Is the Short Version of the Big Five Inventory (BFI-S) Applicable for Use in Telephone Surveys?

Oliver A. Brust[1], Sabine Häder[2], and Michael Häder[1]

The inclusion of psychological indicators in survey research has become more common because they offer the possibility of explaining much of the variance in sociological variables. The Big Five personality dimensions in particular are often used to explain opinions, attitudes, and behavior. However, the short versions of the Big Five Inventory (BFI-S) were developed for face-to-face surveys. Studies have shown distortions in the identification of the Big Five factor structure in subsamples of older respondents in landline telephone surveys. We applied the same BFI-S but with a shorter rating scale in a telephone survey with two subsamples (landline and mobile phone). Using exploratory structural equation modeling (ESEM), we identified the Big Five structure in the subsamples and the age groups. This finding leads us to conclude that the BFI-S is a powerful means of including personality characteristics in telephone surveys.

Key words: Exploratory structural equation modeling; telephone surveys; Big Five personality dimensions.

1. Introduction

Psychological indicators are being used increasingly in survey research as determining factors for the explanation of behavior and attitudes. Rammstedt (2007a), for example, showed that a significant proportion of the variance of sociological variables of interest could be explained using personality dimensions. This holds true for very different social phenomena, such as political attitudes (Heaven and Bucci 2001; Saucier 2000; Van Hiel et al. 2004), educational careers, career choices, the interaction of personality development and social relations, health history, and life course trajectories (Caspi et al. 2005; Goldberg et al. 1998). One empirical question that needs to be addressed in this context is: with which instruments and modes or devices of data collection can personality structure be assessed efficiently – that is, as briefly, reliably, and validly as possible?

In situations where personality is the primary topic of interest, the measurement of these psychological variables requires long inventories. The original German-language version of the Big Five Inventory (BFI), for example, comprises 42 items (Lang et al. 2001; Rammstedt and John 2007). However, in survey situations, where brevity is a high

[1] Technische Universität Dresden, Faculty of Philosophy, Institute of Sociology, D-01062 Dresden, Germany. Emails: oliver.brust@tu-dresden.de and michael.haeder@tu-dresden.de
[2] Leibniz Institute for the Social Sciences, PBox 122155, D-68072 Mannheim, Germany. Email: sabine.haeder@gesis.org

priority, an inventory of this length is too time consuming. Efforts have therefore been made to develop short inventories that can be easily applied in surveys (Gosling et al. 2003; Gerlitz and Schupp 2005, 204).

Short versions of the BFI (with 10 or 15 items) have been used in large studies such as the International Social Survey Programme (ISSP), the German Socio-Economic Panel (SOEP), the British Household Panel Survey, the Household, Income and Labour Dynamics in Australia Survey (HILDA), and the German National Cohort. The Big Five personality dimensions represent a powerful means of analyzing interindividual differences in personality dimensions (Lang et al. 2011). These five dimensions can be described as follows:

 – Neuroticism refers to individual differences in the susceptibility to distress and the experience of negative emotions such as anxiety, anger, and depression.
 – Extraversion refers to individual differences in sociability, gregariousness, level of activity, and the experience of positive emotions.
 – Openness to experience refers to individual differences in the propensity for originality, creativity, and the acceptance of new ideas.
 – Agreeableness refers to individual differences in altruistic behavior, trust, warmth, and kindness.
 – Conscientiousness refers to individual differences in self-control, task orientation, and rule abiding (Taylor et al. 2010, a3-21–a3-22).

As the use of psychological variables in surveys becomes more common, mixed-device surveys are becoming more frequent as well. The utilization of several devices is aimed at taking advantage of the rapid technological progress for survey research (Toepoel and Lugtig 2015). However, existing short scales for the measurement of psychological constructs have been developed and tested only for particular modes and/or devices of data collection – mostly for face-to-face interviewing with CAPI or PAPI. Hence, the question that arises is whether these short scales are also suitable for use in other modes and with other devices – for example mobile phone and landline phone surveys – and are therefore applicable in mixed-device surveys.

Because surveys will be increasingly conducted using both landline and mobile phones, it is very important that the short version of the BFI applied is suitable for users of both devices. To determine whether this is the case, we incorporated a BFI-S into the CELLA2 study (acronym for CELl Phone and LAndline Phone Survey 2), in which telephone numbers for both samples were drawn from different frames, and interviews were conducted by mobile and landline phone. CELLA2 was conducted by GESIS–Leibniz Institute for the Social Sciences, Mannheim, Germany, and Dresden University of Technology, Germany, and was funded by a research grant from the German Research Foundation (DFG).

1.1. Hypotheses

The following hypotheses will be tested:

H1: The Big Five structure of the personality dimensions Neuroticism, Extraversion, Openness to experience, Agreeableness, and Conscientiousness is clearly represented in both the landline and the mobile phone subsamples.

In an evaluation of possible mode effects among landline and mobile phone respondents in the CELLA surveys, Häder and Kühne (2010) showed that differences between the two subsamples in terms of response quality were not significant. Because the device used by respondents (landline vs. mobile phone) does not appear to have an influence on their answers, we expect to find the Big Five personality dimensions in both subsamples.

> **H2:** The Big Five structure of personality dimensions is reproduced in all age groups (age groups: 16–39, $n = 1,244$; 40–59, $n = 1,133$; 60 and older, $n = 514$).

In 2005, Lang et al. (2011) used a short version of the BFI when conducting a landline-only computer-assisted telephone interview (CATI) study with 1,200 respondents in which they applied a 15-item BFI-S with a seven-point rating scale. However, they observed distortions in the CATI assessment of the Big Five personality dimensions among older adults. One possible explanation is the assumption "that the mental workload of the telephone interviewing context would preclude valid self-report responses, since it requires listening to interviewers while reflecting responses on a 7-point rating scale" (Lang et al. 2011, 559).

In CELLA2, we shortened the seven-point scale to a five-point scale in order to lighten the workload of answering the Big Five items. However, although the coarser scale reduces the cognitive effort required to find a satisfactory answer, it could lead to a loss of information.

2. Method

In CELLA2, 3,051 participants (aged 16–93 years, $M = 43.43$, $SD = 16.52$, 48% female) were interviewed about their telephone usage behavior. The questionnaire also included several items aimed at measuring data quality (e.g., question order effects, social desirability, and response stability; see Häder 2012; Kühne et al. 2009). In order to compare the two subsamples, 1,516 interviews were conducted via landline phone and 1,535 via mobile phone. Participants were randomly selected. The same instrument was used nationwide for both subsamples. The landline sample was drawn from the universe of possible landline numbers in Germany using simple random sampling (Gabler and Häder 2002). The sampling frame comprised 139,366,300 numbers, from which 31,358 numbers were selected. For the selection of participants for the mobile phone survey, a modified RDD method was used (Gabler et al. 2012, 147ff.). The sampling frame comprised 197,490,000 mobile phone numbers, from which 44,330 numbers were selected. Both samples were drawn by GESIS – Leibniz Institute for the Social Sciences in Mannheim, Germany. The fieldwork was carried out in the summer of 2010 by a commercial survey research firm and lasted six weeks. A maximum of 15 contact attempts was made for the gross sample ($M = 2.4$, $SD = 2.42$). Of the interviews, 42.3% were conducted on the first contact attempt and 23.3% on the second attempt. The mean duration of the interviews was 12.33 minutes. The following response rates according to AAPOR standards were realized: $RR3_l = 0.148$ for the landline phone sample and $RR3_m = 0.117$ for the mobile phone sample (Schneiderat and Schlinzig 2012, 124).

As is also the case with telephone surveys in other countries, the mobile phone subsample of CELLA2 had a higher percentage of men, was younger than the landline subsample, and more mobile phone respondents were single. Furthermore, Schneiderat and Schlinzig (2012, 129) observed an education bias in the complete sample compared to the 2009 German microcensus, the official reference statistic (complete sample/microcensus: lower secondary educational level: 20.5%/38.8%, intermediate secondary levels: 32.8%/28.5%, higher levels: 36.3%/25.7%, other: 10.4%/7.0%, for further details see Schneiderat and Schlinzig 2012, 129), in such a way that participants with a lower level of education were underrepresented. This is likely to be due only partly to the very low response rate in CELLA2 (approx. 10%), because in the 1990s, when response rates of up to 80% were reached in German social surveys, this bias was also observed (Koch 1998).

In sum, the CELLA2 sample performed well in representing the subgroups of the survey's target population, as a comparison to official reference statistics shows (Schneiderat and Schlinzig 2012, 131). Therefore, Schneiderat and Schlinzig (2012, 131) conclude: "The integration of a mobile sample by applying a dual-frame approach nearly always leads to better sample quality."

To apply the dual-frame approach in our study, the parameters described in Figure 1 are needed.

Following Gabler and Ayhan (2007), the inclusion probability of the target person i is

$$\pi_i \approx k_i^F \frac{m^F}{M^F} \cdot \frac{1}{z_i} + k_i^C \frac{m^C}{M^C}.$$

These inclusion probabilities are used to construct both the Horvitz-Thompson estimator (design weighting) and the GREG estimator (design and adjustment weighting, adjusted for sex, age, and education; see Gabler et al. 2012, 163).

In order to maximize statistical power within age group comparisons, three age groups were defined: young adulthood from 16 to 39 years (total: $n = 1{,}244$, $M = 27.80$, $SD = 6.57$; landline: $n = 516$, $M = 28.73$, $SD = 6.69$; mobile: $n = 728$, $M = 27.14$, $SD = 6.41$), middle adulthood from 40 to 59 years (total: $n = 1133$, $M = 48.80$, $SD = 5.61$; landline: $n = 583$, $M = 48.75$, $SD = 5.71$; mobile: $n = 550$, $M = 48.81$, $SD = 5.50$), and late adulthood comprising people aged 60 years and older (total: $n = 504$, $M = 69.03$, $SD = 6.66$; landline: $n = 343$, $M = 69.81$, $SD = 6.72$; mobile: $n = 171$, $M = 67.45$, $SD = 6.29$). The CELLA2 data can be found in the Data Archive for the Social Sciences at GESIS – Leibniz Institute for the Social Sciences, Germany.

Landline phone		Mobile phone	
M^F	Number of numbers in sampling frame	M^C	Number of numbers in sampling frame
m^F	Number of numbers in sample	m^C	Number of numbers in sample
k_i^F	Number of landline numbers at which target person i can be reached	k_i^C	Number of mobile phone numbers at which target person i can be reached
z_i	Size of household to which target person i belongs		

Fig. 1. Parameters for the dual-frame approach.

To capture the Big Five personality dimensions it was important to choose the instrument most suitable for use in a telephone survey. One possibility was the very short and efficient ten-item BFI-S developed by Rammstedt (2007b). Within the framework of the MOBILEPANEL project (see Häder et al. 2010), we conducted a pretest of this instrument with a panel of $n = 203$ persons who were interviewed via mobile phone. The results of the pretest showed that a more extensive instrument was needed for the telephone-based measurement of the Big Five because, even with ipsative data, exploratory factor analysis (EFA) failed to identify the five-factor structure. The personality dimensions were therefore measured using another short version of the BFI – namely, a 15-item instrument that was constructed for the SOEP and was used in this context for the first time in 2005 (Dehne and Schupp 2007). Whereas the SOEP presented the instrument visually (using a template) within the framework of a face-to-face interview, in our study CELLA2 it was administered by landline and mobile phone. To avoid overburdening the respondents, the response scale was reduced from seven to five scale points (1 – strongly disagree, 5 – strongly agree; Lang et al. 2011; Dehne and Schupp 2007, 8). Similar to Lang et al. (2011, 554), Cronbach's alpha values for the BFI-S scales were low, reflecting the brevity of the three-item scale and the width of these broad constructs (Neuroticism $\alpha = .52$, Extraversion $\alpha = .60$, Openness $\alpha = .55$, Agreeableness $\alpha = .45$, Conscientiousness $\alpha = .52$). As Gosling et al. (2003) demonstrate by comparing Cronbach's alpha values to test-retest reliability values in a ten-item Big Five measure, Cronbach's alpha might not be the right indicator to evaluate reliability for very brief scales due to an underestimation of the true reliability. However, an evaluation of the test-retest reliability was not possible in CELLA2 due to the design of the study. To avoid possible sequence effects, the items of the BFI-S were presented in random order. This randomization was implemented by the CATI software and changed with every participant. Within the questionnaire, the BFI-S was situated after a set of items concerning telephone usage behavior and was followed by other personality measures.

In the 2005 SOEP study, over 20,000 people were interviewed with this instrument. Therefore it was used as a reference for the CELLA2 results. The response rate of the SOEP is about 50% (see Goebel et al. 2008), which is significantly higher than that of CELLA2 (approx. 10%). Table 1 shows the results of the comparisons of the item means of the BFI-S between the two studies.

Overall, it can be seen that the differences between the SOEP and CELLA2 can be regarded as small. They do not exceed less than half a scale point on the seven-point scale. This indicates a satisfactory quality of the realized CELLA2 sample, despite the low response rate. A further comparison of the sample means for the BFI-S items in CELLA2 using the Horvitz-Thompson estimator and the GREG estimator did not reveal a general tendency that could be interpreted as an improvement or deterioration of the estimators (see Table 2). Therefore, in our next model-based analyses we forgo the use of weights.

2.1. Statistical Analysis for Testing Measurement Invariance

To compare the landline and mobile phone samples within the total sample and across the three different age groups (young adulthood, middle adulthood, late adulthood), we conducted exploratory structural equation modeling (ESEM) analyses using Mplus

Table 1.   Means of the BFI-S Items in the SOEP and CELLA2.

Item: I see myself as someone who. . .	CELLA2	SOEP	Difference: CELLA2−SOEP
N: Worries a lot	4.77	4.76	0.01
N: Gets nervous easily	3.58	3.77	− 0.19
N: Is relaxed, handles stress well	3.44	3.47	− 0.03
E: Is talkative	5.33	5.49	0.16
E: Is outgoing, sociable	5.29	5.07	0.22
E: Is reserved	4.21	3.86	0.35
O: Is original, comes up with new ideas	4.71	4.54	0.17
O: Values artistic, aesthetic experiences	4.22	4.09	0.13
O: Has an active imagination	4.78	4.83	− 0.05
A: Is sometimes rude to others	5.35	5.06	0.29
A: Has a forgiving nature	5.59	5.52	0.07
A: Is considerate and kind to almost everyone	5.93	5.78	0.15
C: Does a thorough job	6.08	6.15	− 0.07
C: Tends to be lazy	5.54	5.71	− 0.17
C: Does things efficiently	5.75	5.75	0.00

Note. N = Neuroticism, E = Extraversion, O = Openness to experience, A = Agreeableness, C = Conscientiousness, CELLA2 items adapted by multiplying by 7/5; Sources: data set GREG weighted.

(Version 7, Muthén and Muthén 2012) to test for measurement invariance of the short version of the Big Five Inventory (BFI-S). In the ESEM procedure, a model is estimated on the basis of an a priori postulated number of factors, thereby combining the advantages of EFA and confirmatory factor analysis (CFA). Within the procedure all factor loadings, item intercepts, and item uniquenesses are estimated. It is also possible to evaluate the fit

Table 2.   Comparison of the sample means, the Horvitz-Thompson estimator and the GREG estimator for the BFI-S items in CELLA2.

Item: I see myself as someone who. . .	Sample Mean	HT-Estimator	GREG Estimator
Worries a lot	3.27	3.28	3.41
Gets nervous easily	2.49	2.51	2.55
Is relaxed, handles stress well	2.43	2.45	2.46
Is talkative	3.84	3.80	3.81
Is outgoing, sociable	3.78	3.80	3.78
Is reserved	3.15	3.12	3.01
Is original, comes up with new ideas	3.44	3.42	3.37
Values artistic, aesthetic experiences	3.09	3.10	3.02
Has an active imagination	3.52	3.49	3.42
Is sometimes rude to others	3.83	3.83	3.82
Has a forgiving nature	3.94	3.96	3.99
Is considerate and kind to almost everyone	4.21	4.22	4.23
Does a thorough job	4.33	4.32	4.34
Tends to be lazy	3.86	3.87	3.96
Does things efficiently	4.15	4.13	4.11

of the model to the data and to test for measurement invariance of the estimated parameters across multiple groups (Asparouhov and Muthén 2008). Simultaneously, disadvantages of the individual methods are reduced. Traditional EFA does not offer a method of comparing different factor structures in regard to their equivalence, whereas CFA typically requires indicators to be assigned to single factors. This rules out the possibility of indicators loading on another factor at the same time. Therefore the CFA procedure alone may not be adequate for evaluating the model fit of the Big Five model, because fit indices do not show adequate fit while correlations between the five factors are artificially inflated at the same time (Hopwood and Donnellan 2010).

We used oblique Geomin rotation, following Marsh et al. (2010). Geomin rotation is recommended when indicators have substantial loadings on more than one factor (Browne 2001; Muthén and Muthén 2012), which is often the case with the Big Five model (Hopwood and Donnellan 2010). To evaluate model fit, the maximum-likelihood estimator (ML) with conventional standard errors and chi-square test statistic was used. Compared to other estimators (e.g., maximum-likelihood estimator with robust standard errors, MLR), the ML chi-square test statistic can be used easily for chi-square difference testing and therefore for multiple-group comparisons. However, ML requires a large sample size and multivariate normal distribution. Considering that the response format was ordered categorical, normal distribution could not be ensured for all variables. However, in large samples ML has proved to be relatively robust even when slight deviations from the normal distribution occur (West et al. 1995; Ximénez 2006). Nevertheless, all models were calculated with the maximum-likelihood estimator with robust standard errors (MLR), as well. The results did not differ substantially between ML and MLR. In the following section, we report the results of the calculations using Geomin rotation and the maximum-likelihood estimator. Listwise deletion was used to handle missing data because the percentage of missing data was very small and missing data were missing completely at random (Little's MCAR test: $p = .992$; chi-square $= 13,153$, $df = .28$; Little 1988). Comparably to SEM, within the ESEM procedure we can test whether an a priori postulated model fits the data.

Chi-square tests were used to test for model fit and a nonsignificant result was regarded as the indicator of a fitting model. However, chi-square test results are influenced by sample size (Tucker and Lewis 1973). Therefore, we used goodness-of-fit indices, which are considered to be relatively robust even in the case of sample size differences. To evaluate model fit, the Comparative Fit Index (CFI), the Tucker Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA), and the Standard Root Mean square Residual (SRMR) were used. CFI and TLI values greater than .90 and .95 indicate acceptable and excellent data fits, RMSEA values less than .05 and .08 indicate close and reasonable fits, and SRMR values less than .06 and .10 indicate close and reasonable fits, following common guidelines for the evaluation of model fits to the data (see Marsh et al. 1988; Hu and Bentler 1999; Marsh et al. 2004).

To test for differences between the subsamples (landline vs. mobile phone), ESEM multiple-group analyses were conducted. Five different models of measurement invariance were compared within the total sample and within the three different age groups following Lang et al. (2011; see also Marsh et al. 2013):

(1) Configural invariance is the least demanding model; it imposes no invariance constraints at all. It is used to establish a baseline condition according to which the five-factor structure exists in the two different model groups (landline vs. mobile).

(2) Weak invariance constrains the factor loadings to be invariant across the two different model groups.

(3) Strong invariance constrains the factor loadings and the item intercepts to be invariant across the two groups. A rejection of this model implies different item functioning (i.e., different item means between the two groups cannot be explained merely by differences at the factorial mean levels).

(4) Strict invariance constrains the factor loadings, the item intercepts, and item uniquenesses to be invariant across both groups. A rejection of this model indicates that differences exist in the measurement errors across both models.

(5) In the last model, factor loadings, item intercepts, item uniquenesses and factor means are constrained to be equal.

In order to determine the more parsimonious model, Bentler (1990) suggested the testing of nested models using chi-square difference testing. However, this procedure is also dependent on sample size (Brannick 1995). Therefore, the examination of changes in fit indices is also used as an alternative to this procedure (Cheung and Rensvold 1999; Chen 2007). According to Chen (2007), a more parsimonious model is supported if the CFI change is less than .01 or the RMSEA change is less than .015. According to Marsh et al. (2009), equally good or better TLI and RMSEA values compared to the less restrictive model are a more conservative criterion for the more parsimonious model.

3. Results

Table 3 shows the Geomin rotated loadings for the mobile and the landline phone samples. The Big Five factor structure is clearly identified by the 15 items of the BFI-S in both samples. The solutions show close fit (landline: chi-square/df = 80,884/40, p = .001, CFI/TLI = .987/.967, RMSEA/SRMR = .026/.014; mobile: chi-square/df = 139,827/40, p = .001, CFI/TLI = .969/.918, RMSEA/SRMR = .040/.019) and are almost textbook-like.

The fit indices of the ESEM multiple-group analysis (landline vs. mobile) for the whole sample are reported in Table 4.

In the total sample, fit indices of the ESEM multiple-group analysis (landline vs. mobile) showed close fit for the configural-invariance model and for the weak measurement-invariance model. A comparison of the fit indices of the two models indicated model improvement favoring the weak invariance model. Chi-square difference testing revealed no differences between the two models (chi-square/df = 41,863/50, p = .787, n.s.). The fit indices of the strong-invariance model also showed a close model fit. Compared to the weak invariance model, fit indices remained essentially stable. Chi-square difference testing revealed a significant disparity between the two models (chi-square/df = 25,830/10, p = .004). However, the differences between the CFI values of the strong measurement and the weak measurement-invariance model were less than .01 and a chi-square difference test between the strong- and the configural-invariance model revealed no differences (chi-square/df = 67,693/60, p = .231, n.s.). The strict measurement-invariance model also proved satisfactory. Compared to the

Table 3. *Geomin rotated loadings of the BFI-S in the landline and mobile phone samples.*

Item: I see myself as someone who…	Landline Sample					Mobile Phone Sample				
	N	E	O	A	C	N	E	O	A	C
Worries a lot	**.578**	.052	−.005	.124	.122	**.465**	.016	.009	.118	.168
Gets nervous easily	**.628**	−.023	.031	−.037	−.039	**.627**	−.010	.010	−.023	−.019
Is relaxed, handles stress well[a]	**.440**	−.028	−.048	−.148	−.129	**.424**	−.013	−.151	−.115	.134
Is talkative	.061	**.655**	.062	.069	.036	.058	**.643**	.071	.084	.027
Is outgoing, sociable	.010	**.733**	.008	.008	.036	.023	**.568**	.148	.044	.038
Is reserved[a]	−.111	**.475**	−.016	−.265	−.131	−.073	**.533**	−.069	−.191	−.130
Is original, comes up with new ideas	−.178	.070	**.520**	−.026	.108	−.106	.178	**.502**	−.072	.093
Values artistic, aesthetic experiences	.036	−.107	**.529**	.042	−.019	.022	−.066	**.507**	.076	−.063
Has an active imagination	.004	.049	**.557**	.020	−.057	.030	.047	**.580**	.008	−.054
Is sometimes rude to others[a]	−.141	−.026	−.137	**.485**	−.077	−.159	−.027	−.110	**.474**	−.075
Has a forgiving nature	−.015	.031	.141	**.273**	.038	−.015	.028	.169	**.337**	.037
Is considerate and kind to almost everyone	.018	.046	.032	**.707**	.022	.009	.094	.008	**.576**	.091
Does a thorough job	.038	−.005	−.022	−.031	**.706**	.049	−.023	−.022	.031	**.705**
Tends to be lazy[a]	−.036	.043	−.160	.077	**.286**	−.095	.088	−.116	.044	**.295**
Does things efficiently	−.046	.005	.054	.028	**.707**	−.057	.020	.037	−.021	**.693**

Note. Landline sample: $n = 1,516$, chi-square/$df = 80,884/40$, $p = .001$, CFI/TLI = .987/.967, RMSEA/SRMR = .026/.014; Mobile phone sample: $n = 1,535$, chi-square/$df = 139,827/40$, $p = .001$, CFI/TLI = .969/.918, RMSEA/SRMR = .040/.019; N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness.
[a]Item was recoded (inversed).

Table 4.   Model fit of the ESEM multiple-group analysis (landline vs. mobile) for the whole sample.

Model	ML/*df*	p_{fit} p_{diff}	CFI	TLI	RMSEA	SRMR
(1) Configural invariance	221/80	.001	.978	.942	.034	.017
Differences: (2) vs. (1)	42/50	.787	.001	.025	.008	.004
(2) Weak measurement invariance	263/130	.001	.979	.967	.026	.021
(3) vs. (2)	26/10	.004	.002	.002	.000	.000
(3) vs. (1)	68/60	.231				
(3) Strong invariance	288/140	.001	.977	.965	.026	.021
(4) vs. (3)	30/14	.007	.003	.000	.000	.010
(4) vs. (1)	98/74	.033				
(4) Strict invariance	319/154	.001	.974	.965	.026	.031
(5) vs. (4)	11/5	.045	.001	.000	.001	.001
(5) vs. (1)	109/79	.014				
(5) Strict invariance and fixed factor means	330/159	.001	.973	.965	.027	.032

Note. Total sample: $n = 3{,}051$; ML/*df* = maximum-likelihood chi-square/degrees of freedom; p_{fit} = chi-square test to evaluate model fit; p_{diff} = chi-square difference test between two models; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standard Root Mean square Residual.

strong-invariance model and the configural-invariance model, chi-square difference testing revealed significant differences (chi-square/*df* = 30,204/14, $p = .007$; chi-square/*df* = 97,897/74, $p = .033$). However, compared to the strong-invariance model, there were no CFI changes greater than .01, and TLI and RMSEA values also remained stable. The test of the model for strict invariance and fixed factor means was also satisfactory. Compared to the strict-invariance model and the configural-invariance model, chi-square difference testing revealed significant differences (chi-square/*df* = 11,321/5, $p = .045$; chi-square/*df* = 109,218/79, $p = .014$). However, there were no CFI changes greater than .01 compared to the strict-invariance model, and TLI and RMSEA values also remained essentially stable.

Regarding the complete sample, the comparison of the five models of measurement invariance supports the assumption of equal factor loadings, equal item intercepts, equal item uniquenesses, and equal factor means for the landline and mobile phone samples.

In the next step we tested for measurement invariance for the landline and mobile phone samples for different age groups. The Geomin rotated loadings representing the Big Five factor structure for the three age groups and fit indices are reported in Table 5. TLI and CFI indices show at least acceptable fits; RMSEA and SRMR indicate close fits. Once again, the solutions are almost textbook-like.

3.1.   Measurement Invariance in Young and Middle Adulthood

The fit indices of the ESEM multiple-group analysis (landline vs. mobile) are reported in Table 6 for the young adulthood sample, and in Table 7 for the middle adulthood sample.

With regard to the young and middle adulthood groups, the comparison of the five measurement-invariance models supports the assumption of equal factor loadings, equal item intercepts, equal item uniquenesses, and equal factor means for the landline and mobile

Table 5. Geomin rotated loadings for young, middle-aged and older adults.

Item: I see myself as someone who. . .	Young adults					Middle-aged adults					Older adults				
	N	E	O	A	C	N	E	O	A	C	N	E	O	A	C
Worries a lot	**.558**	.084	.032	.064	.168	**.483**	−.050	.037	.131	.206	**.379**	.030	−.059	.217	.064
Gets nervous easily	**.591**	−.061	.033	−.058	−.016	**.653**	.011	−.014	.003	−.003	**.759**	−.001	.014	−.047	−.028
Is relaxed, handles stress well[a]	**.497**	−.024	−.116	−.151	−.048	**.522**	.020	−.127	−.140	−.109	**.217**	−.100	−.005	.003	−.317
Is talkative	.099	**.656**	.018	.129	.028	.022	**.681**	.087	.037	.058	.031	**.593**	.094	.135	.031
Is outgoing, sociable	−.015	**.660**	.114	.009	.023	.014	**.541**	.189	.040	.073	.021	**.845**	−.078	.011	−.020
Is reserved[a]	−.110	**.535**	−.034	−.098	−.143	−.055	**.545**	−.057	−.261	−.096	−.024	**.276**	.023	−.207	−.127
Is original, comes up with new ideas	−.149	.131	**.540**	−.036	.085	−.155	.020	**.653**	−.007	.063	−.133	.283	**.356**	−.203	.137
Values artistic, aesthetic experiences	.062	−.059	**.480**	.068	−.074	.005	−.013	**.507**	.010	−.135	−.072	−.038	**.715**	.170	−.001
Has an active imagination	.039	.070	**.505**	.021	−.067	.044	.104	**.486**	−.015	−.059	−.024	.162	**.591**	−.008	−.040
Is sometimes rude to others[a]	−.052	−.006	−.242	**.584**	−.035	−.153	−.014	−.095	**.547**	−.159	−.223	−.040	−.037	**.388**	−.10
Has a forgiving nature	−.030	−.084	.189	**.280**	.039	−.010	.055	.186	**.339**	−.006	.022	.184	.060	**.238**	.106
Is considerate and kind to almost everyone	.057	.043	.026	**.573**	.053	.033	.085	.011	**.655**	.069	−.046	.159	.057	**.639**	.064
Does a thorough job	.071	−.010	−.029	−.010	**.689**	.029	.021	−.003	.026	**.622**	.009	−.001	−.049	.031	**.702**
Tends to be lazy[a]	−.009	.065	−.125	.058	**.329**	−.145	.070	−.097	.032	**.272**	−.101	.049	−.071	.166	**.170**
Does things efficiently	−.099	.007	.048	.002	**.704**	−.055	.024	.002	−.006	**.758**	.011	−.007	.106	.008	**.725**

Note. Young adults: $n = 1,244$, chi-square/$df = 109,276/40$, CFI/TLI = .971/.924, $p = .001$, RMSEA/SRMR = .037/.018; Middle-aged adults: $n = 1,133$, chi-square/$df = 104,811/40$, $p = .001$, CFI/TLI = .975/.935, RMSEA/SRMR = .038/.018; Older adults: $n = 504$, chi-square = 44,734/40, $p = .2797$, n.s., CFI/TLI = .996/.989, RMSEA/SRMR = .015/.018, ten respondents with very low education were excluded from the mobile phone sample in this age group; N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness.

[a]Item was recoded (inversed).

Table 6.   Model fit of the ESEM multiple-group analysis (landline vs. mobile) for young adulthood.

Model	ML/df	p_{fit} p_{diff}	CFI	TLI	RMSEA	SRMR
(1) Configural invariance	146/80	.001	.973	.928	.037	.020
Differences: (2) vs. (1)	61/50	.140	.005	.020	.006	.011
(2) Weak measurement invariance	207/130	.001	.968	.948	.031	.031
(3) vs. (2)	8/10	.653	.001	.006	.002	.000
(3) vs. (1)	69/60	.209				
(3) Strong invariance	215/140	.001	.969	.954	.029	.031
(4) vs. (3)	22/14	.083	.003	.001	.000	.004
(4) vs. (1)	90/74	.095				
(4) Strict invariance	237/154	.001	.966	.953	.029	.035
(5) vs. (4)	5/5	.443	.000	.002	.000	.001
(5) vs. (1)	95/79	.104				
(5) Strict invariance and fixed factor means	241/159	.001	.966	.955	.029	.036

Note. Young adults: $n = 1,244$; ML/*df* = maximum-likelihood chi-square/degrees of freedom; p_{fit} = chi-square test to evaluate model fit; p_{diff} = chi-square difference test between two models; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standard Root Mean square Residual.

phone samples, with fit indices showing excellent fits and remaining essentially stable across the different models. Chi-square difference testing also revealed no differences.

3.2.   Measurement Invariance in Late Adulthood

Finally, we tested measurement invariance of the BFI-S across the two sampling modes in later adulthood (i.e., 60 years and older). Neither the configural-invariance model nor the

Table 7.   Model fit of the ESEM multiple-group analysis (landline vs. mobile) for middle adulthood.

Model	ML/df	p_{fit} p_{diff}	CFI	TLI	RMSEA	SRMR
(1) Configural invariance	153/80	.001	.973	.928	.040	.021
Differences: (2) vs. (1)	62/50	.120	.005	.020	.006	.008
(2) Weak measurement invariance	215/130	.001	.968	.948	.034	.029
(3) vs. (2)	10/10	.426	.000	.004	.001	.001
(3) vs. (1)	72/60	.137				
(3) Strong invariance	225/140	.001	.968	.952	.033	.030
(4) vs. (3)	17/14	.247	.001	.003	.001	.009
(4) vs. (1)	89/74	.109				
(4) Strict invariance	242/154	.001	.967	.955	.032	.039
(5) vs. (4)	4/5	.612	.000	.002	.001	.002
(5) vs. (1)	93/79	.137				
(5) Strict invariance and fixed factor means	246/159	.001	.967	.957	.031	.041

Note. Middle-aged adults: $n = 1,133$; ML/*df* = maximum-likelihood chi-square/degrees of freedom; p_{fit} = chi-square test to evaluate model fit; p_{diff} = chi-square difference test between two models; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standard Root Mean square Residual.

single-group ESEM model showed any convergence for the mobile phone respondents, due to one item that had a negative residual variance leading to a nonpositive definite covariance matrix. After checking item covariances and correlations, we were able to rule out multicollinearity and linear dependency as alternative explanations.

These results are consistent with those obtained by Lang et al. (2011), who compared measurement invariance of the 15-item BFI-S in early, middle, and late adulthood across different modes of data collection (CATI, face-to-face interviewing, self-administered questionnaires). They too observed distortions in the results of the CATI assessment of the Big Five dimensions in older adults. Lang et al. (2001) suggested two possible explanations for these distortions – namely:

a) that they might be due to the fact that the mental workload of the telephone interviewing context caused by the seven-point rating scale led to invalid self-reports, as it might be difficult for older adults to listen to the interviewer and reflect on possible responses on a seven-point rating scale at the same time, and

b) that the costs of the greater workload would manifest themselves in greater variability in item responses, which might result in a reduced likelihood of identifying the expected five-factor structure.

To simplify the assessment situation and to reduce the mental workload for older adults, we used a five-point rating scale. In contrast to the results of Lang et al. (2011, 558f.), which showed a nonacceptable model fit for the single-group ESEM model in the landline sample (chi-square/df = 83/40, p < .001, CFI/TLI = .909/.761, RMSEA = .069), the single-group ESEM model in our results showed excellent fit, as evidenced by a nonsignificant chi-square difference test (chi-square/df = 34,202/40, p = .728, n.s.).

Therefore, the simplification of the assessment situation by using a five-point rating scale may have helped older respondents to handle the telephone interviewing situation better – at least in the landline sample. By contrast, the mobile phone setting might be more difficult for older respondents. According to Lang et al. (2011), the distortion of self-report responses due to a higher mental workload might be a problem for less-educated older respondents in particular. They therefore conducted ESEM analyses excluding older adults with only eight or nine years of education from the CATI sample, which allowed them to successfully test for measurement invariance across the three different conditions. Following Lang et al. (2011), we also excluded ten respondents who did not have any school graduation qualifications from the mobile phone sample, and then tested for measurement invariance for the landline and mobile phone samples.

The fit indices of the ESEM multiple-group analysis for the late-adulthood sample are reported in Table 8.

The configural-invariance model, the weak measurement-invariance model, the strong-invariance model, and the strict-invariance model showed excellent fits as evidenced by nonsignificant chi-square difference tests. Multiple-group comparison also revealed no differences between the first three models, as evidenced by nonsignificant chi-square difference tests as well as fit indices which remained essentially stable. However, chi-square difference testing revealed significant differences in the strict-invariance model compared to the strong-invariance model (chi-square/df = 24,963/14, p = .035) and the configural model (chi-square/df = 98,757/74, p = .029). The CFI

Table 8. Model fit of the ESEM multiple-group analysis (landline vs. mobile) for late adulthood.

Model	ML/df	p_{fit} p_{diff}	CFI	TLI	RMSEA	SRMR
(1) Configural invariance	84/80	.370	.997	.992	.013	.023
Differences: (2) vs. (1)	59/50	.180	.008	.010	.007	.021
(2) Weak measurement invariance	143/130	.213	.989	.982	.020	.044
(3) vs. (2)	15/10	.139	.004	.005	.002	.004
(3) vs. (1)	74/60	.109				
(3) Strong invariance	157/140	.149	.985	.977	.022	.048
(4) vs. (3)	25/14	.035	.010	.011	.005	.016
(4) vs. (1)	99/74	.029				
(4) Strict invariance	182/154	.059	.975	.966	.027	.064
(5) vs. (4)	7/5	.207	.002	.001	.001	.001
(5) vs. (1)	106/79	.023				
(5) Strict invariance and fixed factor means	190/159	.049	.973	.965	.028	.065

Note. Older adults: $n = 504$, ten respondents with very low education were excluded from the mobile sample; ML/df = maximum-likelihood chi-square/degrees of freedom; p_{fit} = chi-square test to evaluate model fit; p_{diff} = chi-square difference test between two models; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standard Root Mean square Residual.

change was .01. Differences in fit indices between the strict-invariance model and the strong-invariance model also revealed no improved parsimony. The strict-invariance model (with fixed factor means) showed reasonable fit. Chi-square difference testing revealed no significant differences between the strict-invariance model with fixed factor means and the strict-invariance model (chi-square/df = 7,195/5, p = .207, n.s.). However, compared to the configural-invariance model, differences were significant (chi-square/df = 105,952/79, p = .023). Compared to the strict-invariance model fit indices remained essentially stable.

Comparing the five measurement-invariance models supports the assumption of equal factor loadings and equal item intercepts. However, the assumption of equal item uniquenesses and equal factor means is not supported. This indicates the existence of differences in measurement error in the landline sample compared to the mobile phone sample. Moreover, factor means may not be invariant across these two assessment conditions among older respondents.

4. Discussion

The results support our hypotheses H1 and H2. We were able to demonstrate that the Big Five personality dimensions were represented in both the landline and mobile phone samples. We were also able to show that these dimensions were reproduced in all three age groups of the respondents. Hence, our research results are not consistent with those of Lang et al. (2011), who questioned the suitability of the BFI-S for use in telephone surveys that include older adults. The fact that, in contrast to Lang et al.'s study, the BFI-S also yielded satisfactory results in the older adults group may be due to the fact that we used a five-point rather than a seven-point rating scale, which may have considerably reduced the mental workload of answering the questions.

However, we also found that some models for testing measurement invariance did not fit for the entire older adult population (mobile phone sample). This finding is consistent with the results of Rammstedt et al. (2010), who found that in subgroups with no, low, or intermediate secondary education, the Big Five structure could not be identified as expected. However, in samples with higher secondary education, the five-factor structure replicated clearly. According to the authors, these factor structures appear to be highly sensitive to a person's educational level.

A second methodological problem concerning the Big Five personality dimensions is acquiescence response bias (Rammstedt et al. 2013), which increases with age. Interestingly, the problems representing the Big Five factor structure for older and/or less-educated respondents occur not only in telephone samples, but also in the samples of the 2004 and 2006 ISSP who completed a self-administered questionnaire (Rammstedt et al. 2010). To reduce this effect, it might be helpful to use ipsative data.

Another limitation of our study arises from the possibility that older people's skills in participating in telephone surveys might have changed since 2005, when the data were collected on which Lang et al.'s (2011) study was based. Therefore, it is not possible to determine with certainty whether (a) the five-point rating scale actually works better, or (b) older adults have become more adept at taking phone surveys. The ideal way to test this would be to randomly assign people to a five- or seven-point scale.

In sum, we were not able to substantiate concerns expressed by Lang et al. (2011) regarding the use of the BFI-S in age-heterogeneous telephone samples in general. On the contrary, we would encourage survey researchers to make more use of such inventories because in many cases personality traits are important determinants of behavior and attitudes. Our findings lead us to conclude that in both landline as well as mobile phone surveys the application of the 15-item BFI-S works sufficiently.

5. References

Asparouhov, T. and B. Muthen. 2008. "Multilevel Mixture Models." In *Advances in Latent Variable Mixture Models*, edited by G.R. Hancock and K.M. Samuelsen, 27–51. Charlotte, NC: Information Age Publishing.

Bentler, P. 1990. "Comparative Fit Indexes in Structural Models." *Psychological Bulletin* 107: 238–246.

Brannick, M.T. 1995. "Critical Comments on Applying Covariance Structure Modeling." *Journal of Organizational Behavior* 16: 201–213. Doi: http://dx.doi.org/10.1002/job.4030160303.

Browne, M.W. 2001. "An Overview of Analytic Rotation in Exploratory Factor Analysis." *Multivariate Behavioral Research* 36: 111–150. Doi: http://dx.doi.org/10.1207/S15327906MBR3601_05.

Caspi, A., B.W. Roberts, and R.L. Shiner. 2005. "Personality Development: Stability and Change." *Annual Review of Psychology* 56: 453–484. Doi: http://dx.doi.org/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. Doi: http://dx.doi.org/10.1080/10705510701301834.

Cheung, G.W. and R.B. Rensvold. 1999. "Testing Factorial Invariance Across Groups: A Reconceptualization and Proposed New Method." *Journal of Management* 25: 1–27. Doi: http://dx.doi.org/10.1177/014920639902500101.

Dehne, M. and J. Schupp. 2007. "Persönlichkeitsmerkmale im Sozio-ökonomischen Panel (SOEP) – Konzept, Umsetzung und empirische Eigenschaften". *DIW Research Notes* 26. Berlin: DIW Berlin.

Gabler, S. and Ö. Ayhan. 2007. "Gewichtung bei Erhebungen im Festnetz und über Mobilfunk: Ein Dual-Frame Ansatz". In *Mobilfunktelefonie – Eine Herausforderung für die Umfrageforschung*, edited by S. Gabler and S. Häder, 39–45, ZUMA-Nachrichten Spezial, Vol. 13. Mannheim: GESIS.

Gabler, S., and S. Häder. 2002. "Idiosyncrasies in Telephone Sampling – The Case of Germany." *IJPOR* 14: 339–345. Doi: http://dx.doi.org/10.1093/ijpor/14.3.339.

Gabler, S., S. Haeder, I. Lehnhoff, and E. Mardian. 2012. "Weighting for Unequal Inclusion Probabilities and Nonresponse in Dual Frame Telephone Surveys." In *Telephone Surveys in Europe*, edited by S. Häder, M. Häder, and M. Kühne, 147–168. Heidelberg: Springer Verlag.

Gerlitz, J. and J. Schupp. 2005. "Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP". *DIW Research Notes* 4. Berlin: DIW Berlin.

Goebel, J., M. Grabka, P. Krause, M. Kroh, R. Pischner, I. Sieber, and M. Spiess. 2008. "Mikrodaten, Gewichtung und Datenstruktur der Längsschnittstudie Sozio-oekonomisches Panel (SOEP)." In *Vierteljahreshefte zur Wirtschaftsforschung*, Vol. 3, edited by J. Frick, O. Groh-Samberg, J. Schupp and K. Spiess, 77–109. Berlin: Duncker & Humblot. Doi: http://dx.doi.org/10.3790/vjh.77.3.77.

Goldberg, L.R., D. Sweeney, P.F. Merenda, and J.E. Hughes. 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: http://dx.doi.org/10.1016/S0191-8869(97)00110-4.

Gosling, S., P. Rentfrow, and W. Swann. 2003. "A Very Brief Measure of the Big Five Personality Domains." *Journal of Research in Personality* 37: 504–528. Doi: http://dx.doi.org/10.1016/S0092-6566(03)00046-1)00046-1.

Häder, M. 2012. "Data Quality in Telephone Surveys via Mobile and Landline Phone." In *Telephone Surveys in Europe*, edited by S. Häder, M. Häder, and M. Kühne, 247–262. Heidelberg: Springer.

Häder, M. and M. Kühne. 2010. "Mobiltelefonerfahrung und Antwortqualität bei Umfragen." *Methoden, Daten, Analysen* 4: 105–112.

Häder, S., I. Lehnhoff, and E. Mardian. 2010. "Mobile Phone Surveys: Empirical Findings from a Research Project." *ASK. Society. Research. Methods* 19: 3–19.

Heaven, P. and S. Bucci. 2001. "Right-Wing Authoritarianism, Social Dominance Orientation and Personality: An Analysis Using the IPIP Measure." *European Journal of Personality* 15: 49–56. Doi: http://dx.doi.org/10.1002/per.389.

Hopwood, C.J. and M.B. Donnellan. 2010. "How Should the Internal Structure of Personality Inventories be Evaluated?" *Personality and Social Psychology Review* 14: 332–346. Doi: http://dx.doi.org/10.1177/1088868310361240.

Hu, L. and P.M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Structural Equation Modeling* 6: 1–55. Doi: http://dx.doi.org/10.1080/10705519909540118.

Koch, A. 1998. "Wenn 'mehr' nicht gleichbedeutend mit 'besser' ist: Ausschöpfungsquoten und Stichprobenverzerrungen in allgemeinen Bevölkerungsumfragen." *ZUMA-Nachrichten* 42: 66–90.

Kühne, M., M. Häder, and T. Schlinzig. 2009. "Mode-Effekte bei telefonischen Befragungen über das Festnetz und den Mobilfunk: Auswirkungen auf die Datenqualität." In *Umfrageforschung. Grenzen und Herausforderung*, edited by M. Weichbold, C. Wolf, and J. Bacher, 45–62. Wiesbaden: VS Verlag für Sozialwissenschaften.

Lang, F.R., O. Lüdtke, and J.B. Asendorpf. 2001. "Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen." *Diagnostica* 47: 111–121. Doi: http://dx.doi.org/10.1026//0012-1924.47.3.111.

Lang, F.R., D. John, O. Lüdtke, J. Schupp, and G.B. Wagner. 2011. "Short Assessment of the Big Five: Robust Across Survey Methods Except Telephone Interviewing." *Behavior Research Methods* 43: 548–567. Doi: http://dx.doi.org/10.3758/s13428-011-0066-z.

Little, R.J.A. 1988. "A Test of Missing Completely at Random for Multivariate Data With Missing Values." *Journal of the American Statistical Association* 83: 1198–1202. Doi: http://dx.doi.org/10.2307/2290157.

Marsh, H.W., J.R. Balla, and R.P. McDonald. 1988. "Goodness of Fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Size." *Psychological Bulletin* 103: 391–410. Doi: http://dx.doi.org/10.1037//0033-2909.103.3.391.

Marsh, H.W., K.-T. Hau, and Z. Wen. 2004. "In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers of Overgeneralizing Hu and Bentler's (1999) findings." *Structural Equation Modeling* 11: 320–341. Doi: http://dx.doi.org/10.1207/s15328007sem1103_2.

Marsh, H.W., B. Muthén, A. Asparouhov, O. Lüdtke, A. Robitzsch, A.J.S. Morin, and U. Trautwein. 2009. "Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching." *Structural Equation Modeling* 16: 439–476. Doi: http://dx.doi.org/10.1080/10705510903008220.

Marsh, H.W., O. Lüdtke, B. Muthén, T. Asparouhov, A.J.S. Morin, U. Trautwein, and B. Nagengast. 2010. "A New Look at the Big Five Factor Structure Through Exploratory Structural Equation Modeling." *Psychological Assessment* 22: 471–491. Doi: http://dx.doi.org/10.1037/a0019227.

Marsh, H.W., B. Nagengast, and A.J.S. Morin. 2013. "Measurement Invariance of Big-Five Factors Over the Life Span: ESEM Test of Gender, Age, Plasticity, Maturity, and La Dolce Vita Effects." *Developmental Psychology* 49: 1194–1218. Doi: http://dx.doi.org/10.1037/a0026913.

Muthén, L.K. and B.O. Muthén. 1998–2012. *Mplus User's Guide*, 7th ed. Los Angeles, CA: Muthén and Muthén.

Rammstedt, B. 2007a. "Welche Vorhersagekraft hat die individuelle Persönlichkeit für inhaltliche sozialwissenschaftliche Variablen?" *ZUMA-Arbeitsbericht* 2007/01.

Rammstedt, B. 2007b. "The 10-Item Big Five Inventory (BFI-10): Norm Values and Investigation of Socio-Demographic Effects Based on a German Population Representative Sample." *European Journal of Psychological Assessment* 23: 193–201. Doi: http://dx.doi.org/10.1027/1015-5759.23.3.193.

Rammstedt, B., L.R. Goldberg, and I. Borg. 2010. "The Measurement Equivalence of Big Five Factor Markers for Persons with Different Levels of Education." *Journal of Research in Personality* 44: 53–61. Doi: http://dx.doi.org/10.1016/j.jrp. 2009.10.005.

Rammstedt, B. and O.P. John. 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: http://dx.doi.org/10.1016/j.jrp. 2006.02.001.

Rammstedt, B., C. Kemper, M. Klein, C. Beierlein, and A. Kovaleva. 2013. "A Short Scale for Assessing the Big Five Dimensions of Personality: 10 Item Big Five Inventory (BFI-10)." *Methods, Data, Analyses* 7: 233–249. Doi: http://dx.doi.org/10.12758/mda. 2013.013.

Saucier, G. 2000. "Isms and the Structure of Social Attitudes." *Journal of Personality and Social Psychology* 78: 366–385. Doi: http://dx.doi.org/10.1037//0022-3514.78.2.366.

Schneiderat, G. and T. Schlinzig. 2012. "Mobile- and Landline-Onlys in Dual-Frame-Approaches: Effects on Sample Quality." In *Telephone Surveys in Europe*, edited by S. Häder, M. Häder, and M. Kühne, 122–143. Heidelberg: Springer.

Taylor, M.F., J. Brice, N. Buck, and E. Prentice-Lane. 2010. *British Household Panel Survey User Manual: Volume A. Introduction, Technical Report, and Appendices*. Colchester: University of Essex.

Toepoel, V. and P. Lugtig. 2015. "Online Surveys are Mixed-Device Surveys. Issues Associated With the Use of Different (Mobile) Devices in Web Surveys." *Methods, Data, Analyses* 9: 155–162. Doi: http://dx.doi.org/10.12758/mda.2015.009.

Tucker, L. and C. Lewis. 1973. "A Reliability Coefficient for Maximum Likelihood Factor Analysis." *Psychometrika* 38: 1–10. Doi: http://dx.doi.org/10.1007/BF02291170.

Van Hiel, A., M. Pandelaere, and B. Duriez. 2004. "The Impact of Need for Closure on Conservative Beliefs and Racism: Differential Mediation by Authoritarian Submission and Authoritarian Dominance." *Personality and Social Psychology Bulletin* 30: 824–837. Doi: http://dx.doi.org/10.1177/0146167204264333.

West, S.G., J.F. Finch, and P.J. Curran. 1995. "Structural Equation Modeling with Nonnormal Variables: Problems and Remedies." In *Structural Equation Modeling: Concepts, Issues and Applications*, edited by R.H. Hoyle, 37–55. Thousand Oaks, CA: Sage.

Ximénez, C. 2006. "A Monte Carlo Study of Recovery of Weak Factor Loadings in Confirmatory Factor Analysis." *Structural Equation Modeling* 13: 587–614. Doi: http://dx.doi.org/10.1207/s15328007sem1304_5.