	.34	**.63**
F3 Volatility	.70	.56	**.78**	.57	.48	.48	.63
Open-Mindedness		Fantasy	Aesthetics	Feelings	Actions	Ideas	Values
F1 Aesthetic Sensitivity	.42	.29	**.48**	.18	.08	.35	.08
F2 Intellectual Curiosity	.51	.17	.34	.29	.20	**.73**	.07
F3 Creative Imagination	.51	**.26**	.37	.34	.28	.52	.07

Notes. F = facet; N = 204; correlations of identically labeled facets are bold italicized; distinctive correlations according to Soto and John (2017a) are set in bold. All coefficients > .22 = p < .001.

Table 8 shows the correlations between the German BFI-2-S and BFI-2-XS domains and the various indicators of life outcomes. For both short forms, the correlation patterns were highly similar in terms of direction and size. For both scales, the correlation pattern reflects the associations typically found with the particular outcome variables. For example, the Open-Mindedness domain was positively associated with educational attainment, whereas Negative Emotionality was negatively correlated with life satisfaction and health, and Extraversion was positively correlated with life satisfaction and health. As can be seen from the R^2 coefficients in Table 8, both scales explained a substantial portion of the variance in the outcomes. Furthermore, the proportion of variance explained by the BFI-2-S and BFI-2-XS domains scores did not decrease compared to the domain scores of the full 60-item version of the BFI-2. Also with regard to the facet scales the proportion of variance explained by the BFI-2-S facet scores was only marginally smaller compared to the facet scores of the full BFI-2 (average ΔR^2 = .02).

As an indicator for the differential validity, and thus for the incremental value, of the facets compared to using only the domain scales, we analyzed the criterion validity for the 15 facets separately. As can be seen from Table 8, in several cases, the facets of a domain show a differential correlation pattern. For example, the positive associations between

Extraversion and life satisfaction and health are primarily due to the facet Energy Level, whereas the other two facets of Extraversion show markedly lower correlations with these outcomes. Similarly, the fact that more neurotic persons report lower incomes is primarily caused by the facets Anxiety and Depression and to a much lesser extent by Emotional Volatility. Moreover, the positive association between Open-Mindedness and educational attainment is due primarily to positive associations with the Intellectual Curiosity and Aesthetic Sensitivity facets and to a much lesser extent to the Creative Imagination facet. For all outcome variables, the proportion of the variance explained by the 15 facets (mean R^2 = .19) is greater than that explained by the domain scales (mean R^2 = .14), thus indicating that the BFI-2-S facets provide approximately 36% greater predictive validity compared to the domain scales.

Summary and General Discussion

Results of the present research demonstrate the validity and utility of the German adaptations of the two short forms of the BFI-2, namely, the BFI-2-S and BFI-2-XS. Specifically, the German-language adaptations and the Anglo-American source versions of the BFI-2-S and BFI-2-XS

Table 8. Correlations of the BFI-2-S and BFI-2-XS domain scales and BFI-2-S facet scales with outcome variables

	Education	Intelligence (G_c)	Life Satisfaction	Health	Income
Extraversion					
Domain scales					
BFI-2-S	.16	.01	.28	.26	.16
BFI-2-XS	.15	.03	.24	.24	.16
BFI-2S facet scales					
F1 Sociability	.03	−.06	.14	.13	.06
F2 Assertiveness	.21	.10	.17	.13	.16
F3 Energy	.12	−.03	.34	.34	.13
Agreeableness					
Domain scales					
BFI-2-S	.03	−.03	.18	.04	−.02
BFI-2-XS	.00	−.04	.18	.04	.01
BFI-2 facet scales					
F1 Compassion	.02	−.05	.13	.00	−.04
F2 Respectfulness	.00	−.03	.18	−.08	−.03
F3 Trust	.04	.01	.11	.00	.01
Conscientiousness					
Domain scales					
BFI-2-S	−.03	−.02	.19	.12	.10
BFI-2-XS	−.05	−.01	.16	.10	.11
BFI-2 facet scales					
F1 Organization	−.06	−.08	.11	.05	.07
F2 Productiveness	−.03	−.02	.18	.14	.13
F3 Responsibility	.02	.08	.16	.11	.05
Negative Emotionality					
Domain scales					
BFI-2-S	−.15	−.15	−.45	−.44	−.20
BFI-2-XS	−.15	−.17	−.45	−.47	−.22
BFI-2 facet scales					
F1 Anxiety	−.16	−.20	−.34	−.35	−.21
F2 Depression	−.13	−.10	−.50	−.49	−.19
F3 Volatility	−.09	−.08	−.28	−.25	−.10
Open-Mindedness					
Domain scales					
BFI-2-S	.29	.21	.10	.05	.09
BFI-2-XS	.30	.20	.07	.03	.07
BFI-2 facet scales					
F1 Aesthetic Sensitivity	.24	.22	.04	.01	.02
F2 Intellectual Curiosity	.28	.12	.08	.05	.09
F3 Imagination	.13	.12	.12	.05	.10
R^2 (BFI-2 domains)	.10	.11	.22	.21	.06
R^2 (BFI-2-S domains)	.11	.08	.23	.23	.06
R^2 (BFI-2-XS domains)	.11	.08	.24	.25	.07
R^2 (BFI-2 facets)	.16	.21	.30	.30	.09
R^2 (BFI-2-S facets)	.15	.16	.28	.30	.08

Notes. All facet correlations with education, life satisfaction, health, and income $\geq |.10|$ are $p < .001$; all correlations with $G_c \geq |.20|$ are $p < .001$.

showed quite similar psychometric properties. Moreover, the German adaptations of both short forms proved to be approximately invariant to the Anglo-American source versions. Extending the original scale development and validation study by Soto and John (2017b), we were able to show high retest reliability of the scales and their facets, with sizes being in line with, or even exceeding, those of other Big Five short scales.

By extending the original validation study for the Anglo-American BFI-2-S and BFI-2-XS versions (Soto & John, 2017b), we have also added the body of validity evidence for these scales: First, we demonstrated that both short forms show substantial convergence with the NEO-PI-R domain scales, with a pattern of associations very similar to that shown by the full BFI-2 and only modest (for the BFI-2-S) or moderate (for the BFI-2-XS) decreases in correlation strength. Our results further revealed the construct validity of the BFI-2-S facets: The distinctive correlation pattern of the facets with those of the NEO-PI-R that was found for the Anglo-American BFI-2 (see Soto & John, 2017a) could be largely replicated for the German adaptation of the BFI-2-S.

Second, we investigated the criterion validity of the domain scales for the two short-form versions for central life outcomes. Our results revealed that, for all outcome variables investigated, the known associations with the Big Five traits established by previous research could be replicated. In addition, our analyses at the facet level further highlighted the incremental validity of Big Five facets, and thus the importance of investigating the facet scales in addition to the domain scales. Several life outcomes showed distinctive correlations with particular BFI-2-S facets, and collectively the 15 facet scales provided substantially greater predictive validity than did the five domain scales.

One central limitation of the present study is that our analyses are based on data that extract the BFI-2-S and BFI-2-XS items from an administration of the full BFI-2, rather than data from a separate administration of the short forms (cf. Smith et al., 2000). To address this caveat we compared our findings to those from a small and comparatively homogeneous sample of academics to which only the German BFI-2 short forms were administered (see Lechner, Génois, Strohmaier, & Rammstedt, 2017). Results indicated that scale means, reliabilities, and factor loadings were all very similar for embedded versus separate administrations of the short forms. For example, congruence coefficients comparing the present factor loadings (Table 3) with those obtained by Lechner et al. (2017) were all at least .91 ($M = .95$) for the BFI-2-S. However, future studies using larger and more diverse samples should be conducted to replicate these initial findings; such studies could also serve as a source for benchmark scale scores.

A second limitation concerns the relatively low Cronbach's alpha reliability of some scales, especially for the ultrashort BFI-2-XS. However, this finding is not surprising given that Cronbach's alpha is determined by a scale's length (i.e., number of items) and mean interitem correlation (i.e., content redundancy; Gosling et al., 2003; Rammstedt & Beierlein, 2014; Smith et al., 2000). Because the BFI-2-XS scales were developed to maximize content breadth using only three items per Big Five domain, they tend to have relatively modest alpha reliabilities in both English and German. However, as shown by the present results, as well as by Soto and John (2017b), modest alphas do not prevent the BFI-2 short forms from demonstrating strong retest reliability, structural validity, and – most importantly – predictive power (Tables 1, 3, 4, and 8).

In sum, our results suggest that the German adaptations of the BFI-2-S and BFI-2-XS can serve as useful instruments for assessing the Big Five in settings that severely limit assessment time, such as large-scale surveys. In contrast with previously established Big Five short scale measures, the 30-item BFI-2-S also allows investigation of the main Big Five facets in addition to the domains, which proved to be an important additional source of personality information. We caution that, in settings which are less restrictive to assessment time, or which focus on individuals' scores (e.g., for candidate selection) rather than overall associations between variables, researchers should carefully consider the reduced psychometric properties of these short scales when deciding whether to administer them instead of a longer measure such as the full BFI-2. However, when the administration of a full-length personality measure is not feasible, the German BFI-2-S and BFI-2-XS offer considerable reliability and validity with minimal assessment time.

Acknowledgment

Electronic Supplementary Materials

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/1015-5759/a000481

ESM 1. Tables (.pdf)
Items of the short scales BFI-2-S and BFI-2-XS in German.
ESM 2. Table (.pdf)
Correlations between the manifest facet scores of the BFI-2-S.
ESM 3. Figure (.pdf)
Correlations of the 15 German BFI-2-S compared to the Anglo-American BFI-2 facet scales with the corresponding facet scales of the NEO-PI-R.

References

Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality, 53*, 1–4. https://doi.org/10.1016/j.jrp.2014.07.001

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397–438. https://doi.org/10.1080/10705510903008204

Beierlein, C., Kovaleva, A., László, Z., Kemper, C. J., & Rammstedt, B. (2014). *Eine Single-Item-Skala zur Erfassung der*

Allgemeinen Lebenszufriedenheit: Die Kurzskala Lebenszufrieden-heit-1 (L-1) [A single-item scale measuring general life satisfaction]. Köln, Germany: GESIS. (GESIS Working Papers 2014, 33). https://www.gesis.org/fileadmin/kurzskalen/working_papers/L1_WorkingPapers_2014-33.pdf

Bogg, T., & Roberts, B. W. (2004). Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality. *Psychological Bulletin, 130*, 887–919. https://www.gwern.net/docs/conscientiousness/2004-bogg.pdf

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology, 56*, 453–484. https://www.annualreviews.org/doi/10.1146/annurev.psych.55.090902.141913

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504. https://www.tandfonline.com/doi/abs/10.1080/10705510701301834

Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: A cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology, 5*, 982. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4157555/

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and NEO Five Factor Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Danner, D., & Rammstedt, B. (2016). Facets of acquiescence: Agreeing with negations is not the same as accepting inconsistency. *Journal of Research in Personality, 65*, 120–129. https://doi.org/10.1016/j.jrp.2016.10.010

Danner, D., Rammstedt, B., Bluemke, M., Treiber, L., Berres, S., Soto, C., & John, O. P. (2016). *Die deutsche Version des Big Five Inventory 2 (BFI-2). Zusammenstellung sozialwissenschaftlicher Items und Skalen* [German version of the Big Five Inventory 2 (BFI-2)]. Mannheim, Germany: GESIS. https://doi.org/10.6102/zis247

DeNeve, K. M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin, 124*, 197–229. https://www.ncbi.nlm.nih.gov/pubmed/9747186

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment, 18*, 192–203. http://psycnet.apa.org/doi/10.1037/1040-3590.18.2.192

George, L. G., Helson, R., & John, O. P. (2011). The "CEO" of women's work lives: How Big Five Conscientiousness, Extraversion, and Openness predict 50 years of work experiences in a changing sociocultural context. *Journal of Personality and Social Psychology, 101*, 812–830. https://doi.org/10.1037/a0024290

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of Personality and Social Psychology* (pp. 141–165). Beverly Hills, CA: Sage.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*, 1216–1229. https://www.ncbi.nlm.nih.gov/pubmed/2283588

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. Jr. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*, 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Gutierrez, J. L. G., Jimenez, B. M., Hernández, E. G., & Puente, C. P. (2005). Personality and subjective well-being: Big Five correlates and demographic variables. *Personality and Individual Differences, 38*, 1561–1569. https://doi.org/10.1016/j.paid.2004.09.015

Harkness, J. (2003). Questionnaire translation. In J. Harkness, F. van de Vijver, & P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Hoboken, NJ: Wiley.

Harkness, J. A., Villar, A., & Edwards, B. (2010). Translation, adaptation, and design. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. Ph. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multi-regional, and multicultural contexts* (pp. 117–140). Hoboken, NJ: Wiley-Blackwell.

Herzberg, P. Y., & Brähler, E. (2006). Assessing the Big-Five personality domains via short forms. A cautionary note and a proposal. *European Journal of Psychological Assessment, 22*, 139–148. https://doi.org/10.1027/1015-5759.22.3.139

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory – Versions 4a and 54*. Berkeley, CA: 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* (pp. 114–158). New York, NY: Guilford Press.

Judge, T. A., Higgins, C. A., Thoresen, C. J., & Barrick, M. R. (1999). The Big Five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*, 621–652. https://doi.org/10.1111/j.1744-6570.1999.tb00174.x

Lechner, C. M., Génois, M., Strohmaier, M., & Rammstedt, B. (2017). *Personality and social behavior on academic conferences: A network approach*. Manuscript in preparation.

Ostendorf, F., & Angleitner, A. (2003). *NEO-Persönlichkeitsinventar nach Costa und McCrae, Revidierte Fassung (NEO-PI-R). Manual* [German adaptation of the revised NEO Personality Inventory by Costa and McCrae (NEO-PI-R)]. Göttingen, Germany: Hogrefe.

Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology, 81*, 524–539. https://www.ncbi.nlm.nih.gov/pubmed/11554651

Rammstedt, B. (2007a). The 10-Item Big Five Inventory (BFI-10). *European Journal of Psychological Assessment, 23*, 193–201. https://doi.org/10.1027/1015-5759.23.3.193

Rammstedt, B. (2007b). Who worries and who is happy? Explaining individual differences in worries and satisfaction by personality. *Personality and Individual Differences, 43*, 1626–1634. https://doi.org/10.1016/j.paid.2007.04.031

Rammstedt, B., & Beierlein, C. (2014). Can't we make it any shorter? The limits of personality assessment and ways to overcome them. *Journal of Individual Difference, 35*, 212–220. https://doi.org/10.1027/1614-0001/a000141

Rammstedt, B., & Danner, D. (2017). Acquiescent responding. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 1–3). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_1276-1

Rammstedt, B., Danner, D., & Lechner, C. (2017). The association between personality and life outcomes – Results from the PIAAC longitudinal study in Germany. *Large-Scale Assessment in Education, 5*, 2. https://doi.org/10.1186/s40536-017-0035-9

Rammstedt, B., Danner, D., & Martin, S. (2016). The association between personality and cognitive ability: Going beyond simple effects. *Journal of Research in Personality, 62*, 39–44. https://doi.org/10.1016/j.jrp.2016.03.005

Rammstedt, B., & John, O. P. (2005). Kurzversion des Big Five Inventory (BFI-K): Entwicklung und Validierung eines ökonomischen Inventars zur Erfassung der fünf Faktoren der Persönlichkeit [Short version of the Big Five Inventory (BFI-K): Development and validation of an economic inventory

for assessment of the five factors of personality]. *Diagnostica, 51*, 195–206. http://psycnet.apa.org/doi/10.1026/0012-1924.51.4.195

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*, 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: 10 Item Big Five Inventory (BFI-10) [A short-scale measure assessing the Big Five dimensions of personality: the 10 item Big Five Inventory (BFI-10)]. *Methoden, Daten, Analysen, 7*, 235–251. https://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/mda/Vol.7_Heft_2/MDA_Vol7_2013-2_Rammstedt.pdf

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of Conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology, 58*, 103–139. http://psycnet.apa.org/doi/10.1111/j.1744-6570.2005.00301.x

Schipolowski, S., Wilhelm, O., Schroeders, U., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2014). BEFKI GC-K: Eine Kurzskala zur Messung kristalliner Intelligenz [BEFKI GC-K: A short-scale measure assessing crystallized intelligence]. *Methoden, Daten, Analysen, 7*, 155–183. https://doi.org/10.12758/mda.2013.010

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment,* *12*, 102–111. http://psycnet.apa.org/doi/10.1037/1040-3590.12.1.102

Soto, C. J., & John, O. P. (2017a). 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*, 117–143. https://doi.org/10.1037/pspp0000096

Soto, C. J., & John, O. P. (2017b). Short and extra-short forms of the Big Five Inventory-2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality, 68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Von Stumm, S., & Ackerman, P. L. (2013). Investment and intellect: A review and meta-analysis. *Psychological Bulletin, 139*, 841–869. http://psycnet.apa.org/doi/10.1037/a0030746

Beatrice Rammstedt
GESIS – Leibniz Institute for the Social Sciences
PO Box 12 21 55
68072 Mannheim
Germany
beatrice.rammstedt@gesis.org