

## EVALUATION OF A VERY SHORT TEST TO MEASURE THE BIG FIVE PERSONALITY FACTORS ON A FLEMISH SAMPLE

**Anikó Lovik \***  
*University of Leuven,  
Belgium*

**Geert Verbeke**  
*University of Leuven,  
Hasselt University,  
Belgium*

**Geert Molenberghs**  
*Hasselt University,  
University of Leuven,  
Belgium*

### *Abstract*

*In this paper, we study the psychometric properties of the ten-item version of the Big Five Inventory as a subset of the original BFI in a Flemish sample. The data come from the Divorce in Flanders study and consist of a full sample of 7533 individuals from 4460 families. Factor analysis shows the presence of the Big Five factor structure with very high primary loadings for most items. However, one of the Agreeableness items loads exclusively on the Extraversion factor and within-factor correlations are also low. Despite this, the BFI-10 correlates well with the BFI. Therefore, while more research is needed before validity and reliability of the Dutch-language version of the test can be concluded, it is clear that the BFI-10 may prove very effective in the assessment of the Big Five factors in the Flemish and Dutch cultures when assessment with longer questionnaires is not feasible.*

Keywords: Big Five; Big Five Inventory; BFI-10; divorce in Flanders; short personality test

### **Introduction**

In recent years, personality inventories have been shortened considerably. While long tests such as the California Psychological Inventory

---

Correspondence concerning this paper should be addressed to:

\* Anikó Lovik MA, MSc., I-BioStat, KU Leuven - University of Leuven. Address: Kapucijnenvoer 33 blok d, 3000 Leuven, Belgium. Website: [www.ibiostat.be](http://www.ibiostat.be) Phone: 003216373361 Email: [aniko.lovik@kuleuven.be](mailto:aniko.lovik@kuleuven.be) [orcid.org/0000-0002-6397-5011](https://orcid.org/0000-0002-6397-5011)

(434 items, Gough, 1987) or the Minnesota Multiphasic Personality Inventory (338 items; Ben-Porath & Tellegen, 2008) are still in use, there is also a huge demand for shorter and thus more accessible tests. Many of the well-established tests became shorter over time or were supplemented by a shorter version. For example, the Revised NEO Personality Inventory has 240 items while its shorter counterpart (NEO Five-Factor-Inventory, NEO-FFI) contains 60 items (Costa & McCrae, 1992; 2010).

Accordingly, to make the already brief questionnaire even shorter, Rammstedt and John (2007) created a ten-item version (BFI-10) of the 44-item Big Five Inventory (BFI; John & Srivastava, 1999) by selecting two items (one positive and one negative) from each factor from the original inventory. This was not the first attempt to create an extremely short questionnaire, Gosling, Rentfrow and Swann published a Big Five measure called Ten-Item Personality Inventory (TIPI) in 2003, consisting of pairs of adjectives. The TIPI has been adapted and validated for several languages including German (Herzberg & Brähler, 2006), Dutch (Hofmans, Kuppens, & Allik, 2008) and Japanese (Oshio, Abe, Cutrone, & Gosling, 2013). In fact, even a five-item version was considered but the authors advise against using it for several reasons as it is less reliable, has weaker correlations with other Big Five measures and other variables, it estimates the latent factors with a single item only (resulting in not being able to estimate the error terms) and also cannot balance positive and negative items (and hereby assessing and controlling for acquiescence bias) (Gosling, Rentfrow, & Swann, 2003). Furthermore, Rammstedt, Koch, Borg and Reitz (2004) note that good test-retest measures may be caused by memory effects which is especially problematic for five-item scales. The ten-item test still only requires a couple of minutes from the participants while avoiding most of the aforementioned limitations (Gosling, Rentfrow, & Swann, 2003; Herzberg & Brähler, 2006). In summary, both the TIPI and the BFI-10 have acceptable psychometric properties, especially considering their lengths. In one (relatively small) study, Furnham (2008) found that the validity of the TIPI was even better than that of the NEO-FFI.

Naturally, creating very brief measures has its drawbacks, a loss of validity and reliability are most often mentioned. In recent years, the negative consequences of using short (ened) tests have been extensively studied in the literature (Credé et al., 2012; Krueger, Emons, & Sijtsma, 2013), which is not too surprising looking at the noticeable rise in the number and use of very brief

tests. It is out of the scope of the present work to discuss pros and cons of short measures in detail, nevertheless, it is important to remember that the aim of the shortened versions is not necessarily to replace the longer versions but to provide researchers and other users with an effective alternative when the use of a longer test is not feasible.

However, one limitation that cannot be avoided with a ten-item test is that these tests can still only assess the highest level of the personality structure (the five factors) and are unable to address all facets of the factors separately (each of the five factors collects six facets, thus a minimum of 30 items would be needed). While these facets are all related, they represent different aspects of the factors and therefore, it would not be surprising to see lower correlations between the items within each of the factors than for the long scale. Since these correlations are sometimes used as reliability measures of the factors for two-item scales, these reliability measures are expected to be low as well.

Despite lower reliability, an argument in favour of the BFI-10 (and short tests in general) is parsimony. As Furnham (2008, p. 316) noted, "If it can be shown that 1-, 2-, or 3-item measures have as high a predictive validity as 12-, 20-, or 30-item measures, it seems difficult to justify the use of the latter." In the same paper, he found that the TIPI has better validity than the NEO-FFI, which has 12 items per scale instead of two. However, it should be noted that this study was limited by the small and homogenous sample and the fact that the participants completed only four personality tests of which three were very short.

To increase validity and reliability, the BFI-10 has been simultaneously developed in English and German by Rammstedt and John (2007) and has also been translated and tested in a Chinese sample (Carciofo, Yang, Song, Du, & Zhang, 2016). The latter results were inconclusive with an unclear factor structure for all but one sample but otherwise acceptable reliability measures. Therefore, it remains to be seen whether the BFI-10 translates well to languages that are very different from Germanic languages. Results from the Japanese TIPI suggest it may be possible to adapt very short Big Five questionnaires (Oshio, Abe, Cutrone, & Gosling, 2013).

## **Objectives**

While the BFI has been translated into Dutch and validated in The

Netherlands (Denissen, Geenen, van Aken, Gosling, & Potter, 2008) and, to some extent, in Flanders (Lovik, Nassiri, Verbeke, Molenberghs, & Sodermans, 2017), to our knowledge, no data has been published regarding the ten-item version. The aim of this paper is to assess the psychometric properties of the short version of the BFI in a Flemish sample.

## Methods

### *Participants*

For this analysis, a subset of the Divorce in Flanders (DiF) project has been used. In this huge regional study, couples who married between 1971 and 2008 and some of their family members were surveyed using a battery of tests. The original sample was drawn randomly from the marriages registered in the Belgian National Registry with a 2/3 oversampling of divorces and stratified for the year of the marriage. It contains data from more than 10,000 individuals from about 4500 families. Our sample is restricted to the original couples and their children (N=7533) excluding grandparents and new partners. Of the 7533 participants 3999 are female and 3529 are male with an average age of 42.4 years (range 14 to 72, SD=11.9). A detailed description of sampling and data collection can be found in Mortelmans et al. (2011).

### *Instruments*

As part of the DiF study, the personality of each participant was assessed with the validated Dutch version (Denissen et al., 2008) of the Big Five Inventory (BFI; John & Srivastava, 1999). The BFI is a short personality test, which is commonly used to assess personality measuring the Big Five personality factors (Neuroticism (N), Extraversion (E), Openness to Experience (O), Conscientiousness(C) and Agreeableness (A)). The inventory contains 44 items and participants rated their agreement with each item regarding their self-perception using a five-point Likert scale (coded 1 (“strongly disagree”) to 5 (“strongly agree”). A subset of these responses was used to assess the BFI-10. The consequences are elaborated on in the Discussion. The negatively framed items were reversed before any analysis has taken place. It is important to note that the BFI-10 does not contain any (grammatically) negatively worded items, which could make the test cognitively taxing.

*Procedure and design*

The BFI-10 has been evaluated using descriptive statistics, Pearson product-moment correlations, Cronbach’s alphas, and exploratory (EFA) and confirmatory (CFA) factor analyses. The EFA has been performed using principal component extraction with both varimax and direct oblimin rotations.

To avoid problems with clustering (responses from members of the same family are expected to be correlated), especially considering the heritability of personality traits (*see* Loehlin, McCrae, Costa, & John, 1998) an uncorrelated, random subsample with 4457 individuals sampled from the original couples who were invited to the study, was also analysed. The analyses did not differ substantially from those presented in this paper.

**Results and discussion**

For each item, less than 1% of the responses was missing and, by using case-wise deletion, this results in the omission of 4% of the observations. 7223 (out of 7533) observations were used for the analysis. Correlations between factors can be seen in Table 1. Pearson correlations are slightly higher for the smaller dataset but are in general low but significant (except for the factor pair O-A). The highest correlation is between factor N and E (-0.22). The correlations between the two items per factor (all are significant <0.05) 0.45, 0.27, 0.16, 0.24, 0.05 for factors N, E, O, C and A, respectively. These correlations are much lower than the correlations of the factors in the BFI (Lovik et al., 2017). The rest of the 10x10 inter-item correlations are all negligible.

Table 1. Pearson correlations per factor

Factor	N	E	O	C	A
Neuroticism					
Extraversion	-0.22**				
Openness to Experience	-0.09**	0.09**			
Conscientiousness	-0.14**	0.13**	-0.03*		
Agreeableness	-0.15**	0.15**	0.00	0.06**	

Note: \* p<0.05; \*\* p<0.0001

Table 2 has the complete-case factor analysis with principal component extraction after oblimin rotation of the BFI-10 for the entire dataset (7223 out

of 7533 people from 4460 families. Retaining five factors, this model explains 65.29% of the variance, which is much higher than what was seen in the BFI for the same data (42.21%). Oblimin rotation was chosen because it is valid even if the factors are correlated. It should be noted that using varimax rotation results in only very slight changes in the factor loadings.

Table 2 also contains the means and standard deviations for each item. Means vary between 2.36 (item N-: *is relaxed, handles stress well*) and 4.30 (item C+: *does a thorough job*) with in total three items above 4, two of these are from factor C. The higher means for factor C may be caused by the social desirability or cultural aspects of the items but also because of the age of the respondents is higher than in most studies as it is known that factor C means tend to increase with age (Rammstedt, 2007).

Table 2. Means and standard deviations, factor analysis of the full dataset

Questionnaire Item	M	SD	N	E	O	C	A
(N-) is relaxed, handles stress well	2.36	1.24	<b>.83</b>	-.04	.00	-.03	.13
(N+) gets nervous easily	2.99	1.27	<b>.87</b>	.07	-.06	.04	-.15
(E-) is reserved	3.24	1.28	-.21	<b>.53</b>	-.17	.07	.03
(E+) is outgoing, sociable	4.06	.92	.06	<b>.79</b>	-.05	.16	-.03
(O-) has few artistic interests	3.11	1.44	-.04	-.10	<b>.93</b>	.09	.12
(O+) has an active imagination	3.52	1.23	-.04	.29	<b>.43</b>	-.06	<b>-.46</b>
(C-) tends to be lazy	4.02	1.15	.01	.03	-.02	<b>.71</b>	<b>.41</b>
(C+) does a thorough job	4.3	.94	-.02	.02	.10	<b>.82</b>	-.26
(A-) tends to find fault with others	3.12	1.24	-.05	.13	.11	-.04	<b>.85</b>
(A+) is generally trusting	3.73	1.12	.05	<b>.63</b>	.12	-.17	.09

For most items, the primary loadings are moderate to high with an upper range of 0.93 (item O-: *has few artistic interests*) but two items have disturbing cross loadings. First, the other Openness item (O+: *has an active imagination*) has a -0.46 loading on Agreeableness which is higher than the primary loading and is surprising considering the results of the BFI were the cross-loadings were rather low. However, this cross loading is also present in the results of Rammstedt, Kemper, Klein, Beierlein and Kovaleva (2012) in a German sample (there the primary loading for this item is -0.54 and the cross-loading 0.45 on factor E). Second, the positive item of factor A (*is generally trusting*) has a nil primary loading and a high cross loading of 0.63 on Extraversion. This latter item was already problematic in the BFI loading on three factors (E, O and A). However, cross-loadings are very low compared to

the BFI, in general. Interestingly, the Dutch TIPI - while containing different items - also shows a similar structure with an Openness item with high and multiple cross-loadings and a problematic Agreeableness item (Hofmans, Kuppens, & Allik, 2008).

The primary factor loadings presented are similar to those presented by Rammstedt and colleagues (2012) with the notable exception of item A+ (*is generally trusting*) which in the German sample has a very high primary loading (-0.81) and has negligible cross-loadings. It is notable that the item is different in the German version as it is longer (*Ich schenke anderen leicht Vertrauen, glaube an das Gute im Menschen* compared to ... *is generally trusting*). This could have been due to the different rotation. However, comparison with the varimax-rotated factors gave the same results<sup>1</sup>.

The confirmatory factor analysis confirms the above findings ( $\chi^2 = 1691.44, df = 30, p < 0.001, GFI = 0.95, RMSEA = 0.088$ ).

Reliability measures per factor are presented in Table 3. While the BFI has high Cronbach's alphas with 8-10 items per scale, the two-item factors are not very high at all despite the high factor loadings. While lower Cronbach's alphas are to be expected in such a short survey, even the lowest Cronbach's alpha reported for the TIPI by Herzberg and Brähler (2006) is higher (0.24 for Agreeableness). It should be noted that, since there are only two items per scale, the inter-item correlation cannot be high, especially considering how diverse traits each factor represents (Ziegler, Kemper, & Krueger, 2014). Thus, for two-item scales Cronbach's alphas consistently underestimate the reliability and Pearson correlations of the items within the factors are sometimes used in this context. However, these are even lower (*see* Table 3). Eisinga, Grotenhuis and Pelzer (2013) found in a simulation study that the Pearson correlations are not an adequate measure for reliability of two-item scales. As the factors collect a bunch of personality traits that are related but can still be quite different (for example the two items belonging to factor A are related to trust (*is generally trusting*) and being critical (*tends to find fault with others*), this is not surprising.

---

<sup>1</sup> This analysis, along with further results, can be found on the website of the Interuniversity Institute for Biostatistics and statistical Bioinformatics (<https://ibiostat.be>).

Table 3. Cronbach's alphas and Pearson correlations per factor

Factor	Cronbach's alpha of the BFI-10	Pearson correlations for the two items of a factor	Cronbach's alpha of the BFI
Neuroticism	.62	.45*	.80
Extraversion	.42	.27*	.79
Openness to Experience	.28	.16*	.78
Conscientiousness	.39	.24*	.78
Agreeableness	.09	.05*	.67

Note: \* p<0.0001

Table 4 compares descriptive per factor for the BFI-10 and the BFI and provides part-whole correlations per factor. These latter are rather low compared to the ones given in Rammstedt and John (2007), where most of these are above 0.80 and none below 0.70. However, the pattern is similar with lower correlations for the factor Openness and especially Agreeableness and higher ones for the rest of the factors.

Table 4. Means, standard deviations and part-whole correlations of the BFI-10 and the BFI

	BFI-10	BFI	Pearson correlation between BFI-10 and BFI
Neuroticism	2.68 (1.07)	2.77 (.78)	.84
Extraversion	3.65 (.89)	3.73 (.71)	.79
Openness to Experience	3.31 (1.03)	3.58 (.65)	.76
Conscientiousness	4.16 (.83)	3.87 (.64)	.78
Agreeableness	3.42 (.86)	3.92 (.56)	.69

## Conclusion

In this analysis, some of the psychometric properties of the BFI-10 were checked in a sample of respondents who completed the BFI questionnaire. While in the English and German language samples the BFI-10 is a reliable and good substitute of its longer version in case an extremely short measure is needed, the Dutch language version is conflicting. Potentially, the usability of the BFI-10 in other cultures can be limited if the translations cling too much to the original items. The Dutch BFI-10 may be more useful with at least the



problematic Agreeableness item replaced with something else. Rammstedt and John (2007) suggested adding a third factor A item: "...is considerate and kind to almost everyone."). Unfortunately, this is not a viable solution for the Dutch BFI-10 (both scale, as item level results are worse than for the ten-item test). An obvious first choice to test would be the longer version of the item, which is used in the German version.

While the sample is rather heterogeneous and big, without exploring the test-retest reliability and the external validity of the BFI-10 the generalisability of these results is limited. Another limitation is that the BFI-10 was measured as part of the original Big Five Inventory so the participants originally responded to the BFI and not to both inventories. While the test-retest reliability of the BFI is generally very high, it is not perfect as assumed here when calculating the correlation between the BFI and the BFI-10. Therefore, these correlations are overestimated. It should be noted that we do not expect the presence of the other items to have a major influence on the responses to the items of the BFI-10. Partly, because many of the target items are in the beginning of the BFI (e.g. items 2, 3, 6, 9 of the BFI are also the first items of a specific factor on the BFI - for factors A, C, E and N, respectively) but also because these items are rather different from each other to be able to assess the different aspects of personality and the personality factors. Therefore, a priming effect is rather unlikely to occur.

In summary, while this analysis has limitations it provides information about what one may expect when adapting the BFI-10 to the Dutch language and also gives ground to further research.

### References

- Ben-Porath, Y. S., & Tellegen, A. (2008). *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2 Restructured Form): Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press.
- Carciofo, R., Yang, J., Song, N., Du, F., & Zhang, K. (2016). Psychometric Evaluation of Chinese-Language 44-Item and 10-Item Big Five Personality Inventories, Including Correlations with Chronotype, Mindfulness and Mind Wandering. *PloS one*, *11*(2), e0149963. doi: 10.1371/journal.pone.0149963

- 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.
- 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(4), 874-888. doi: 10.1037/a0027403.
- Denissen, J. J. A., Geenen, R., van Aken, M. A. G., Gosling, S. D., & Potter, J. (2008). Development and Validation of a Dutch Translation of the Big Five Inventory (BFI). *Journal of Personality Assessment*, 90(2), 152-157.
- Eisinga, R., Grotenhuis, M. T., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58, 637-642. doi: 10.1007/s00038-012-0416-3
- Furnham, A. (2008). Relationship among four Big Five measures of different length. *Psychological Reports*, 102(1), 312-316.
- Gough, H. G. (1987). *California Psychological Inventory Administrator's Guide*. Palo Alto, CA: Consulting Psychologists Press, Inc.
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, 37(6), 504-528. doi: 10.1016/S0092-6566(03)00046-1
- Herzberg, P. Y., & Brähler, E. (2006). Assessing the Big-Five personality domains via short forms. *European Journal of Psychological Assessment*, 22(3), 139-148. doi: 10.1027/1015-5759.22.3.139
- Hofmans, J., Kuppens, P., & Allik, J. (2008). Is short in length short in content? An examination of the domain representation of the Ten Item Personality Inventory scales in Dutch language. *Personality and Individual Differences*, 45(8), 750-755. doi: 10.1016/j.paid.2008.08.004
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102-138). New York, NY: Guilford.
- Kruyen, P. M., Emons, W. H., & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing*, 13(3), 223-248.
- Loehlin, J. C., McCrae, R. R., Costa, P. T., & John, O. P. (1998). Heritabilities of common and measure-specific components of the Big Five personality

- factors. *Journal of Research in Personality*, 32(4), 431-453.
- Lovik, A., Nassiri, V., Verbeke, G., Molenberghs, G., & Sodermans, A. K. (2017). Psychometric properties and comparison of different techniques for factor analysis on the Big Five Inventory from a Flemish sample. *Personality and Individual Differences*, 117, 122-129.
- McCrae, R. R., & Costa, P. T. (2010). *NEO Inventories professional manual*. Lutz, FL: Psychological Assessment Resources.
- Mortelmans, D., Pasteels, I., Bracke, P., Matthijs, K., Van Bavel, J., & Van Peer, C. (2011). *Scheiding in Vlaanderen [Divorce in Flanders]*. Leuven, Belgium: Acco.
- Oshio, A., Abe, S., Cutrone, P., & Gosling, S. D. (2013). Big Five content representation of the Japanese version of the Ten-Item Personality Inventory. *Psychology*, 4, 924-929. doi: 10.4236/psych.2013.412133
- Rammstedt, B. (2007). The 10-Item Big Five Inventory. *European Journal of Psychological Assessment*, 23(3), 193-201.
- 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(1), 203-212.
- Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2012). Eine kurze Skala zur Messung der fünf Dimensionen der Persönlichkeit: Big-Five-Inventory-10 (BFI-10) [A short test to measure the five dimensions of personality: Big Five Inventory 10 (BFI-10)]. *GESIS-Working Papers*, 2012(22).
- Rammstedt, B., Koch, K., Borg, I., & Reitz, T. (2004). *Entwicklung und validierung einer kurzskala für die messung der big-five-persönlichkeitsdimensionen in umfragen [Development and validation of a short scale for measuring the big 5 personality dimensions in surveys]*. PID: <http://nbn-resolving.de/urn:nbn:de:0168-ssoar-207618>  
<http://www.ssoar.info/ssoar/handle/document/20761>
- Ziegler, M., Kemper, C. J., & Krueger, R. F. (2014). Short scales - Five misunderstandings and ways to overcome them. *Journal of Individual Differences*, 35(4), 185-189. doi:10.1027/1614-0001/a000148

Received August 23, 2017

Revision September 19, 2017

Accepted October 10, 2017