

Psychometric Properties of the BFI-K: A Cross-Validation Study

Anastassiya Kovaleva

Department of Didactics of Biology, University of Bielefeld, Germany

Constanze Beierlein

GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany

Christoph J. Kemper

*Institute for Medical and Pharmaceutical Proficiency Assessment, Mainz,
Germany*

Beatrice Rammstedt

GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany

Abstract

In recent years, the BFI-K, a short version of the well-established Big Five Inventory (BFI), has been employed in a number of studies investigating samples from heterogeneous populations. As the BFI-K had been validated only in homogeneous student samples, cross-validation in an extended, heterogeneous sample was called for. The present study ($N = 5,178$) aimed to meet this demand. Factor analysis, which was carried out using Exploratory Structural Equation Modelling (ESEM), validated the factorial structure. A three-factor MANOVA conducted with the variables age, gender, and education replicated the well-known effects of sociodemographic variables on the Big Five. Hence, our results reveal that the psychometric properties of the BFI-K scales are in line with those of the initial validation studies.

Keywords: Big Five, BFI-K, short scale, Exploratory Structural Equation Modelling (ESEM), sociodemographic variables

Introduction

During the last decades, the Big Five have become well-established and widely accepted dimensions of personality description (e. g., Barrick, Mount, & Judge, 2001; Caspi, Roberts, & Shiner, 2005; Credé, Harms, Niehorster, & Gaye-Valentine, 2012; De Raad, 2000; John, Naumann, & Soto, 2008; Lang, John, Lüdtke, Schupp, & Wagner, 2011; Marsh et al., 2010). This high level of acceptance is evidenced by the popularity of these five dimensions not only in personality psychology but also in many other scientific fields (e. g., Hurtz & Donovan, 2000; Lang & Lüdtke, 2005). For example, the Big Five have been shown to be valid predictors of survey participation (Rogelberg, Conway, Sederburg, Spitzmüller, Aziz, & Knight, 2003), political party preferences, political attitudes, and voting behaviour (e. g., van Hiel, Kossowska, & Mervielde, 2000; Kunz, 2005). Empirical evidence of the predictive ability of the Big Five were also found for career success and job performance (e. g., Barrick et al., 2001; Hurtz & Donovan, 2000; Salgado, 1997; Tett, Jackson, & Rothstein, 1991), and for subjective well-being (e. g., DeNeve & Cooper, 1998; Haslam,

Whelan, & Bastian, 2009; Rammstedt & Schupp, 2008; Robinson, Solberg, Vargas, & Tamir, 2003).

There are many well-established instruments for the measurement of the Big Five, e. g., the NEO-Personality Inventory (NEO-PI-R, Costa & McCrae, 1985, 1992; Ostendorf & Angleitner, 2004), the NEO-Five Factor Inventory (NEO-FFI, Costa & McCrae, 1992; Borkenau & Ostendorf, 2008) and the Big Five Inventory (BFI; John, Donahue, & Kentle, 1991; Lang, Lüdtke, & Asendorpf, 2001; Rammstedt, 1997). However, for researchers outside the core field of personality psychology these instruments are too lengthy. These researchers are primarily interested in correlation analyses, i. e., in the impact of the personality dimensions on dependent variables such as voting behaviour or career success. In such studies the Big Five are usually just one part of an extensive questionnaire battery that comprises several potential predictors. As a result, the time available for the measurement of each individual construct is often limited. This is the case in population surveys, for example, where each response is viewed in terms of what it costs to collect. Temporal or financial constraints prompt most researchers in these fields either to opt for a short but crude measure of personality or to forgo the collection of personality data altogether. For the above-mentioned reasons, studies outside the core field of personality research are unable to accommodate Big Five measures comprising of sometimes more than 200 items as in the NEO-PI-R. But also shorter measures of 45 to 60 items are still too lengthy to be of use in their research contexts. Thus, there is a clear demand for substantially shorter Big Five measures.

In recent years, several short scales for the measurement of the Big Five have been developed to meet this demand (e.g., Aronson, Reilly, & Lynn, 2006; Bernard, Walsh, & Mills, 2005; Donnellan, Oswald, Baird, & Lucas, 2006; Herzberg & Brähler, 2006; Gosling, Rentfrow, & Swann, 2003; Rammstedt 2007; Rammstedt & John, 2005, 2007; Rammstedt, Koch, Borg & Reitz, 2004; Woods & Hampson, 2005). These instruments range in length from 5 items – where each domain is measured by only one item – to 21 items. Due to their efficiency, these short Big Five scales have already been used in a number of studies, including several surveys such as the International Social Survey Programme (ISSP), the German Longitudinal Election Study (GLES), and the German Socio-Economic Panel (GSOEP).

Because reliability and validity inevitably decrease when the number of items is reduced, short scales face the problem of lower psychometric properties compared to longer scales. Of the short scales mentioned above, the BFI-K is the most comprehensive one and possesses the highest psychometric qualities. Although each domain is measured by at least four items, it still enables the Big Five to be assessed in less than two minutes (cf. Rammstedt & John, 2005). The efficiency of the BFI-K and its comparatively good psychometric properties have made it the instrument of choice for several studies during the last few years. These studies include investigations into the impact of personality and political skills on job performance (Blickle et al., 2008), the role of forgiveness in action orientation (Allemand, Job, Christen, & Keller, 2008), and the dimensional structure of personality, ideological beliefs, social attitudes, and personal values (von Collani & Grumm, 2009).

The BFI-K was initially validated on the basis of two samples consisting of university students (cf. Rammstedt & John, 2005). This is common practice in

psychological research. Several of the well-established Big Five scales were primarily validated in student samples, e.g., the BFI (Benet-Martinez & John, 1998), the Mini-IPIP (Donnellan et al., 2006), and the TIPI (Gosling et al., 2003). Even though this approach is far accepted it clearly lacks the investigation of the generalizability to other populational groups. Student samples are highly homogeneous with regard to age and education. Coefficients of the psychometric quality of a scale depend on the sample used and are affected by restrictions of sample variance (Nunnally & Bernstein, 2004). Thus, it cannot be taken for granted that coefficients calculated for student samples equally apply to samples from the general population. For example, Rammstedt, Goldberg, and Borg (2010; Rammstedt & Kemper, 2011) could demonstrate that the Big Five factor structure is highly sensitive to educational effects. Therefore, scales developed to assess a construct in the general population need to be validated in samples comparable to this general population and thus in samples more heterogeneous than student samples.

As an increasing number of studies assesses the BFI-K in non-student samples (Blickle et al., 2008, von Collani & Grumm, 2009), evidence for the construct validity of the scale for the general population is necessary. In order to ensure the applicability of the BFI-K in studies that target the general population, key psychometric properties of the BFI-K must be calculated based on samples that are more representative of the general population especially with regard to the key sociodemographic variables gender, age, and education. The present study aims to address this issue by testing two important aspects of the construct validity of the BFI-K in a large and heterogeneous sample - (1) the factorial validity of the BFI-K and (2) the criterion (known-groups) validity of BFI-K scales. The factorial validity of the BFI-K has already been demonstrated for university students (Rammstedt & John, 2005). However, it has not been demonstrated for heterogeneous samples so far. To test the factorial validity of the BFI-K, a novel multivariate method was applied - Exploratory Structural Equation Modeling (ESEM; Asparouhov & Muthén, 2009). To test the criterion (known groups) validity of the BFI-K scales, the sample was divided into groups based on the most important variables to describe a population's age, gender, and education. Applying a MANOVA design, the effects on group means were investigated and compared to corresponding results of previous studies (e.g., Costa, Terracciano, & McCrae, 2001; Feingold, 1994; Goldberg, Sweeney, Merenda, & Hughes, 1998; Rammstedt, 2007; Srivastava, John, Gosling, & Potter, 2003; Viken, Rose, Kapiro, & Koskenvuo, 1994 etc.).

This study thus aims to benefit the usage of the BFI-K in wider research contexts as it can be done at present.

Method

Sample and Procedure

The BFI-K was employed within the framework of an online study conducted by Respondi AG, a commercial provider of sampling services. Participants were recruited by means of Web-based advertising. Participation was financially rewarded. The data for the present study was collected from the sample comprising 5,178 respondents ranging in age between 16 and 77 years ($M = 32.9$, $SD = 11.1$). Respondents under the age of 18 years were excluded

from our analyses ($N = 15$), thereby yielding a reduced sample of $N = 5,163$. Sociodemographic characteristics of the sample divided by age (18 to 29, 30 to 44, 45 to 59, 60 years or above), gender (male, female), and education (lower or no secondary education, intermediate secondary education, entrance qualification for universities) are depicted in Table 1.

Table 1
Sociodemographic characteristics of the sample ($N = 5,163$)

	Total Sample			Male Subsample		Female Subsample	
	Abs.	%	% German Microcensus 2007	Abs.	%	Abs.	%
<i>Gender</i>							
Male	2382	46.1	48.4	2382	100		
Female	2781	53.9	51.6			2781	100
<i>Age</i>							
18-29 years	2491	48.2	20.5	1080	45.3	1411	50.7
30-44 years	1826	35.4	25.5	891	37.4	935	33.6
45-59 years	733	14.2	24.4	343	14.4	390	14.1
> 60 years	113	2.2	29.6	68	2.9	45	1.6
<i>M (SD)</i>	32.9	(11.1)	42.5 (22.3)	33.5	(11.5)	32.3	(10.7)
<i>Level of education¹</i>							
Low	565	10.9	48.7	269	11.3	296	10.6
Medium	1653	32.0	27.6	700	29.4	953	34.3
High	2945	57.0	23.7	1413	59.3	1532	55.1
Total	5163	100.0	100.0	2382	100.0	2781	100.0

Note. Abs.= absolute frequency, % = relative frequency. ¹Low level of education - lower or no secondary education; medium level of education - intermediate secondary education; high level of education - entrance qualification for the universities of applied sciences and general universities.

Questionnaire

The BFI-K (Rammstedt & John, 2005), which consists of 21 items, is a short version of the well-established 44-item Big Five Inventory (BFI; John et al., 1991; Lang et al., 2001; Rammstedt, 1997). Previous research (Rammstedt & John, 2005) has demonstrated that the BFI-K possesses sufficient reliability coefficients and factorial validity. Moreover, several indicators of construct validity such as convergence coefficients of self-reports with partner ratings and with other inventories assessing the Big Five were all shown to be substantial and comparatively high. The items of the BFI-K are rated on a five-point Likert scale ranging from 'strongly disagree' (1) to 'strongly agree' (5). Four scales consist of four items whereas Openness comprises five items. Standardized normal values were calculated by weighting the present sample to the German Microcensus2007 distribution of gender, age, and education presented in

Table 1. These values are available as supplementary online material at: <http://www.gesis.org/kurzskalen-psychologischer-merkmale/supplemental-online-material>

Statistical Analysis

Different aspects of the psychometric quality of the BFI-K were evaluated. (1) At first, descriptive statistics for all BFI-K scales as well as reliability coefficients (Cronbach's Alpha) were calculated. (2) Afterwards, we tested whether the Big Five factor structure could be replicated with the BFI-K based on a heterogeneous sample (factorial validity). A new approach called Exploratory Structural Equation Modelling (ESEM; Marsh et al., 2010; cf. Asparouhov&Muthén, 2009) was used for the validation of the factor structure. ESEM not only enables factor loadings to be identified as in traditional Exploratory Factor Analysis (EFA), but also provides parameter estimates, standard errors, and goodness-of-fit statistics that are typically associated only with CFA (cf. Lang et al., 2011). Although Confirmatory Factor Analysis (CFA) is regarded as a strong analytical tool in construct validation, several authors question its appropriateness as a mean of evaluating the internal structure of personality inventories (Hopwood &Donnellan, 2010; Marsh et al., 2009, 2010; see also Lang et al., 2011). CFA models generally allow indicators to load on only one specific factor disrupting structures with secondary loadings. In this case, all secondary loadings are constrained to zero. However, McCrae, Zonderman, Costa, Bond, and Paunonen (1996) state that the Five Factors Model does not postulate such a perfect simple structure. Traits may rather have meaningful secondary loadings on several factors. These misgivings prompted us to use the benefits that are suggested with ESEM algorithms in order to maintain the most realistic estimation of the Big Five structure. (3) In a third step, we investigated whether the effects of sociodemographic variables observed in previous studies can be replicated with the BFI-K in a large and heterogeneous sample. A three-factor (age x gender x education) MANOVA was conducted using SPSS 19 with age, gender, and education as independent variables and the BFI-K scales as dependent variables. Overall, 35 tests of significance for main and interaction effects were computed. The alpha level for multiple comparisons was adjusted according to Bonferroni ($p < .00145$).

Results

Descriptive Statistics and Reliability

Means, standard deviations, Cronbach's Alpha coefficients, and skewness and kurtosis of the BFI-K scales are presented in Table 2. For comparison purposes, means and standard deviations of two studies reported by Rammstedt and John (2005) are included in the latter 6 columns of the table.

In the present study, scale means ranged between 2.92 and 3.62, with an overall mean of 3.29 indicating no severe ceiling or bottom effects. Scale means are highly similar to those reported by Rammstedt and John (2005). The standard deviations found in the present study -ranging between .53 and .65 with a mean of .58 - are slightly lower than those reported in the previous validation studies (where they ranged between .62 and .92).

The skewness of the scales varied between $-.26$ and $.11$ with an absolute mean of $.13$; kurtosis ranged from $-.04$ to $.54$, yielding an absolute mean of $.25$, thus indicating, that normal distribution can be assumed for all five scales (cf. Muthén & Kaplan, 1985).

With regard to the scale reliability in the present study, Cronbach's Alpha coefficients ranged between $.58$ for Agreeableness and $.80$ for Extraversion with a mean of $.70$. These results are also highly comparable in size and rank-order to those reported in the initial validation studies.

Table 2

Descriptive statistics and internal consistency of the BFI-K scales for the total sample of the present study and for the initial validation samples

	Present Study(N)					Rammstedt & John (2005)					
	M	SD	α	Skew.	Kurt.	Sample 1 (N_1)			Sample 2 (N_2)		
	M	SD	α	Skew.	Kurt.	M	SD	α	M	SD	α
Extraversion	3.48	.65	.80	-.08	-.22	3.48	.87	.86	3.59	.87	.81
Agreeableness	2.93	.56	.58	-.04	.16	3.02	.73	.64	2.89	.77	.59
Conscientiousness	3.62	.53	.69	-.26	.54	3.53	.69	.70	3.52	.73	.69
Neuroticism	2.92	.57	.74	.11	-.04	2.88	.77	.74	3.12	.92	.77
Openness	3.52	.57	.69	-.14	-.30	3.96	.62	.66	4.02	.64	.70
Mean	3.29	.58	.70	.13	.25	3.37	.74	.72	3.43	.79	.72

Note: $N = 5,163$; $N_1 = 459$; $N_2 = 391$; M = means, SD = standard deviations, α = Cronbach's Alpha, Skew. = skewness, Kurt = kurtosis.

Construct Validity

(1) Factorial validity

The factorial structure of the 21 BFI-K items was investigated using ESEM analysis with oblique rotation. ESEM analysis was conducted with *Mplus* (Version 5.2; Muthén & Muthén, 2008). The ESEM model encompassed five factors and 21 indicators in total. We chose the Maximum Likelihood Robust (MLR) estimator, which is robust against violations of normality assumptions (cf. Lang et al., 2011). Following Marsh et al. (2010, p. 475), we employed an oblique geomin rotation with an epsilon value of $.5$ (cf. Asparouhov & Muthén, 2009). In sum, the fit indices revealed an acceptable fit of the five factor model ($\chi^2/df = 14.76$, $p = .01$; CFI = $.94$; RMSEA = $.05$; SRMR = $.03$; cf. Kline, 2005). The geomin rotated standardized factor loadings are presented in Table 3, with highest loadings per item in bold print. All factors were clearly interpretable in terms of the Big Five. The results show that - except of the two Openness items (No. 18: "...is curious about many different things", No. 20: "...is ingenious, a deep thinker") - all of the items loaded highest on their corresponding factors. With the exception of the two Openness items and the Agreeableness item No. 5 ("...is generally trusting") the factor loading structure

was unambiguous and converged to a rather clear simple structure (cf. Brown, 2006). This is also reflected in the low factor intercorrelations shown in Table 4. These correlations ranged between -.02 (Conscientiousness and Agreeableness) and -.23 (Extraversion and Neuroticism) with a mean of .12.

Table 3
Standardized factor loadings of the BFI-K items (ESEM with geomin rotation)

I see myself as someone who...	Factorloadings				
	E	A	C	N	O
1. ...is outgoing, sociable.	.66	-.01	.18	-.04	.05
2. ...generates a lot of enthusiasm.	.44	-.03	.29	-.01	.18
3. ...tends to be quiet. ^a	.81	.06	-.04	-.05	.03
4. ...is reserved. ^a	.76	.02	-.11	-.13	-.01
5. ...is generally trusting.	.07	.20	.18	.19	.06
6. ...tends to find fault with others. ^a	-.14	.38	.02	-.21	.01
7. ...can be cold and aloof. ^a	.11	.74	-.06	.04	.03
8. ...is sometimes rude to others. ^a	-.04	.81	.03	-.02	.03
9. ...does things efficiently.	.05	-.04	.65	-.08	-.03
10. ...does a thorough job.	-.01	.03	.71	-.02	.01
11. ...makes plans and follows through with them.	.10	-.07	.54	-.12	.06
12. ...tends to be lazy. ^a	.10	.17	.28	-.19	-.01
13. ...gets nervous easily.	-.20	-.03	-.08	.62	-.02
14. ...worrys a lot.	-.03	-.08	.12	.70	.01
15. ...is depressed, blue.	-.11	-.07	-.11	.66	.03
16. ...is relaxed, handles stress well. ^a	.12	-.01	-.22	.46	-.09
17. ...values artistic, aesthetic experiences.	-.04	.01	-.02	.01	.79
18. ...is curious about many different things.	.13	-.03	.38	-.02	.29
19. ...has an active imagination.	.17	-.06	.20	.03	.46
20. ...is ingenious, a deep thinker.	-.08	-.06	.31	.27	.28
21. ...has few artistic interests. ^a	.01	.07	-.15	-.05	.75

Note: N = 5,163; E = Extraversion, A = Agreeableness, C = Conscientiousness, N = Neuroticism, O = Openness. Highest loadings in bold print. Item was inverted.

Table 4
Standardized factor correlations of the BFI-K factors

	E	A	C	N
Agreeableness (A)	.05**			
Conscientiousness (C)	.18**	-.02		
Neuroticism (N)	-.23**	-.18**	-.14**	
Openness (O)	.16**	.02	.22	.02

Note: $N = 5,163$; E = Extraversion.

** $p < .01$.

(2) Mean differences in sociodemographic characteristics

As a second indicator for the construct validity of the BFI-K scales, the impact of the core sociodemographic variables age, gender, and education on the scales was investigated and compared with the results of previous studies in the field (e. g., Costa et al., 2001; Feingold, 1994; Goldberg et al., 1998; Rammstedt, 2007; Srivastava et al., 2003; Viken et al. 1994 etc.). A three-factor MANOVA was conducted using SPSS 19; the sociodemographic variables gender, age, and education were simultaneously considered as independent variables and the five BFI-K scales as dependent variables. Overall, 35 tests of significance for main and interaction effects were computed. The alpha level for multiple comparisons was adjusted according to Bonferroni ($p < .00145$). Table 5 presents the resulting sociodemographic effects. Descriptive statistics are reported for the total sample and for both sexes separately, for the four age groups and the three educational groups described in the Method section above. The multivariate tests yielded significant overall main effects for age (Hotelling's Trace: $F_{(15, 15401)} = 14.64$, $p < .00145$, $\eta^2 = .01$), gender ($F_{(5, 5135)} = 14.16$, $p < .00145$, $\eta^2 = .01$), and educational level ($F_{(10, 10268)} = 3.84$, $p < .00145$, $\eta^2 < .01$). No interaction effects were identified.

Post-hoc tests conducted for all three main effects revealed the following differences: With regard to gender, substantial effects were found for all BFI-K scales with females scoring constantly higher than males. However, the size of the effects was rather small ($\eta^2 < .01$). In line with previous findings (Costa et al., 2001; Feingold, 1994; Rammstedt, 2007; Srivastava et al., 2003; Viken et al. 1994), the strongest gender effect emerged for Neuroticism ($F_{(1, 5139)} = 25.42$, $p < .00145$) followed by Conscientiousness ($F_{(1, 5139)} = 21.18$, $p < .00145$).

Age groups differed significantly in (1) Neuroticism ($F_{(3, 5139)} = 36.92$, $p < .00145$, $\eta^2 = .02$) and (2) Agreeableness ($F_{(3, 5139)} = 20.91$, $p < .00145$, $\eta^2 = .01$), with respondents in the youngest age group (18 to 29 yrs.) scoring higher on Neuroticism and lower on Agreeableness than all other participants, and (3) on Extraversion ($F_{(3, 5139)} = 11.14$, $p < .00145$, $\eta^2 = .01$), with the youngest (18 to 29 yrs.) age group scoring lower than the second age group (30 to 44 yrs.). These findings are also in good agreement with results typically reported in the literature (cf. Caspi et al., 2005; McCrae et al., 1999; Rammstedt, 2007).

With regard to education, a significant effect was found for Openness ($F_{(2, 5139)} = 10.98$, $p < .00145$, $\eta^2 < .01$), with the highest educated group scoring

higher than the intermediate and lower-educated groups. This effect is also in line with results of previous studies (e. g., Goldberg et al., 1998; Rammstedt, 2007; Vassend & Skrandal, 1995).

In sum, results revealed that the **BFI-K** scales show generally the same relation to sociodemographic variables as it is usually reported in the corresponding literature.

Discussion

The wide acceptance of the Big Five for the description of personality on a global level has made personality assessment popular outside the core field of personality research. This has led to an intensified demand for short scales for the measurement of these factors. Numerous short measures ranging in length between 5 and 21 items have been developed to meet this demand. However, most of these instruments have not been validated in the populations of interest to researchers outside the core field of psychology, namely in non-student samples. The **BFI-K** is the most comprehensive of the short Big Five measures. The present study aimed to replicate and extend the initial **BFI-K** validation studies by using a more comprehensive and heterogeneous sample. Item statistics, reliability coefficients, results of an ESEM analysis on the structure of the **BFI-K** and of an analysis of sociodemographic effects were presented. These findings were compared with the results of the initial validation study of the **BFI-K** (Rammstedt & John, 2005).

Results from our large and heterogeneous sample indicate that means and standard deviations of the **BFI-K** scales are comparable in size to those reported in the initial validation studies. Thus, no deviations in scale properties indicating potential bottom or ceiling effects were detectable in the present study. Nor did kurtosis or skewness give any hints of deviations from the expected normal distribution. With regard to reliability, coefficients were found to be highly comparable in size and rank-order to those reported by Rammstedt and John (2005). The internal consistencies (Cronbach's Alphas) of the **BFI-K** scales, especially of Agreeableness, may be considered rather low compared to longer Big Five scales, e.g., the IPIP-FFM (Donnellan et al., 2006). As the **BFI-K** scales contain considerably less items, a lower Cronbach's Alpha is inevitable. However, reliability coefficients may still be considered as sufficient for research purposes, e. g., for the comparison of groups rather than individuals (Aiken & Groth-Marnat, 2006). In addition, McCrae, Kurtz, Yamagata, and Terraciano (2011) recently showed that test-retest reliability, rather than internal consistencies, have a crucial impact on validity coefficients of a scale. Acceptably high test-retest correlations have been demonstrated for **BFI-K** scales in prior research, ranging between .76 and .93 (Rammstedt & John, 2005). Taken together, our results concerning item statistics and reliability are consistent with the initial validation of the **BFI-K** which was based on considerably smaller and more homogenous samples.

Table 5
MANOVA results, descriptive statistics by age, gender, and education separately

	E		A		C		N		O	
	<i>M</i>	<i>SD</i>								
<i>Gender</i>										
Male	3.40	.67	2.81	.54	3.52	.52	2.89	.56	3.48	.60
Female	3.52	.68	2.97	.55	3.68	.52	3.03	.58	3.62	.60
$F_{(1, 5139)}$	12.86*		16.48*		21.18*		25.42*		16.05*	
η^2	.002		.003		.004		.005		.003	
<i>Age</i>										
18-29 years	3.43	.69	2.84	.56	3.61	.55	3.07	.57	3.58	.63
30-44 years	3.52	.66	2.94	.54	3.60	.51	2.87	.56	3.53	.58
45-59 years	3.50	.66	3.00	.53	3.64	.49	2.84	.57	3.55	.59
≥ 60 years	3.38	.60	3.05	.50	3.57	.50	2.90	.57	3.54	.53
$F_{(3, 5139)}$	11.14*		20.91*		.43		36.92*		1.73	
η^2	.01		.01		<.001		.02		.001	
<i>Level of education¹</i>										
Low	3.44	.69	2.86	.55	3.61	.60	2.96	.59	3.47	.61
Medium	3.48	.67	2.92	.56	2.96	.60	2.96	.60	3.47	.61
High	3.47	.68	2.90	.54	2.70	.56	2.97	.56	3.62	.60
$F_{(2, 5139)}$.45		2.35		1.79		.14		10.98*	
η^2	<.001		.001		.001		<.001		.004	
<i>Age & Gender^F</i>										
$F_{(3, 5139)}$	1.65		.32		.88		2.02		1.13	
η^2	.001		<.001		.001		.001		.001	
<i>Age & Education</i>										
$F_{(6, 5139)}$	2.74		.79		.50		.60		2.70	
η^2	.003		.001		.001		.001		.003	
<i>Education & Gender</i>										
$F_{(2, 5139)}$.32		.83		1.11		.88		.10	
η^2	<.001		<.001		<.001		<.001		<.001	
<i>Age, Gender & Education</i>										
$F_{(6, 5139)}$.11		2.06		1.38		1.41		1.20	
η^2	<.001		.002		.002		.002		.001	

Note. *M* = means, *SD* = standard deviation.¹Low level of education - lower or no secondary education; medium level of education - intermediate secondary education; high level of education - entrance qualification for universities of applied sciences and general universities.

* $p < .00145$

The outcomes of the ESEM analysis indicate an unambiguous five-factor structure for the BFI-K items, which were clearly interpretable in terms of the Big Five. All except of two items loaded highest on their corresponding factors. The cross-loadings detected in the present study were all already found in the initial validation study indicating that they appear to be specific for the BFI instrument (cf. Rammstedt & John, 2005). Although we could replicate the work of Rammstedt and John (2005) to a large extent, the here-found factorial structure is less explicit than the one reported by them. This difference is most likely accounted for by differences in the sample compositions with the here investigated sample being much more heterogeneous than the former ones. Rammstedt und colleagues (2010; see also Rammstedt & Kemper, 2011) explain the general difficulties in replicating the Big Five in non-student samples. They demonstrated that the factorial structure is particularly sensitive to educational effects. While the factorial structure of the Big Five emerged “with textbook-like clarity” (Rammstedt et al., 2010, p. 53) in student samples, it was found to be strongly blurred in samples heterogeneous with regard to education. As the present results are based on a non-student sample, we assume that the secondary loadings and low loadings may be due to the detrimental effect described by Rammstedt et al. (2010). In summary, however, the results of the present study using a heterogeneous sample largely replicate previous findings in more homogeneous samples (cf. Rammstedt & John, 2005) and support the factorial validity of the BFI-K.

Results of the analyses of sociodemographic effects clearly replicated the most frequently reported effects: (1) in accordance to the recent findings on gender effect on the Big Five (Marsh et al., 2010) we found that women score higher on all Big Five factors. The most explicit differences were found in Neuroticism and Conscientiousness, these differences had been also already reported in literature (e.g., Costa et al., 2001; Rammstedt, 2007; Srivastava et al., 2003). (2) In correspondence with earlier findings (e.g., Caspi et al., 2005; McCrae et al., 1999; Rammstedt, 2007) the elder respondents in our study scored higher on Agreeableness and lower on Neuroticism than younger ones. (3) Also in our study higher educated respondents scored higher on Openness than participants with a low level of education (e.g., Goldberg et al., 1998; Rammstedt 2007; Vassend & Skrandal, 1995). Finally, also in line with previous findings in this field is the fact that no interaction effects of age, gender or education on the Big Five were detected (cf. Rammstedt, 2007).

Although a substantial number of very short personality measures have been developed to allow the assessment of individuals' standing on personality traits in such research settings (e.g., Aronson et al., 2006; Bernard et al., 2005; Herzberg & Brähler, 2006; Gosling et al., 2003; Rammstedt 2007; Rammstedt & John, 2005, 2007; Rammstedt et al., 2004; Woods & Hampson, 2005) the construct validity of such very brief measures have become increasingly criticized (Credé et al., 2012). Nevertheless, the Big Five are still often used to explain and predict content variables of various research issues, which are investigated in samples different from the student samples. Aronson and colleagues (2006) explored the efficiency of team work and confirmed the influence of leader personality profile on new product development. Sanz, García-Vera, and Magán (2010) investigated the individual differences in trait anger and hostility in an extended heterogeneous Spanish sample. They report

that trait anger to be primarily associated with Neuroticism, whereas mistrust and confrontational attitude are principally related to low Agreeableness. McCann (2010) analyzed the data of the large-scale state-based survey taken from the National Center for Chronic Disease Prevention of the USA and found Neuroticism to be the prime predictor of well-being. For such studies the BFI-K with its four items per dimension promises to be more reliable and still very efficient measurement tool.

In sum, this study provides evidence that the BFI-K is a sufficiently reliable and valid short scale for the measurement of the Big Five for research purposes. Results of the initial validation studies were replicated and extended in a large and heterogeneous sample. Furthermore, the standardized normal values provided as supplementary online material will enable researchers to compare their findings with values for population as a whole and for subpopulations differentiated according to the core sociodemographic variables age, gender, and education. However, compared to full-length Big Five instruments there are substantial losses with regard to the psychometric quality of the scales. Thus, if testing time is not extremely limited, full-length Big Five measures possess clear psychometric advantages. But, of the several short-scale instruments assessing the Big Five the BFI-K proved to provide the highest psychometric quality. It thus allows an acceptable reliable and valid assessment of the Big Five in studies which do not allow for a comprehensive assessment of the Big Five due to monetary or time constraints.

References

- Aiken, L. R. & Groth-Marnat, G. (2006). *Psychological testing and assessment*. Boston: Pearson.
- Allemand, M., Job, V., Christen, S., & Keller, M. (2008). Forgiveness and action orientation. *Personality and Individual Differences* 45, 762-766. doi: 10.1016/j.paid.2008.08.002
- Aronson, Z. H., Reilly, R. R., & Lynn, G. S. (2006). The impact of leaderpersonality on new product development teamwork and performance: The moderating role of uncertainty. *Journal of Engineering and Technology Management*, 23, 221-247. doi: 10.1016/j.jengtecman.2006.06.003
- Asparouhov, T. & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397-438.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9, 9-20. doi: 10.1111/1468-2389.00160
- Benet-Martínez, V. & John, O. P. (1998). Los cincograndes across cultures and ethnic groups: Multitraitmultimethod analysis of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75, 729-750. doi: 10.1037//0022-3514.75.3.729.
- Bernard, L. C., Walsh, R. P., & Mills, M. (2005). Ask once, may tell: Comparative validity of single and multiple item measurement of the Big-Five personality factors. *Counseling and Clinical Psychology Journal*, 2, 40-57.

- Blickle, G., Meurs, J. A., Zettler, I., Solga, J., Noethen D., Kramer, J., & Ferris, G. R. (2008). Personality, political skill, and job performance. *Journal of Vocational Behavior, 72*, 377-387. doi: 10.1016/j.jvb.2007.11.008
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford.
- Borkenau, P. & Ostendorf, F. (2008). *Neo-FFI: Neo-Fünf-Faktoren-Inventar nach Costa und McCrae (Manual)*(2. neu normierte und vollständig überarbeitete Auflage). Göttingen: Hogrefe.
- Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality development: Stability and change. *Annual Review of Psychology, 56*, 453-484. doi: 10.1146/annurev.psych.55.090902.141913
- Costa, P. T., Jr. & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., Jr. & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322-331. doi:10.1037/0022-3514.81.2.322
- Credé, M., Harms, P., Niehorster, S., & Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and Social Psychology, 102*, 874-888. doi:10.1037/a0027403.
- 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. doi: 10.1037/0033-2909.124.2.197
- De Raad, B. (2000). *The Big Five personality factors*. Seattle, WA: Hogrefe.
- 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. doi: 10.1037/1040-3590.18.2.192.
- Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116*, 429-456.
- Goldberg, L. R., Sweeney, D., Merenda, P. F., & Hughes, J. E. (1998). Demographic variables and personality: The effects of gender, age, education, and ethnic/racial status on self-descriptions of personality attributes. *Personality and Individual Differences, 24*, 393-403. doi: 10.1016/S0191-8869(97)00110-4
- 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. doi: 10.1016/S0092-6566(03)00046-1
- Haslam, N., Whelan, J., & Bastian, B. (2009). Big Five traits mediate associations between values and subjective well-being. *Personality and Individual Differences, 46*, 40-42. doi: 10.1016/j.paid.2008.09.001
- 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* (3), 139-148. doi: 10.1027/1015-5759.22.3.139.

- Hopwood, C. J. & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review, 14*, 332-346. doi: 10.1177/1088868310361240
- Hurtz, G. M., & Donovan, J. J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology, 85*, 869-879. doi: 10.1037/0021-9010.85.6.869
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The "Big Five" Inventory - Versions 4a and 54*. Berkeley University of California, Berkeley, Institute of Personality and Social Research.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm Shift to the Big Five Trait Taxonomy: History, Measurement, and Conceptual Issues. In O. P. John, R. W. Robins, & L. A. Pervin (Hrsg.), *Handbook of personality: theory and research* (S. 114-158). New York: Guilford Press.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford.
- Kunz, V. (2005). Persönlichkeitseigenschaften und Wahlverhalten in den alten und neuen Bundesländern nach der Theorie des überlegten Handelns. In S. Schumann & H. Schoen (Eds.), *Persönlichkeit. Eine vergessene Größe der empirischen Sozialforschung* (pp. 117-136). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.
- Lang, F., John, D., Lüdtke, O., Schupp, J., & Wagner, G. (2011). Short assessment of the Big Five: robust across survey methods except telephone interviewing. *Behavior Research Methods, 43*, 548-567. Verfügbar unter: 10.3758/s13428-011-0066-z
- Lang, F. R., & Lüdtke, O. (2005). Der Big Five-Ansatz der Persönlichkeitsforschung: Instrumente und Vorgehen. In S. Schumann & H. Schoen (Eds.), *Persönlichkeit. Eine vergessene Größe der empirischen Sozialforschung* (pp. 29-39). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Lang, F. R., Lüdtke, O., & Asendorpf, J. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica, 47*, 111-112. doi: 10.1026//0012-1924.47.3.111
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the Big Five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*, 471-491. doi: 10.1037/a0019227
- Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16*, 439-476. doi: 10.1080/10705510903008220
- McCann, S. J. H. (2010). Subjective well-being, personality, demographic variables, and American state differences in smoking prevalence. *Nicotine & Tobacco Research, 12*, 895-904. doi: 10.1093/ntr/ntq113.
- McCrae, R. R., Zonderman, A. B., Costa, P. T., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology, 3*, 552-566.

- McCrae, R. R., Costa P. T., de Lima, M. P., Simoes, A., Ostendorf, F., Angleitner, A., & Piedmont, R. L. (1999). Age differences in personality across the adult life span: Parallels in five cultures. *Developmental Psychology, 35*, 466-477. doi: 10.1037/0012-1649.35.2.466
- McCrae, R. R., Kurtz, J. E., Yamagata, S., & Terracciano, A. (2011). Internal Consistency, Retest Reliability, and Their Implications for Personality Scale Validity. *Personality and Social Psychology Review, 15*, 28-50. doi: 10.1177/1088868310366253
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171-189.
- Muthén, L. K. & Muthén, B. (2008). *Mplus user's guide*. Los Angeles CA: Muthén & Muthén.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Ostendorf, F., & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar nach Costa und McCrae, Revidierte Fassung (NEO-PI-R)*. Göttingen: Hogrefe.
- Rammstedt, B. (1997). *Die deutsche Version des Big Five Inventory (BFI): Übersetzung und Validierung eines Fragebogens zur Erfassung des Fünf-Faktoren-Modells der Persönlichkeit*. (Unpublished thesis). University of Bielefeld, Germany.
- Rammstedt, B. (2007). The 10-Item Big Five Inventory norm values and investigation of sociodemographic effects based on a German population representative sample. *European Journal of Psychological Assessment, 23*, 193-201. doi: 10.1027/1015-5759.23.3.193
- Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markets for persons with different levels of education. *Journal of Research in Personality, 44*, 53-61. doi: 10.1016/j.jrp.2009.10.005
- 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. *Diagnostica, 51*, 195-206. 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. doi: 10.1016/j.jrp.2006.02.001
- Rammstedt, B., & Kemper, C. J. (2011). Measurement equivalence of the Big Five: Shedding light on potential causes of the educational bias. *Journal of Research in Personality, 45*, 121-125. doi: 10.1016/j.jrp.2010.11.006
- Rammstedt, B., Koch, K., Borg, I., & Reitz, T. (2004). Entwicklung und Validierung einer Kurzsкала für die Messung der Big Five Persönlichkeitsdimensionen in Umfragen. *ZUMA Nachrichten, 55*, 5 - 28.
- Rammstedt, B., & Schupp, J. (2008). Only the congruent survives - Personality similarities in couples. *Personality and Individual Differences, 45*, 533-535. doi: 10.1016/j.paid.2008.06.007
- Robinson, M. D., Solberg, E. C., Vargas, P. T., & Tamir, M. (2003). Trait as default: Extraversion, subjective well-being, and the distinction between

- neutral and positive events. *Journal of Personality and Social Psychology*, *85*, 517-527. doi: 10.1037/0022-3514.85.3.517
- Rogelberg, S. G., Conway, J. M., Sederburg, M. E., Spitzmüller, C., Aziz, S., & Knight, W. E. (2003). Profiling active and passive nonrespondents to an organizational survey. *Journal of Applied Psychology*, *88*, 1104-1114. doi: 10.1037/0021-9010.88.6.1104
- Salgado, J. F. (1997). The Five Factor Model of Personality and job performance in the European community. *Journal of Applied Psychology*, *82*, 30-43. doi: 10.1037/0021-9010.82.1.30
- Sanz, J., García-Vera, M. P., & Magán, I. (2010). Anger and hostility from the perspective of the Big Five personality model. *Scandinavian Journal of Psychology*, *51*, 262-270. doi: 10.1111/j.1467-9450.2009.00771.x.
- Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Personality processes and individual differences - development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology*, *84*, 1041-1053. doi: 10.1037/0022-3514.84.5.1041
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, *44*, 703-742.
- Van Hiel, A., Kossowska, M., & Mervielde, I. (2000). The relationship between Openness to Experience and Political Ideology. *Personality and Individual Differences*, *28*, 741-751. doi: 10.1016/S0191-8869(99)00135-X
- Vassend, O. & Skrandal, A. (1995). Factor analytic studies of the NEO-Personality-Inventory and the Five-Factor model: The problem of high structural complexity and conceptual indeterminacy. *Personality and Individual Differences*, *19*, 135-147. doi: 10.1016/0191-8869(95)00041-4
- Viken, R. J., Rose, R. J., Kapiro, J., & Koskenvuo, M. (1994). A developmental genetic analysis of adult personality: Extraversion and neuroticism from 18 to 59 years of age. *Journal of Personality and Social Psychology*, *4*, 722-730. doi: 10.1037/0022-3514.66.4.722
- Von Collani, G., & Grumm, M. (2009). On the dimensional structure of personality, ideological beliefs, social attitudes, and personal values. *Journal of Individual Differences*, *30*, 107-119. doi: 10.1027/1614-0001.30.2.107
- Woods, S. A., & Hampson, S. E. (2005). Measuring the Big Five with single items using a bipolar response scale. *European Journal of Personality*, *19*, 373-390. doi: 10.1002/per.542

Author's Note

Correspondence concerning this manuscript should be addressed to Beatrice Rammstedt, GESIS - Leibniz Institute for the Social Sciences, PO-Box 12 21 55, D-68072 Mannheim, Germany.
 Fon.: +49-621-1246-155
 E-mail: beatrice.rammstedt@gesis.org.

About the Authors

Dr. Anastasiya Kovaleva studied Psychology at the University of Bielefeld. She worked at the GESIS, Leibniz Institute for the Social Sciences, in the Department Survey Design and Methodology and taught at the University of Mannheim. In her doctoral thesis she constructed and validated a short scale for the assessment of locus of control. Currently she is a researcher at the University of Bielefeld. Her research interests concern research methodology and development of efficient psychological measurement tools.

Contact: anstasiya.kovaleva@uni-bielefeld.de

Website: <http://didaktik.biologie.uni-bielefeld.de/index.php/de/mitarbeiter/dr-anastasiya-kovaleva>

Dr. Constanze Beierlein studied Psychology and Political Science at the Universities of Koblenz-Landau, Darmstadt, and Frankfurt. In her doctoral thesis she investigated gender differences in motivation and strategy use when learning a new computer software. She is senior researcher at the GESIS, Leibniz Institute for the Social Sciences. In her research she is concerned with the construction and validation of short scales for large-scale social surveys. Her further research interests include methodological as well as substantive research questions with regard to basic human values, political participation, and justice psychology. She is co-editor of the "Journal of Political Psychology" in Germany.

Contact: constanze.beierlein@gesis.org

Website: <http://www.gesis.org/das-institut/mitarbeiterverzeichnis>

Dr. Christoph J. Kemper is a psychologist and senior researcher who is concerned with research, lectures, consulting, and providing services in the field of psychological assessment. He worked at different universities and research institutes in Germany, e.g. GESIS. In his research he focuses on the assessment of normal and abnormal personality, attitudes, cognitive assessment as well as different aspects of data quality and related interindividual differences, e.g. response styles.

Contact: contact@christoph-kemper.net

Website: <http://www.christoph-kemper.net>

Prof. Dr. Beatrice Rammstedt studied Psychology at the University of Bielefeld, University of California at Berkeley and Université Paris X. She completed her dissertation on gender differences in self-estimated intelligence at the University of Göttingen and her habilitation on the assessment of the Big Five personality factors in situations with extreme time limitations at the University of Mannheim. Currently she holds the position of a scientific director of the Department Survey Design and Methodology at the GESIS - Leibniz Institute for the Social Sciences and is full professor at the University of Mannheim.

Contact: beatrice.rammstedt@gesis.org

Website: <http://www.gesis.org/das-institut/mitarbeiterverzeichnis>