

Brief report

Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German ☆

Beatrice Rammstedt ^{a,*}, Oliver P. John ^b

^a Center for Survey Research and Methodologies (ZUMA), P.O. Box 12 21 55, D-68072 Mannheim, Germany

^b Department of Psychology, University of California, Berkeley MC 1650, Berkeley, CA 94720-1650, USA

Available online 3 April 2006

Abstract

To provide a measure of the Big Five for contexts in which participant time is severely limited, we abbreviated the Big Five Inventory (BFI-44) to a 10-item version, the BFI-10. To permit its use in cross-cultural research, the BFI-10 was developed simultaneously in several samples in both English and German. Results focus on the psychometric characteristics of the 2-item scales on the BFI-10, including their part-whole correlations with the BFI-44 scales, retest reliability, structural validity, convergent validity with the NEO-PI-R and its facets, and external validity using peer ratings. Overall, results indicate that the BFI-10 scales retain significant levels of reliability and validity. Thus, reducing the items of the BFI-44 to less than a fourth yielded effect sizes that were lower than those for the full BFI-44 but still sufficient for research settings with truly limited time constraints.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Big Five personality dimensions; Five-Factor Model; Short measures; Reliability; Validity; Test construction

☆ This research was supported, in part, by Grant MH49255 from the National Institute of Mental Health and a Faculty Research Grant from the University of California, Berkeley, to Oliver P. John. We wish to thank the Center for Survey Research and Methodologies (ZUMA) for making possible a guest professorship in Mannheim for Oliver P. John to help work on this project. We are grateful to Anna Maria Baltes, Michael Bosnjak, James Gross, Chris Soto, and Sanjay Srivastava for their help with data collection and analyses, and to Sam Gosling for spirited discussion and helpful comments on an earlier draft.

* Corresponding author. Fax: +496211246100.

E-mail address: rammstedt@zuma-mannheim.de (B. Rammstedt).

1. Introduction

The Big Five Inventory (BFI) was constructed in the late 1980s (John, Donahue, & Kentle, 1991) as an extremely short instrument. At that time, it seemed quite radical to suggest that 44 short-phrase items, answered in about 5 min response time, were sufficient to measure the Big Five dimensions. Most instruments (Goldberg, 1992) then in use were much longer; even the short form of the NEO-PI-R (Costa & McCrae, 1992) included 60 items.

But time has changed. What then seemed radically short now seems tediously long as researchers are faced with limited assessment time; in fact, there has been an accelerating trend towards shorter and shorter personality instruments. The demand for super-short measures is growing, and even researchers using the BFI are asking for a shorter version. Examples of this trend toward minimal measurement are the single-item self-esteem scale (Robins, Hendin, & Trzesniewski, 2001), single-item ability ratings (Rammstedt & Rammseyer, 2002), and even a 10-item measure of the Big Five (Gosling, Rentfrow, & Swann, 2003b). As Burisch (1997) predicted, many of these super-short instruments show respectable psychometric characteristics, suggesting that a shorter version of the BFI may be feasible.

We began with the existing and well-proven 44 items from the BFI and asked whether that item set could be abbreviated to 10 items, with just 2 items per scale. To guard against capitalization on chance when selecting the “best” two items in any one sample or measurement context, and to make the resulting measure more useful, we broadened its generalizability in two ways: we used multiple samples and required that the short scales had to hold up not only in US samples but also in another language-and-culture context, namely Germany. Our results focus on the psychometric characteristics of the 2-item abbreviated BFI-10 scales, including their part-whole correlations with the full BFI scales, retest reliability, structural validity, convergent validity with the NEO-PI-R and its facets, and external validity using peer ratings. In each analysis, we emphasize results across samples and cultures. Moreover, we compare the results for the abbreviated scales directly with those for the full BFI. In the Discussion, we summarize our results and compare them with those for other short measures in the literature.

2. Method

2.1. Participants

The first US sample (US-1) consisted of 726 students (68% females; mean age = 21 years) at a large public university, whereas the second (US-2) consisted of 726 students at a private university (56% females; mean age = 18). The first German sample (G-1) consisted of 457 students (56% females; mean age = 25) and the second (G-2) of 376 students (66% females; mean age = 24). A third US sample (US-3), originally collected by Gosling, Kwan, and John (2003a), consisted of dog owners ($N = 75$) who rated themselves on the BFI and were rated by a friend, thus providing peer rating data to examine external validity.

2.2. Measures: BFI-44, the abbreviated 10-item short version (BFI-10), and the NEO-PI-R

The standard BFI (John & Srivastava, 1999; John et al., 1991; reprinted in Benet-Martínez & John, 1998) consists of 44 short-phrase items, rated on a five-step scale from

1 = “disagree strongly” to 5 = “agree strongly”. The items were selected using both consensual expert judgment and empirical item analyses to represent the core (i.e., most prototypical) traits that define each Big Five domain (see John, 1989, 1990). The BFI was carefully translated and adapted to German (Rammstedt & John, 2006), and the German BFI has psychometric properties similar to the original (see also Lang, Lüdtke, & Asendorpf, 2001).

We selected 2 BFI items for each Big Five dimension following five criteria: (1) We represented both the high and low pole of each factor, so that each BFI-10 scale would consist of one true-scored and one false-scored item. (2) We covered as broad a bandwidth as possible for each scale by selecting two items that both measured core aspects of a Big Five dimension but were not highly redundant in content. (3) We constructed identical English-language and German-language versions, so that the resulting instrument would be usable for cross-cultural research and to minimize capitalizing on chance. (4) To the extent that there still were item choices to be made, we selected items on the basis of two empirical criteria, namely their corrected item-total correlations with the full BFI scales (thus favoring more central over more peripheral item content) and the simple-structure pattern of their loadings in factor analyses of all 44 items (thus favoring items related uniquely to one factor and not to the other four factors). The final ten items are shown in the Appendix A.

To examine convergent validity, we also used the English NEO-PI-R (Costa & McCrae, 1992) as well as its German adaptation (Ostendorf & Angleitner, 2004). In our student samples, α reliabilities for the 48-item long domain scales averaged .85; α s for the 12-item facet scales were, as expected, lower (mean $\alpha = .75$). The NEO-PI-R was intended to measure the Five-Factor Model rather than the lexical “Big Five” tradition (e.g., Goldberg, 1992) that underpins the development of the BFI. Thus, the NEO-PI-R differs somewhat from the BFI (and Goldberg’s, 1992, measures) in how the constructs are defined, especially for Openness, Agreeableness, and Extraversion (see John & Srivastava, 1999).

2.3. Procedure: Self-reports, retest, NEO-PI-R, and peer ratings in five samples

The US-1, US-2, US-3, and G-1 samples all provided self-reports on the full BFI, allowing us to score both BFI-10 and full BFI-44 scales. To test whether the results for the BFI-10 replicate when the 10 items are not embedded in the full BFI, participants in the G-2 sample were administered only a subset, rather than the full 44-item set. To assess retest reliability, a subsample of US-1 completed the BFI a second time 8 weeks later ($N = 178$), and a subsample of G-1 completed the BFI again 6 weeks later ($N = 57$). NEO-PI-R data were available for subsamples of US-1 ($N = 233$) and G-1 ($N = 184$). Finally, to examine external validity, we used peer ratings as validity criteria in three subsamples: 231 US-1 participants were rated by a friend who knew them well; 158 G-1 participants were rated by a dating partner; and all 75 dog owners in US-3 were rated by a friend or partner.

3. Results and discussion

3.1. Generalizability across items: How well do the BFI-10 scales represent the full scales?

The most crucial question is how well the 2-item scales on the BFI-10 can stand in for the full BFI-44 scales—that is, how well do they generalize to the full scales they were designed to represent. Table 1 presents the part-whole correlations of the short scales with the full scales in the three large samples that completed the full BFI. Results indicate

Table 1

How well can 2-item scales represent the standard 9-item BFI scales? Part-whole correlations of the BFI-10 with the BFI-44 scales, test–retest stability, and self-peer external validity correlations in US and German samples

BFI-10 scales	Part-whole correlations				Test–retest stability			Self-peer convergent validity correlations			
	US-1	US-2	G-1	Mean	US-1	G-1	Mean	US-1	US-3	G-1	Mean
Extraversion	.87	.90	.90	.89	.79	.87	.83	.46	.65	.59	.57
Agreeableness	.74	.78	.70	.74	.69	.66	.68	.29	.44	.46	.40
Conscientiousness	.84	.77	.83	.82	.70	.83	.77	.43	.26	.44	.38
Neuroticism	.88	.85	.86	.86	.76	.71	.74	.36	.30	.45	.37
Openness	.79	.78	.80	.79	.65	.78	.72	.45	.44	.45	.45
Mean BFI-10	.83	.82	.83	.83	.72	.78	.75	.40	.43	.48	.44
Mean BFI-44	—	—	—	—	.83	.85	.84	.53	.52	.62	.56

Note. US-1 and US-2 refer to the first two US samples, whereas G-1 is the first German sample, as described in the Method section. Retest, retest correlations across an 8-week interval in the US-1 sample and a 6-week interval in the G-1 sample. Mean correlations computed via Fisher *r*-to-*Z* transformation.

substantial correlations in both the US and the German samples; the overall mean correlation was .83 (as computed with Fisher's *r*-to-*Z* transformation, as were all further computations using correlations). That is, although the BFI-10 scales include less than 25% of the full BFI-44 scales, they predicted almost 70% of the variance of the full scales. Table 1 also shows that the BFI-10 scales differed in their part-whole correlations, just as one would expect from the internal consistency values of the original BFI-44 scales from which they were derived. The three most homogeneous BFI-44 scales, Extraversion, Neuroticism, and Conscientiousness, were best represented by their 2-item versions (average correlations of .89, .86, and .82, respectively). The two least homogeneous BFI-44 scales were least well represented, Agreeableness (.74) and Openness (.79); although their correlations in the mid-to-high .70s look respectable, it is important to note that in variance terms, these two BFI-10 scales lost 45 and 38% of the variance of the full scales, illustrating that using these abbreviated scales come at a cost.

3.2. Generalizability across time: Test–retest stability

How generalizable are scores on these brief scales over time? Table 1 shows test–retest correlations for the BFI-10 scales in the two retest samples. Mean retest stability coefficients were .72 in US-1, .78 in G-1, and .75 overall, suggesting that the BFI-10 scales achieved respectable levels of stability over 6–8 weeks in both cultures. In comparison, the temporal stabilities of the full BFI-44 scales averaged .84. In variance terms, then, the average BFI-44 scale had 71% stable variance, whereas the average BFI-10 scale had 56% stable variance, a difference of 15%. Again, the scales differed somewhat from each other; as shown in Table 1, Extraversion, Conscientiousness, and Neuroticism showed greater stability, and Agreeableness and Openness somewhat less stability.

3.3. Structural validity: Intercorrelations among the scales and item factor analysis

There has been concern about the intercorrelations among the Big Five dimensions (e.g., Block, 1995); the highest intercorrelations among the NEO-PI-R domain scales exceed .40 (Costa & McCrae, 1992), as do those among Goldberg's (1992) Big Five adjective scales. The BFI-10 scales, however, proved to be quite independent in both US and German

samples; mean intercorrelations in our samples ranged from .08 to .13 and averaged .11. Not a single correlation even reached an absolute value of .25, and only four of the total of 40 coefficients even exceeded .20. These extremely low intercorrelations provide strong evidence of discriminant validity and compare favorably with the full BFI-44 scales, which showed an overall mean intercorrelation of .21.

To test whether the Big Five structure could be replicated in this abbreviated item set, we used common-factor analysis and found the expected five-factor structure in each of our four samples. Moreover, the loadings of the 10 items on the five varimax-rotated factors showed clear simple-structure solutions in all US and German samples, with substantial loadings on the one expected or convergent factor (mean loading = .64) and negligible secondary loadings on the four other factors (mean = .08). This pattern of mean loadings was virtually the same as we found for the full BFI-44 (.63 and .10, respectively).

3.4. Convergent validation: Correlations with the NEO-PI-R domain and facet scales

Table 2 shows the correlations with the NEO-PI-R. Differences in construct definitions (see Section 2) necessarily limit the absolute size of these correlations and introduce some apparent discriminant validity issues. Thus, the *pattern* of the correlations is of greater importance here than their absolute size. Overall, the convergent validity correlations with the NEO-PI-R domain scales averaged .67 across Big Five domains and samples, as compared to .78 for the full BFI-44 scales, indicating a loss in convergent validity of $r = .23$. In variance terms, the brief BFI-10 scales share 45% of their variance with the NEO-PI-R domain scales, whereas the full BFI-44 scales share 61%. Comparing the BFI-10 scales to each other, convergent validity with the NEO-PI-R domain scales was highest for Extraversion, Neuroticism, and Conscientiousness, and somewhat lower for Openness and Agreeableness. The same ordering held for the BFI-44 scales (cf. Rammstedt & John, 2006), suggesting that the lower values for Openness and Agreeableness were not caused by the particular item selection for the BFI-10 but rather reflect conceptual differences between the NEO-PI-R and the BFI in their definitions of these two constructs.

To examine the breadth and content of the BFI-10 scales, Table 2 also shows their correlations with the 6 facet scales defining each NEO-PI-R domain as well as the mean of these six facet correlations. Overall, as shown by the convergent validity correlations, each BFI-10 scale correlated substantially with the relevant NEO-PI-R facet scales; of these 60 correlations (30 facets times 2 samples), all but one were significant at $p < .01$, and 50 (83%) exceeded .30. Overall, they averaged .48 for the BFI-10, as compared to .56 for the full BFI-44. This loss in convergent validity amounts to $r = .11$ or a drop from 31 to 23% of variance.

Again, there were some differences among the BFI-10 scales. The correlations of the BFI-10 Extraversion scale with the six NEO-PI-R Extraversion facets ranged from .33 to .72 and averaged .52. Similarly comprehensive construct coverage was also found for Neuroticism (only Impulsiveness showed smaller correlations) and Conscientiousness (only Deliberation showed smaller correlations). Not surprisingly, the BFI-10 Agreeableness scale had the lowest correlations with the NEO-PI-R facets, averaging .38. For the full BFI-44 Agreeableness scale, that correlation was .45, so the drop in the average correlation was not unusually large but the individual facet correlations showed a more complex pattern: whereas BFI-10 Agreeableness correlated on average .63 with the Trust facet (similar to .60 for the BFI-44), the correlations with Altruism and Compliance (both .40) were lower than for the BFI-44 (.63 and .55, respectively), suggesting that BFI-10 Agreeableness

Table 2

Convergent validity: correlations of the BFI-10 with the NEO-PI-R domain and facet scales in US and German samples

NEO-PI-R domain and facet scales	BFI-10 scales									
	Extraversion		Agreeableness		Conscientiousness		Neuroticism		Openness	
	US-1	G-1	US-1	G-1	US-1	G-1	US-1	G-1	US-1	G-1
<i>Extraversion</i>	.69	.79	.24	.11	.29	.08	-.24	-.34	.20	.27
Warmth	.50	.53	.36	.34	.24	.08	-.08	-.23	.14	.22
Gregariousness	.54	.55	.16	.11	.08	-.03	-.06	-.16	.05	.09
Assertiveness	.57	.72	.00	-.13	.26	.02	-.35	-.39	.09	.21
Activity	.44	.58	.08	.01	.42	.33	-.09	-.12	.16	.29
Excitement-seeking	.33	.34	.04	.04	.00	-.17	-.13	-.07	.18	.05
Positive emotions	.47	.53	.35	.14	.21	.12	-.24	-.41	.20	.24
Facet mean	.48	.55	.17	.09	.21	.06	-.16	-.23	.14	.18
<i>Agreeableness</i>	.02	-.05	.51	.65	.18	.14	.03	-.07	.05	-.06
Trust	.24	.25	.62	.64	.21	.06	-.23	-.22	.06	.05
Straightforwardness	-.10	-.15	.25	.34	.19	.20	.10	.00	-.03	-.08
Altruism	.19	.18	.35	.44	.28	.07	-.02	-.11	.14	.06
Compliance	-.13	-.21	.37	.43	.00	.04	.00	-.10	-.08	-.08
Modesty	-.18	-.28	.19	.22	-.03	.05	.17	.19	-.03	-.19
Tender-mindedness	.08	.04	.25	.30	.08	.08	.09	.02	.17	.00
Facet mean	.02	-.03	.35	.40	.12	.08	.02	-.04	.04	-.04
<i>Conscientiousness</i>	.09	.20	.15	.02	.70	.70	-.14	-.25	.04	.08
Competence	.22	.32	.13	.05	.56	.37	-.34	-.47	.13	.18
Order	-.02	.03	.02	.01	.42	.44	.06	-.07	.01	-.04
Dutifulness	.03	.10	.21	.05	.56	.59	-.06	-.08	.03	.02
Achievement striving	.18	.27	.06	-.01	.61	.60	-.04	-.08	.10	.18
Self-discipline	.15	.24	.21	.12	.62	.70	-.22	-.30	.01	.12
Deliberation	-.13	-.08	.06	-.12	.36	.22	-.03	-.07	-.07	-.11
Facet mean	.07	.15	.12	.02	.53	.50	-.11	-.18	.04	.06
<i>Neuroticism</i>	-.22	-.31	-.35	-.10	-.28	-.05	.73	.71	-.02	-.04
Anxiety	-.13	-.21	-.23	-.06	-.06	.01	.72	.68	.01	-.01
Angry hostility	-.06	-.06	-.47	-.30	-.10	-.04	.43	.56	-.02	.06
Depression	-.26	-.36	-.29	-.05	-.17	-.07	.61	.54	-.02	-.09
Self-consciousness	-.36	-.46	-.22	-.09	-.09	.04	.56	.59	-.07	-.14
Impulsiveness	.06	.02	-.17	.00	-.17	-.13	.27	.26	.12	.19
Vulnerability	-.22	-.31	-.22	-.01	-.20	-.09	.70	.67	-.10	-.15
Facet mean	-.16	-.24	-.27	-.09	-.13	-.05	.57	.56	-.01	-.02
<i>Openness</i>	.24	.26	.06	.17	.07	-.02	-.12	-.17	.63	.61
Fantasy	.14	.13	.02	.10	-.16	-.20	-.05	.02	.57	.58
Aesthetics	.16	.17	.08	.22	.14	.10	-.01	-.05	.58	.56
Feelings	.25	.26	.02	.12	.13	.08	.10	.06	.38	.46
Actions	.20	.23	.08	.08	.06	.02	-.18	-.26	.33	.32
Ideas	.15	.20	.01	.00	.09	.01	-.25	-.32	.44	.33
Values	.14	.09	.05	.19	.00	-.10	-.11	-.13	.28	.09
Facet mean	.17	.18	.04	.12	.04	-.02	-.08	-.12	.44	.40

Note. BFI, Big Five Inventory; NEO-PI-R, NEO Personality Inventory Revised. Correlations of .30 or higher are set in bold. Correlations larger than .11 in US-1 and .18 in G-1 were significant at the $p < .01$ level. Correlations for the five NEO-PI-R domain scales are set in italics to differentiate them from the facet scales and their means.

provides less broad construct coverage than the full BFI-44. Like the BFI-44, BFI-10 Agreeableness did correlate with Warmth (.53 and .35, respectively) which the NEO-PI-R includes in the Extraversion domain. For BFI-10 Openness, the facet correlations averaged .42; here the correlations with Openness to Fantasy (.58) and Aesthetics (.57) were similar to the full BFI-44 (.50 and .63, respectively) whereas the correlations with Openness to Ideas and Actions were lower (.39 and .33 vs. .62 and .48), suggesting that imagination and artistic interests are most central to the abbreviated scale. In sum, the BFI-10 scales showed substantial convergent and discriminant validity; with some exceptions for Agreeableness and Openness, the pattern of correlations for the BFI-10 was generally similar to that for the BFI-44.

3.5. External validation: Correlations with peer ratings

One central way to evaluate construct validity is to demonstrate generalizability to ratings by knowledgeable informants. Table 1 reports convergent validity correlations between the self-report and peer-report BFI-10 scales. Overall these correlations averaged .44, as compared to .56 for the BFI-44, which represents a loss of $r = .16$ in external validity. Moreover, the convergent correlations for the BFI-10 were always much higher than the discriminant correlations, which averaged .09 in absolute values. Even the highest discriminant correlations were modest in size: .19 in US-1, .24 in US-2 and .18 in G-1. Thus, reducing the number of BFI items to less than one fourth did lower external validity, but convergent validity remained substantial and discriminant validity excellent.

4. General discussion

How does the BFI-10, which uses established items, compare to the Ten Item Personality Inventory (TIPI; Gosling et al., 2003b), which introduced new adjectival items based on a review of the existing Big Five instruments? The retest reliabilities across six weeks averaged .72 for the TIPI, similar to .75 for the BFI-10. In terms of discriminant validity (Gosling, 2004, personal communication) the absolute intercorrelations among the TIPI scales averaged .20, larger than the .11 for the BFI-10; two TIPI intercorrelations even exceeded .30, namely Extraversion and Openness (.36) and Agreeableness and Neuroticism (–.31). Thus, not surprisingly, when we factor analyzed the item intercorrelations reported by Gosling et al. (2003b), the expected five-factor structure did not emerge, whereas the BFI-10 always showed a clear five-factor structure. Convergent validity with the NEO-PI-R domains averaged .63 for the TIPI and .67 for the BFI-10, even though the BFI-10 correlations were likely attenuated in the German sample due to the use of translated instruments. Self-peer validity information is not yet available for the TIPI; for the BFI-10, correlations with peer ratings revealed good external validity coefficients. Overall, then, our results for the BFI-10 were at least as good as those for the TIPI, suggesting that new items are not necessary to obtain a reliable and valid Big Five measure. Moreover, in comparing the results for the two measures, it is important to realize that we averaged across both English and German versions of the BFI-10. It remains to be seen, for example, how well a German translation of the TIPI will converge with the German version of the NEO-PI-R. Nonetheless, future research should compare the two instruments directly, using multiple samples and language contexts.

How much reliability and validity is lost by reducing the BFI scales to just 2 items? Overall, effect sizes and convergences with the BFI-44 generalized across multiple samples

and two languages. On average, the BFI-10 scales captured 70% of the full BFI variance and retained 85% of the retest reliability. Discriminant and structural validity, however, remained essentially the same. Validity correlations with the NEO-PI-R stayed substantial but the averaged loss in convergent validity from BFI-44 to BFI-10 amounted to $r = .23$. Correlations with the NEO-PI-R facet scales revealed generally good bandwidth of the five BFI-10 scales, with a mean validity loss of $r = .11$. Self-peer validity correlations were still substantial but reduced from $.56$ to $.44$, a loss of $r = .16$. Overall, these findings show that the BFI-10 retains a substantial portion of the reliability and validity of the original BFI-44 and thus support for the construct validity of the BFI-10.

However, the losses are also noticeable and were most substantial for the BFI-10 Agreeableness scale. What should researchers do if they need to use the BFI-10 and for whom the Agreeableness construct is crucial? For these contexts, we recommend to add a third Agreeableness item, namely “Is considerate and kind to almost everyone.” The resulting 3-item scale increased the part-whole correlation with the BFI-44 from $.74$ to $.81$ (now representing 66% of the BFI-44) and the retest correlation from $.68$ to $.70$. Validity also increased: the correlation with the NEO-PI-R Agreeableness scale went from $.58$ to $.63$, and the representation of Altruism improved, as the correlation with this NEO-PI-R facet went from $.40$ to $.52$; even external validity increased from $.40$ to $.50$.

In conclusion, results from multiple samples and for two languages, namely English and German, suggest that, given its brevity, the BFI-10 possesses acceptable psychometric properties. However, there were substantial losses in comparison to the full-scale BFI. Thus, if testing time is not extremely limited, full-length Big Five measures possess clear psychometric advantages. That is, we agree with Gosling et al. (2003b) that ultra-short measures should not and cannot be used as substitutes for regular personality assessments. Only for research settings in which participant time is truly limited and when personality assessment would otherwise be impossible, such as in telephone surveys, the BFI-10 offers an adequate assessment of personality.

Appendix A. Big Five Inventory-10 (BFI-10)

English version.

Instruction: How well do the following statements describe your personality?

I see myself as someone who ...	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
... is reserved	(1)	(2)	(3)	(4)	(5)
... is generally trusting	(1)	(2)	(3)	(4)	(5)
... tends to be lazy	(1)	(2)	(3)	(4)	(5)
... is relaxed, handles stress well	(1)	(2)	(3)	(4)	(5)
... has few artistic interests	(1)	(2)	(3)	(4)	(5)
... is outgoing, sociable	(1)	(2)	(3)	(4)	(5)
... tends to find fault with others	(1)	(2)	(3)	(4)	(5)
... does a thorough job	(1)	(2)	(3)	(4)	(5)
... gets nervous easily	(1)	(2)	(3)	(4)	(5)
... has an active imagination	(1)	(2)	(3)	(4)	(5)

German version.

Instruction: Inwieweit treffen die folgenden Aussagen auf Sie zu?

Ich...	trifft überhaupt nicht zu	trifft eher nicht zu	weder noch	eher zutreffend	trifft voll und ganz zu
...bin eher zurückhaltend, reserviert.	(1)	(2)	(3)	(4)	(5)
...schenke anderen leicht Vertrauen, glaube an das Gute im Menschen.	(1)	(2)	(3)	(4)	(5)
...bin bequem, neige zur Faulheit.	(1)	(2)	(3)	(4)	(5)
...bin entspannt, lasse mich durch Stress nicht aus der Ruhe bringen.	(1)	(2)	(3)	(4)	(5)
...habe nur wenig künstlerisches Interesse.	(1)	(2)	(3)	(4)	(5)
...gehe aus mir heraus, bin gesellig.	(1)	(2)	(3)	(4)	(5)
... neige dazu, andere zu kritisieren.	(1)	(2)	(3)	(4)	(5)
...erledige Aufgaben gründlich.	(1)	(2)	(3)	(4)	(5)
...werde leicht nervös und unsicher.	(1)	(2)	(3)	(4)	(5)
...habe eine aktive Vorstellungskraft, bin phantasievoll.	(1)	(2)	(3)	(4)	(5)

Scoring the BFI-10 scales:

Extraversion: 1R, 6; Agreeableness: 2, 7R; Conscientiousness: 3R, 8; Neuroticism: 4R, 9; Openness: 5R; 10 (R = item is reversed-scored).

Optional additional Agreeableness item (true-scored):

English version	(1)	(2)	(3)	(4)	(5)
...is considerate and kind to almost everyone.	(1)	(2)	(3)	(4)	(5)
German version					
...bin rücksichtsvoll zu anderen, einfühlsam.	(1)	(2)	(3)	(4)	(5)

References

- Benet-Martínez, V., & John, O. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75, 729–750.
- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117, 187–215.
- Burisch, M. (1997). Test length and validity revisited. *European Journal of Personality*, 11, 303–315.
- Costa, P. T., & McCrae, R. R. (1992). *NEO-PI-R Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Goldberg, L. R. (1992). The development of markers for the Big five factor structure. *Psychological Assessment*, 4, 26–42.
- Gosling, S. D., Kwan, V. S. Y., & John, O. P. (2003a). A dog's got personality: A cross-species comparative approach to evaluating personality judgments. *Journal of Personality and Social Psychology*, 85, 1161–1169.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003b). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37, 504–528.
- John, O. P. (1989). Towards a taxonomy of personality descriptors. In D. M. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 261–271). New York: Springer-Verlag.

- 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). New York: Guilford Press.
- 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., & 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). New York: Guilford Press.
- Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. [Validity and psychometric equivalence of the German version of the Big Five Inventory in young, middle-aged and old adults]. *Diagnostica*, 47, 111–121.
- Ostendorf, F., & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae, revidierte Form (NEO-PI-R)*. [NEO Personality Inventory based on Costa and McCrae, revised version (NEO-PI-R)]. Göttingen: Hogrefe.
- Rammstedt, B., & John, O. P. (2006). *Die großen Fünf: Nomological net and cultural differences in the U.S. and Germany*. Manuscript in preparation.
- Rammstedt, B., & Rammsayer, T. (2002). Gender differences in self-estimated intelligence and their relation to gender-role orientation. *European Journal of Personality*, 16, 369–382.
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg self-esteem scale. *Personality and Social Psychology Bulletin*, 27, 151–161.