

Psychometric properties and comparison of different techniques for
factor analysis on the Big Five Inventory from a Flemish sample

Peer-reviewed author version

Lovik, Aniko; NASSIRI, Vahid; VERBEKE, Geert; MOLENBERGHS, Geert &
Sodermans, An Katrien (2017) Psychometric properties and comparison of different
techniques for factor analysis on the Big Five Inventory from a Flemish sample. In:
PERSONALITY AND INDIVIDUAL DIFFERENCES, 117, p. 122-129.

DOI: 10.1016/j.paid.2017.05.048

Handle: <http://hdl.handle.net/1942/24449>

Psychometric properties and comparison of different techniques for factor analysis on the Big Five Inventory from a Flemish sample

Second draft

10 March 2016

Abstract

In this paper we examine the Dutch version of the Big Five Inventory, a short questionnaire used to measure the Big Five factors of personality, on a Flemish sample coming from the Divorce in Flanders study. Our aim is twofold. First, we will show that the sample has good psychometric properties and a clear factor structure comparable to what has been found in a previous Dutch sample (Denissen et al, 2008) and in other international samples of the Big Five research literature.

Second, we compare the usual method of analysis, namely principal components analysis (PCA) with varimax rotation, to several methods which each address a common problem in factor analysis. We compare the original PCA results to PCA with a different, non-orthogonal rotation (addressing the problem of correlated factors), after normalisation (considering individual differences in response style) and with a sample that was completed using multiple imputation (to account for missing observations and bias due to listwise deletion). The four factor analysis results differ minimally.

Keywords: Divorce in Flanders, Big Five Inventory, factor analysis, missing data, multiple imputation, varimax rotation, oblimin rotation, acquiescence score

1 Introduction

Questionnaires have been favoured tools in psychological and sociological research from the beginning of the last century for several reasons. Among others, they allow for relatively cheap and fast data collection and the possibility to reach populations that are unavailable via other methods. The increasing popularity of survey methods resulted in the development of statistical methods that allow reliable ways for analysing the collected data. Factor analysis has been a widely used method in social sciences to handle the massive amounts of survey data since the beginning. From the first half of the 20th century more and more personality questionnaires were developed to help clinicians and researchers to describe and understand the structure of personality (McCrae & John, 1992). The emergence of the Big Five model of personality (McCrae & Costa, 1987; Digman, 1990) resulted in the development of several questionnaires (for a review see Widiger & Trull, 1997; John & Strivastava, 1999) such as the Big Five Inventory (John & Strivastava, 1999).

The Big Five is the most established and well-validated model of personality (John & Srivastava, 1999; McCrae & Costa, 1987). It consists of Neuroticism (Ne), Extraversion (Ex), Openness to

Experience (Op), Conscientiousness (Co), and Agreeableness (Ag) (Costa & McCrae, 1992; Digman, 1990; John & Strivastava, 1999; McCrae & Costa, 1987). Neuroticism is characterised by anxiety, nervousity, sadness and is the polar opposite of emotional stability. Extraversion is linked to sociability, assertiveness, and energy. Openness to Experience (Op) is refers to originality, curiosity, creativity and intelligence. Conscientiousness (Co) is related to orderliness, responsibility, and dependability. Agreeableness (Ag) implies characteristics such as good-naturedness, modesty, compliance, cooperativeness, and trust (John & Srivastava, 1999).

The five factors are generally found across cultures (Hofstee, Kiers, de Raad, Goldberg, & Ostendorf, 1997) and do not change considerably with age (Asendorpf & van Aken, 2003).

Although there are several well-known questionnaires to measure the Big Five factors, most of them are rather long which can seriously limit their usability as most survey research has to take into account time and space constraints. John and Strivastata developed the Big Five Inventory (BFI) in 1999 to address these problems. Their questionnaire is available on the internet, making it useful for web-surveys, and short, only consisting of 44 items which can be rated in less than 15 minutes, while still reliable.

The analysis of questionnaires consisting of many items raises several methodological problems. First of all, data reduction is needed. Questionnaires, even those composed of a dozen questions, ensue a complex structure of variables where covariances between responses have to be taken into account. Longer inventories are therefore often analysed using principal components or factor analysis which reduces the dimensionality by imposing a certain number of latent factors based on inter-item correlations. This is especially useful in personality research where the different personality traits can be reduced to personality factors. For example, personality traits expressed with adjectives such as "assertive", "active", "energetic", "adventurous", "outspoken" and "enthusiastic" can be collected in one term: extraversion.

To be able to easily interpret these unobserved factors it is useful to assume factors do not share common variance. In fact, the Big Five personality factors make this assumption: the five factors are independent and this view is also implied by using an orthogonal (varimax) rotation which assumes uncorrelated factors. Using varimax rotation while the factors have high correlations may result in a false interpretation of the results.

Another issue that may cause distortion in the results is related to the personal differences in filling-in questionnaires. Some participants tend to give extreme answers while others try to stay close to the middle and never use the two endpoints of a scale. At the same time some participants may prefer to give overwhelmingly positive or overwhelmingly negative responses. While with a big sample size many of the individual differences are balanced out, some analyses are sensitive to these kinds of extreme responses. Normalising the data by taking into account the individual differences in the centre and the variability of the answers may give a better understanding of the population. A special normalisation, called the acquiescence index, has been presented by the inventors of the BFI in 2008 to account for this type of bias (Oliver, Naumann, & Soto, 2008) in this inventory. The method is also called ipsatisation.

Lastly, a general problem with survey data is the incompleteness of the collected data. The participants may forget or may not wish to fill in all answers resulting in a certain amount of missing data. An important factor is the type of missingness which can be divided into three categories according to Rubin (1976). Data is missing completely at random (MCAR) if the values that are missing are independent of both observed and unobserved data. If the missingness can be

fully accounted for by variables where there is complete information then the data are missing at random (MAR). Else, data are missing not at random (MNAR).

While some statistical methods can cope with a certain amount of (randomly) missing data, others cannot which often results in the omission of observations from the analysis or even in the exclusion of an entire set of data. For example, because of one missing item we might need to exclude all answers of a participant from the analysis (this is called listwise deletion) even though more than 95% of the items are answered. Listwise deletion by definition reduces the power of the study. In case of the BFI an even bigger concern is the bias resulting from excluding participants with one or more missing items. Because we want to measure personality dimensions, participants who give themselves higher scores on the following items and can be somewhat careless (item 8), lazy (item 23), less organised (item 18) and/or easily distracted (item 43) may be more likely to miss a response which will be reflected on the same factor: Conscientiousness.

A possible solution is multiple imputation (explained in Section 6), a statistical method that allows the use of all observed data. A major difference between the analysis with listwise deletion and with multiple imputation can be an indicator of aforementioned bias.

The paper is organized as follows. The Divorce in Flanders dataset is first introduced in Section 2. The psychometric properties and factor analysis of the original data using the usual varimax rotation is presented next in Section 3 followed by results of the analysis with oblimin rotation in Section 4. In Section 6 and Section 5 we present the results of the ipsatised (or normalised) and the multiply imputed data. Discussion and concluding remarks are given in Section 7.

2 Divorce in Flanders study

The dataset we use for analysis is a subsample from the 'Divorce in Flanders' (DiF) project which contains a sample of marriages registered between 1971 and 2008 with oversampling of divorces (1/3 intact and 2/3 dissolved marriages at the sampling date) drawn from the Belgian National Register. Family members across three generations were surveyed during the original data collection, more than 10.000 people. In this paper we use data from 4457 families, 7533 people in total (3362 mothers, 2920 fathers and 1251 children). We excluded new partners and parents of the selected sample. The reason for the data collection, among others, was to assess eventual patterns of matching personality traits between family members, predicting personality traits by studying the intergenerational transmission of personality, associating personality traits with fertility and personality traits with divorce.

As part of this study the personality of each participant was assessed with the validated Dutch version (Denissen, Geenen, van Aken, Gosling, & Potter, 2008) of the Big Five Inventory (BFI, John & Strivastava, 1999), a personality test which is a commonly used tool to assess personality measuring the five factors of personality (Goldberg, 1990; Widiger & Trull, 2010, etc.).

We considered individuals and not families as units thereby neglecting the potential correlatedness of the responses of family members. Consequences of this decision will be discussed in the Discussion.

From each family one child (aged 14 or more) was selected randomly to participate. If there was only one child above the age of 14 in the family that child was automatically chosen. If the children in a specific family were all younger than 14 years at the time of the study they were not included in the study. However, the age of the children is of no concern as children give relatively stable

responses to Big Five questionnaires from middle childhood (Asendorpf & van Aken, 2003). In a longitudinal study they found that all same-factor reliability measures at age 4-6 (teacher Q-scores) and at age 10 (parental Q-scores) were above 0.6. According to the Dutch BFI data, the factor structure is not changing substantially with age (Denissen, Geenen, van Aken, Gosling, & Potter, 2008).

The data were collected using oral interviews, web-surveys and mail questionnaires. While incompleteness is very common in web-surveys where the researchers have no control over the participants who can abandon the questionnaire at any time, according to Gosling, Vazire, Srivastava, and John (2004), the collected data is not necessarily of lower quality than for paper-and-pen questionnaires and internet-based findings do not differ substantially from data collected in other ways. We compared the three subgroups and found no difference in descriptives and factor loadings between the different types of data collection methods in this particular case.

Participants were asked to rate their agreement with each item regarding their perceptions of themselves using a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The 44-item questionnaire contains 14 reversed items (indicated in the tables with r after the item number). The items were reversed before any analysis has taken place.

3 Psychometric properties of the Divorce in Flanders data

3.1 Descriptive statistics of the original dataset

Item means (presented in Table 3 and in Table 5) fell out of the 2 to 4 range for nearly half of the items compared to the three extreme items in the validated Dutch BFI scale (Denissen, Geenen, van Aken, Gosling, & Potter, 2008) meaning that our sample tended to give more extreme answers (standard deviations were similar). Of these 18, all were above 4. As suggested by Denissen et al (2008) such high values may be related to high levels of social desirability of the items. In general, the means in our study were higher than those found by Denissen et al (1999). The standard deviations were all above 0.80 (ranging from 0.83-1.44) and in general very similar to the ones by Denissen et al (1999).

Descriptives at factor level are in the upper part of Table 1. They were calculated based on the non-missing values. There is one missing value in the Neuroticism and one missing value in the factor Conscientiousness meaning that participants failed to answer any questions belonging to these factors. The items are not per factor but mixed in the inventory so it not likely to have been caused by forgetfulness or problems with the data collection. Factor means are rather high with four factors (all except Ne) having a mean above 3.50 and the remaining factor, Neuroticism below 3.00. Not surprisingly, factor Agreeableness has the highest average (3.92) most likely because of the high social desirability of the personality traits associated with this factor. Standard deviations range from 1.05 to 1.14. All minima are 1 and all maxima are 5 which means that for each factor at least one participant chose or only 'strongly disagree' or only 'strongly agree' for each item belonging to that factor.

While participants were asked to respond to all items, 747 questionnaires had at least one missing response. For this reason, out of 7533 participants only 6786 subjects' data could be used for the factor analysis meaning that approximately 10% of all cases were excluded from the analysis. Given the sample size, this is technically not a concern. However, it could be a source of bias as explained previously.

Table 1: *Descriptives and correlations between the factors.*

Variable	N	Mean	SD	Min	Max
Neuroticism	7532	2.76	1.22	1.00	5.00
Extraversion	7533	3.73	1.11	1.00	5.00
Openness to Experience	7533	3.58	1.14	1.00	5.00
Conscientiousness	7532	3.87	1.05	1.00	5.00
Agreeableness	7533	3.92	1.06	1.00	5.00

Correlations					
	Ne	Ex	Op	Co	Ag
Neuroticism	1.00				
Extraversion	-0.37	1.00			
Openness to Experience	-0.15	0.32	1.00		
Conscientiousness	-0.23	0.30	0.14	1.00	
Agreeableness	-0.30	0.27	0.13	0.35	1.00

All correlations are significant <.0001

3.2 Factor analytic results and comparison to other samples

We used principal components analysis with varimax rotation as it is customary for validating the Big Five Inventory (Szirmák, & De Raad, 1994; Benet-Martinez, & John, 1998; Lang, Lüdtke, & Asendorpf, 2001; Denissen, Geenen, van Aken, Gosling, & Potter, 2008; Plaisant, Courtois, Réveillère, Mendelsohn, & John, 2010;). Orthogonal rotations such as the varimax rotation are used when the factors are supposed to be uncorrelated or have low interfactor correlations. The varimax rotation has the property of maximizing the variance of the factor loading thus making the (small amount of) high factor loadings higher and the (larger amount of) low factor loadings lower (Tabachnick, & Fidell, 2001), resulting in a clear(er) factor structure. The uncorrelated factors are also easier to interpret and replicate. In our data the correlations between the factors are relatively high, ranging in absolute value between 0.13 and 0.37 (see lower part of Table 1) which indicates that a non-orthogonal rotation may be better.

The results of the factor analysis together with the item means and standard deviations are presented in Table 3. This model explained 42.2% of the variance, with the primary loadings ranging from 0.21 to 0.72 and with an average value of 0.54 which is nearly 0.1 lower than the mean primary factor loading in the previous Dutch sample. However, the average primary loadings per factor (Table 2) are very similar to those reported by Benet-Martinez and John (1998) for an English sample (0.58; 0.67; 0.50; 0.57 and 0.51, respectively, with a mean of 0.57) and better than the German counterpart (0.50; Lang, Lüdtke, & Asendorpf, 2001).

All factors have at least one item with a high (above 0.65) primary loading and negligible cross-loadings (below 0.15) for all other factors. At the same time all, except Agreeableness have complex items meaning that an item has loadings equal or higher than 0.3 on more than one

Table 2: *Mean factor loadings and cross-loadings.*

	N	E	O	C	A	Mean
Average primary loadings	0.60	0.57	0.51	0.57	0.46	0.54
Average cross-loadings	0.12	0.14	0.17	0.14	0.12	0.14

factor. In total, there are 12 complex factors and 3 items that have higher cross-loadings than primary loadings. Item 29 ("can be moody") has a primary loading of 0.4 (Ne) and a -0.50 cross-loading on Ag. Item 35r has a primary loading of 0.22 on Openness and a 0.23 on Neuroticism but this is a problematic item in the Dutch version, too, having no factor loading above 0.23 at all. Lastly, item 42. ("likes to cooperate with others") has a primary loading of 0.25 on Agreeableness but seems more to belong to Extraversion (0.39). There are two items, 35r (Op, "prefers work that is routine") and 7(Ag, "is helpful and unselfish with others") without a factor loading above 0.3. Item 35r is similar in the Dutch version of the BFI while item 7 has there a high primary loading (0.62) and null-loadings otherwise. However, the highest loading of item 7 is still at the factor Agreeableness where it is supposed to be according to the intentions of the creators of the BFI. In general, while the primary loadings are lower than expected from the previous Dutch sample, the factor structure seems quite clear.

4 Non-orthogonal rotation (oblimin)

The threshold to choose non-orthogonal rotation suggested by Tabachnick and Fidell (2001) is 0.32 for at least one interfactor correlation. Considering the values presented in Table 1, we also performed principal components analysis with oblimin rotation. However, it is to be noted that the factor loadings were very similar as presented in Table 5 (with a maximum difference of 0.06, mean difference of 0.01 and standard deviation of 0.02 on the primary factor loadings).

This model five-factor solution explained 42.2 % of the variance, with the primary loadings ranging from 0.23 to 0.75 and with an average value of 0.53. The average primary and cross-loadings per factor (Table 4) are very similar to our previous results with varimax rotation with a maximal difference of 0.02. However, the loadings in case of the oblimin rotation are always lower (for primary and cross-loadings). The average primary factor loadings are slightly lower for three factors (N, E, O) each time with 0.01 for the oblimin rotation and the average cross-loadings are also lower for all factors (with a difference of 0.01 for N, C and A factors and 0.02 for E and O). There are 12 complex items and 10 items with negligible (below 0.16) cross-loadings and high primary loadings.

5 Ipsatised scores

In their 2008 paper the creators of the Big Five Inventory introduced a new scoring method to account and compensate for the individual differences in answering questionnaires, the so called Acquiescence Index (Soto, John, Gosling, & Potter, 2008; John, Naumann, & Soto, 2008) based on the concept of response styles and acquiescence (Jackson & Messick, 1958). The correction is for the tendency of agreement or disagreement with the items, independently of the item content (Jackson & Messick, 1958).

Table 3: *Factor loadings for PCA with varimax rotation.*

Questionnaire Item	Mean	SD	N	E	O	C	A
19. worries a lot	3.30	1.27	0.72	-0.09	0.05	0.04	-0.05
14. can be tense	3.56	1.15	0.70	-0.06	0.08	0.10	-0.17
9r. is relaxed, handles stress well	2.36	1.24	0.63	-0.15	-0.23	-0.15	-0.01
39. gets nervous easily	2.99	1.27	0.75	-0.05	-0.02	-0.04	-0.13
24r. is emotionally stable, not easily upset	2.36	1.24	0.51	-0.12	-0.23	-0.19	-0.01
34r. remains calm in tense situations	2.24	1.13	0.56	-0.03	-0.25	-0.22	-0.09
4. is depressed, blue	2.12	1.22	0.49	-0.32	0.04	-0.07	-0.14
29. can be moody	3.17	1.21	0.40	-0.01	0.09	0.01	-0.50
1. is talkative	4.07	1.18	0.03	0.68	0.12	0.10	-0.04
21r. tends to be quiet	3.27	1.29	-0.14	0.71	-0.07	-0.04	-0.04
16. generates a lot of enthusiasm	4.00	0.93	-0.07	0.57	0.38	0.28	0.10
36. is outgoing, sociable	4.06	0.92	0.05	0.58	0.19	0.13	0.33
6r. is reserved	3.24	1.28	-0.26	0.64	-0.12	0.00	0.06
31r. is sometimes shy, inhibited	3.03	1.28	-0.36	0.55	-0.17	0.05	0.01
11. is full of energy	4.05	1.01	-0.29	0.43	0.25	0.34	0.00
26. has an assertive personality	4.09	1.00	-0.29	0.39	0.18	0.31	-0.21
40. likes to reflect, play with ideas	4.03	0.96	-0.04	0.07	0.56	0.33	-0.04
25. is inventive	3.89	0.98	-0.21	0.22	0.56	0.29	-0.06
30. values artistic, aesthetic experiences	3.40	1.26	0.01	-0.12	0.67	0.02	0.13
5. is original, creative	3.76	1.05	-0.10	0.22	0.52	0.22	-0.08
15. is ingenious, a deep thinker	3.95	0.98	0.09	0.02	0.47	0.41	-0.12
20. has an active imagination	3.52	1.23	0.02	0.22	0.56	-0.07	-0.10
10. is curious about many different things	4.37	0.85	-0.14	0.28	0.50	0.23	-0.07
44. is sophisticated in art, music, or literature	2.81	1.32	-0.03	-0.09	0.58	-0.09	0.10
41r. has few artistic interests	3.11	1.44	-0.11	-0.14	0.51	-0.07	0.16
35r. prefers work that is routine	2.97	1.32	-0.23	0.00	0.22	-0.06	-0.10
3. does a thorough job	4.30	0.94	-0.01	0.08	0.12	0.63	-0.05
28. perseveres until the task is finished	4.28	0.94	-0.05	0.08	0.09	0.69	0.07
18r. tends to be disorganised	3.50	1.31	-0.07	-0.08	-0.36	0.53	0.23
23r. tends to be lazy	4.02	1.15	-0.09	0.01	-0.24	0.55	0.27
13. is a reliable worker	4.43	0.85	0.03	0.15	0.14	0.58	0.08
33. does things efficiently	4.26	0.83	-0.08	0.12	0.14	0.68	0.10
38. makes plans and follows through with them	4.07	0.94	-0.12	0.24	0.23	0.60	-0.02
43r. is easily distracted	3.19	1.24	-0.35	-0.08	-0.21	0.45	0.20
8r. can be somewhat careless	2.82	1.29	-0.01	-0.14	-0.36	0.42	0.25
32. is considerate and kind to almost everyone	4.19	0.85	0.15	0.26	0.22	0.21	0.50
17. has a forgiving nature	4.19	0.99	0.06	0.20	0.25	0.09	0.45
7. is helpful and unselfish with others	4.27	0.96	0.11	0.11	0.20	0.20	0.26
12r. starts quarrels with others	4.23	0.99	-0.26	-0.09	-0.04	0.10	0.55
37r. is sometimes rude to others	3.61	1.24	-0.15	-0.12	-0.08	0.12	0.68
27r. can be cold and aloof	3.65	1.26	-0.05	0.27	-0.08	0.03	0.58
22. is generally trusting	3.73	1.12	0.01	0.22	0.24	-0.06	0.34
2r. tends to find fault with others	3.12	1.24	-0.17	-0.17	-0.16	0.04	0.56
42. likes to cooperate with others	4.27	0.92	0.01	0.39	0.17	0.20	0.25

N = Neuroticism, E = Extraversion, O = Openness to Experience, C = Conscientiousness, A = Agreeableness

Items are ordered as in Appendix of Denissen, Geenen, van Aken, Gosling, & Potter, 2008.

Negatively framed items were reversed before analysis.

Table 4: Mean factor loadings for PCA with oblimin rotation.

	N	E	O	C	A	Mean
Average primary loadings	0.59	0.56	0.50	0.57	0.46	0.53
Average cross-loadings	0.11	0.12	0.15	0.13	0.11	0.12

Specifically for the BFI, Soto, John, Gosling and Potter (2008) devised 16 pairs of opposite items based on item content and the size of the interitem correlations to correct the value of each individual response (for 44 items) with the average and standard deviation of these pairs for that specific individual. In practice, this means that each individual response is standardised by dividing the difference of the result and individual mean by the individual's standard deviation for the selected 32 items.

$$Item_i = \frac{Item_i - mean(BFI32)}{std(BFI32)}$$

We performed factor analysis with oblimin rotation on the ipsatised dataset after calculating the acquiescence score for each individual. This model explains only 38.4% of the variance. Looking at the mean factor loadings and cross-loadings, the difference between 'normal' and ipsatised data seems relatively small (Table 6). However, when looking at the factor loadings per item (Table 7) we can observe several differences. While factors 'Neuroticism' and 'Openness to Experience' seem only slightly worse than for the original data, even here we can see cross-loadings over 0.3 (item 34r (Co) and item 15 (Co), respectively). For the three other factors there are several problematic items where the cross-loadings are higher than the expected primary loadings (3 items for Extraversion and 4 each for the other two factors). There are 2 factors with two cross loadings higher than the primary factor (item 18: Op and Ag instead of Co and item 27: Ne and Co instead of Ag), and three items (2r, 12r, 37r) with three cross-loadings higher than the primary loading, all three items have above 0.3 loadings for Neuroticism and Extraversion and a lower loading on Conscientiousness while the primary loadings are as low as 0.13, 0.09 and 0.10, respectively.

6 Multiple imputation

As described in Section 3.1, 747 participants were omitted from the analysis due to missing items. Considering listwise deletion is not unbiased unless the missingness is missing completely at random (MCAR) in which case the estimates will be less precise but unbiased or only a small fraction of the data is affected, the results could be biased (Graham & Hofer, 2000 cited by). It is notable that with 10% of missing data the multiple imputation of Likert-type data yields acceptable estimates with both MCAR and MAR (missing at random) data independently from eventual inter-item correlations and violations of normality (Leite & Beretvas, 2010). A possible solution, providing valid results under less strict assumptions about the nature of missingness is multiple imputation (MI).

Multiple imputation (MI) is a procedure that allows a practical solution to problems caused by the incompleteness of the data by imputing the missing values multiple times. Multiply imputing a dataset with missing values k times results in k different, complete datasets. The analysis proceeds

Table 5: *Factor loadings for PCA with oblimin rotation.*

Questionnaire Item	Mean	SD	N	E	O	C	A
19. worries a lot	3.30	1.27	0.72	-0.06	0.05	0.08	-0.02
14. can be tense	3.56	1.15	0.70	-0.04	0.07	0.15	-0.15
9r. is relaxed, handles stress well	2.36	1.24	0.62	-0.09	-0.20	-0.12	0.04
39. gets nervous easily	2.99	1.27	0.75	0.00	-0.02	0.01	-0.10
24r. is emotionally stable, not easily upset	2.36	1.24	0.50	-0.07	-0.19	-0.17	0.03
34r. remains calm in tense situations	2.24	1.13	0.55	0.04	-0.23	-0.20	-0.05
4. is depressed, blue	2.12	1.22	0.47	-0.30	0.09	-0.02	-0.11
29. can be moody	3.17	1.21	0.38	0.01	0.06	0.06	-0.49
1. is talkative	4.07	1.18	0.08	0.69	0.01	0.05	-0.05
21r. tends to be quiet	3.27	1.29	-0.10	0.74	-0.17	-0.10	-0.04
16. generates a lot of enthusiasm	4.00	0.93	-0.01	0.53	0.28	0.24	0.07
36. is outgoing, sociable	4.06	0.92	0.11	0.57	0.12	0.07	0.32
6r. is reserved	3.24	1.28	-0.21	0.66	-0.21	-0.06	0.04
31r. is sometimes shy, inhibited	3.03	1.28	-0.32	0.55	-0.25	0.00	-0.02
11. is full of energy	4.05	1.01	-0.24	0.38	0.15	0.31	-0.04
26. has an assertive personality	4.09	1.00	-0.25	0.36	0.09	0.30	-0.26
40. likes to reflect, play with ideas	4.03	0.96	-0.02	0.01	0.52	0.34	-0.07
25. is inventive	3.89	0.98	-0.18	0.16	0.50	0.28	-0.10
30. values artistic, aesthetic experiences	3.40	1.26	0.01	-0.17	0.70	0.02	0.14
5. is original, comes up with new ideas	3.76	1.05	-0.07	0.18	0.47	0.21	-0.10
15. is ingenious, a deep thinker	3.95	0.98	0.11	-0.03	0.42	0.43	-0.15
20. has an active imagination	3.52	1.23	0.03	0.21	0.54	-0.08	-0.09
10. is curious about many different things	4.37	0.85	-0.11	0.24	0.44	0.21	-0.09
44. is sophisticated in art, music, or literature	2.81	1.32	-0.03	-0.12	0.61	-0.10	0.11
41r. has few artistic interests	3.11	1.44	-0.12	-0.17	0.55	-0.07	0.16
35r. prefers work that is routine	2.97	1.32	-0.24	-0.02	0.22	-0.06	-0.10
3. does a thorough job	4.30	0.94	0.03	0.02	0.04	0.65	-0.10
28. perseveres until the task is finished	4.28	0.94	-0.01	0.01	0.01	0.70	0.01
18r. tends to be disorganized	3.50	1.31	-0.04	-0.12	-0.39	0.53	0.19
23r. tends to be lazy	4.02	1.15	-0.05	-0.04	-0.29	0.54	0.22
13. is a reliable worker	4.43	0.85	0.08	0.09	0.06	0.59	0.03
33. does things efficiently	4.26	0.83	-0.03	0.04	0.05	0.69	0.04
38. makes plans and follows through with them	4.07	0.94	-0.07	0.17	0.13	0.59	-0.08
43r. is easily distracted	3.19	1.24	-0.33	-0.14	-0.23	0.45	0.15
8r. can be somewhat careless	2.82	1.29	0.01	-0.17	-0.38	0.43	0.21
32. is considerate and kind to almost everyone	4.19	0.85	0.20	0.23	0.19	0.17	0.49
17. has a forgiving nature	4.19	0.99	0.10	0.18	0.23	0.05	0.45
7. is helpful and unselfish with others	4.27	0.96	0.14	0.09	0.18	0.19	0.25
12r. starts quarrels with others	4.23	0.99	-0.25	-0.13	-0.01	0.06	0.54
37r. is sometimes rude to others	3.61	1.24	-0.12	-0.15	-0.04	0.08	0.68
27r. can be cold and aloof	3.65	1.26	-0.01	0.26	-0.09	-0.04	0.59
22. is generally trusting	3.73	1.12	0.04	0.22	0.24	-0.11	0.35
2r. tends to find fault with others	3.12	1.24	-0.17	-0.19	-0.11	0.00	0.55
42. likes to cooperate with others	4.27	0.92	0.06	0.37	0.10	0.16	0.23

N = Neuroticism, E = Extraversion, O = Openness to Experience, C = Conscientiousness, A = Agreeableness

Items are ordered as in Appendix of Denissen, Geenen, van Aken, Gosling, & Potter, 2008.

Negatively framed items were reversed before analysis.

Table 6: *Mean factor loadings and cross-loadings (ipsatized scores).*

	N	E	O	C	A	Mean
Average primary loadings	0.61	0.49	0.53	0.41	0.32	0.47
Average cross-loadings	0.07	0.08	0.02	0.13	0.13	0.09

in two steps. First, each completed data set is analyzed. Second, the results are combined. Several statistical software packages, including SAS, have implemented MI procedures.

Given that the imputed datasets are complete, listwise deletion is not necessary any more which means we can use the data of all 7533 participants in the factor analysis.

A general assumption of multiple imputation is that the data are MAR or MCAR the completeness for each BFI item ranges between 98,3% and 99,8% and items are more likely to be absent if the item belongs to the factor 'Openness to Experience' (Op). We assume the data are (at least) MAR.

We performed multiple imputation on the family level assuming the variables were normally distributed. We obtained five complete datasets where missing values were imputed based on the means and variances of, and correlations between the variables. The imputed datasets were slightly different. We set the limits were set to 1 and 5 with imputed values rounded to the nearest multiple of 0.01. Thus, all imputed values were between the range available to the subjects to choose from when they participated in the survey (we also checked if this caused a boundary problem. It did not). Imputing on the family level meant that we also imputed the values for the persons who did not participate in the study. After the imputation we removed these people from the analysis (the results we obtained this way allowed to consider eventual family effects. However, results obtained after performing multiple imputation on individual level were only slightly different from the presented numbers.

After performing the principal components analysis on each completed dataset the results of the 5 analyses were only minimally different (average difference 0.01). To combine the results of the five PCA's we took the unweighted average of every factor loading. They explained 41.8% of the variance in all five analyses. Table 9 presents the results of the PCA with oblimin rotation together with the means and standard deviations per item (also the mean of the 5 datasets).

In Table 8 are the average factor loading and cross-loadings for each of the five factors. The biggest difference is at the Openness factor (0.47 compared to 0.50) where the most missing responses were observed, but overall the mean factor loadings and cross-loading show no notable difference from the previous analyses.

7 Discussion and concluding Remarks

In this study, we evaluated the psychometric properties of the Dutch version of the Big Five Inventory on a Flemish sample and compared the factor analysis results of a dataset from the Divorce in Flanders project with missing data with the multiply imputed version of the same dataset.

We evaluated the quality of the data by comparing the results to the validated Dutch translation of

Table 7: *Factor loadings for PCA with oblimin rotation for the ipsatized scores.*

Questionnaire Item	Mean	SD	N	E	O	C	A
19. worries a lot	-0.18	0.97	0.67	-0.17	-0.07	-0.06	0.05
14. can be tense	0.02	0.90	0.71	-0.10	-0.02	0.06	0.01
9r. is relaxed, handles stress well"	-0.92	0.95	0.59	-0.14	-0.15	-0.29	-0.19
39. gets nervous easily	-0.44	0.98	0.73	-0.08	-0.11	-0.06	0.06
24r. is emotionally stable, not easily upset"	-0.95	0.95	0.49	-0.10	-0.12	-0.29	-0.13
34r. remains calm in tense situations	-1.03	0.86	0.57	0.02	-0.15	-0.30	-0.15
4. is depressed, blue"	-1.17	1.04	0.52	-0.27	0.07	0.01	0.03
29. can be moody	-0.30	0.99	0.60	0.11	0.05	0.16	0.02
1. is talkative	0.43	0.87	0.03	0.63	-0.03	-0.07	0.06
21r. tends to be quiet	-0.22	0.94	-0.14	0.74	-0.07	-0.22	-0.08
16. generates a lot of enthusiasm	0.33	0.65	-0.10	0.37	0.17	0.18	0.37
36. is outgoing, sociable"	0.39	0.64	-0.07	0.28	-0.10	-0.07	0.52
6r. is reserved	-0.26	0.94	-0.28	0.63	-0.10	-0.18	-0.06
31r. is sometimes shy, inhibited"	-0.42	0.95	-0.35	0.58	-0.08	-0.10	-0.18
11. is full of energy	0.38	0.78	-0.21	0.29	0.10	0.35	0.26
26. has an assertive personality	0.42	0.79	-0.08	0.38	0.12	0.33	0.04
40. likes to reflect, play with ideas"	0.36	0.73	0.04	-0.04	0.54	0.26	0.02
25. is inventive	0.25	0.74	-0.10	0.11	0.51	0.28	0.11
30. values artistic, aesthetic experiences"	-0.12	0.94	-0.04	-0.23	0.73	-0.09	-0.05
5. is original, comes up with new ideas"	0.15	0.79	-0.01	0.14	0.50	0.16	0.05
15. is ingenious, a deep thinker"	0.31	0.75	0.18	-0.06	0.40	0.35	-0.02
20. has an active imagination	-0.04	0.99	0.12	0.14	0.49	0.00	0.20
10. is curious about many different things	0.65	0.67	0.02	0.14	0.41	0.21	0.18
44. is sophisticated in art, music, or literature"	-0.61	1.06	-0.04	-0.15	0.66	-0.13	-0.02
41r. has few artistic interests	-0.33	1.08	-0.17	-0.19	0.69	-0.23	-0.23
35r. prefers work that is routine	-0.44	1.01	-0.10	0.08	0.40	-0.09	-0.22
3. does a thorough job	0.60	0.68	0.05	-0.03	0.00	0.53	-0.16
28. perseveres until the task is finished	0.57	0.65	-0.07	-0.11	-0.08	0.60	-0.04
18r. tends to be disorganized	-0.04	0.95	-0.21	-0.19	-0.35	0.26	-0.36
23r. tends to be lazy	0.37	0.82	-0.24	-0.15	-0.30	0.26	-0.27
13. is a reliable worker	0.71	0.60	0.06	-0.06	-0.06	0.48	0.04
33. does things efficiently	0.55	0.56	-0.09	-0.11	-0.07	0.61	0.05
38. makes plans and follows through with them	0.40	0.68	-0.08	0.09	0.08	0.56	0.04
43r. is easily distracted	-0.30	0.91	-0.42	-0.15	-0.10	0.26	-0.30
8r. can be somewhat careless	-0.59	0.95	-0.16	-0.23	-0.30	0.17	-0.32
32. is considerate and kind to almost everyone	0.48	0.59	-0.03	-0.18	-0.10	-0.02	0.59
17. has a forgiving nature	0.50	0.73	-0.09	-0.16	0.01	-0.08	0.50
7. is helpful and unselfish with others	0.58	0.70	0.07	-0.16	0.00	0.08	0.31
12r. starts quarrels with others	0.54	0.77	-0.38	-0.36	-0.01	-0.17	0.10
37r. is sometimes rude to others	0.05	0.90	-0.44	-0.45	-0.07	-0.20	0.09
27r. can be cold and aloof	0.09	0.89	-0.35	-0.04	-0.16	-0.31	0.16
22. is generally trusting	0.12	0.90	-0.06	-0.08	0.06	-0.05	0.52
2r. tends to find fault with others	-0.36	0.96	-0.34	-0.40	-0.11	-0.17	0.14
42. likes to cooperate with others	0.56	0.67	-0.04	0.11	-0.09	0.10	0.44

N = Neuroticism, E = Extraversion, O = Openness to Experience, C = Conscientiousness, A = Agreeableness

Items are ordered as in Appendix of Denissen, Geenen, van Aken, Gosling, & Potter, 2008.

Negatively framed items were reversed before analysis.

Table 8: Mean factor loadings for PCA after multiple imputation.

	N	E	O	C	A	Mean
Average primary loadings	0.58	0.55	0.47	0.56	0.46	0.52
Average cross-loadings	0.11	0.12	0.15	0.14	0.11	0.13

the BFI which supported the imposed five-factor structure. Although the factor loadings are lower than for the Dutch validation sample and there were some cross-loadings on the less stable factors, the results of the factor analysis are similar to the Dutch, German, English and French samples supporting the findings of Denissen, Geenen, van Aken, Gosling, & Potter (2008).

In essence, all four factor analyses lead to the same factor structure as stated by the Big Five theory. In general, they explain about 41-42 percent of the variance which is an unremarkable result.

We found that the multiple imputation method works well in this context. The imputation allowed us to use the entire sample instead of a heavily reduced one in the factor analysis and by calculating the imputations per family we also took into account the possible similarities between the family members. The results of the factor analysis and the factor structure are basically the same for the imputed and the original data set. This should not be interpreted as evidence that multiple imputation is not of any use. Indeed, case-wise deletion excludes participants who forget to complete all items, which obviously may interfere with several facets of the factor 'Conscientiousness' (e.g., Order, Dutifulness, and Self-Discipline, John & Strivastata, 1999).

An issue to be addressed is that of clustering. The 7533 subjects (full dataset) are from 4457 families implying that subjects participating in the study cannot be assumed to be independent of one another. We chose to neglect this clustering for several reasons. First of all, to make our results easily comparable to other BFI reports we wanted to keep the analysis as close to the usual methods as possible. Second, we saw from the correlation structure of the 15 factor*role variables that the correlations between family members are negligible compared to the correlations observed within participants. Finally, the extension of factor analysis to multilevel data (Longford & Muthén, 1992) is computationally very demanding, in particular for large data sets with small clusters (families) and many items per subject. Development of fast and efficient numerical algorithms is needed and will be addressed in future research.

References

- Asendorpf, J. B., & van Aken, M. A. G. (2003). Validity of Big Five personality judgments in childhood: A 9-year longitudinal study. *European Journal of Personality*, 17, 117.
- Benet-Martinez, V., & John, O. P. (1998). Los cinco grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75, 729-750.
- Costa, P. T., & 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.

Table 9: *Factor analysis after multiple imputation (average of 5 imputed datasets)* .

Questionnaire Item	Mean	SD	N	E	O	C	A
19. worries a lot	3.30	1.28	0.71	-0.07	0.06	0.08	-0.02
14. can be tense	3.56	1.16	0.70	-0.04	0.06	0.16	-0.15
9r. is relaxed, handles stress well"	2.33	1.25	0.61	-0.09	-0.20	-0.13	0.03
39. gets nervous easily	2.95	1.28	0.74	0.00	-0.02	0.01	-0.11
24r. is emotionally stable, not easily upset"	2.33	1.24	0.49	-0.07	-0.19	-0.19	0.04
34r. remains calm in tense situations	2.20	1.12	0.54	0.04	-0.22	-0.22	-0.06
4. is depressed, blue"	2.12	1.24	0.46	-0.31	0.08	-0.01	-0.12
29. can be moody	3.07	1.23	0.38	0.00	0.07	0.08	-0.48
1. is talkative	4.08	1.19	0.07	0.68	0.01	0.07	-0.05
21r. tends to be quiet	3.26	1.30	-0.09	0.74	-0.17	-0.10	-0.04
16. generates a lot of enthusiasm	4.00	0.94	-0.02	0.52	0.27	0.25	0.07
36. is outgoing, sociable"	4.07	0.93	0.13	0.56	0.13	0.08	0.32
6r. is reserved	3.24	1.29	-0.21	0.65	-0.20	-0.07	0.05
31r. is sometimes shy, inhibited"	3.08	1.29	-0.32	0.56	-0.26	-0.01	-0.03
11. is full of energy	4.04	1.02	-0.25	0.37	0.15	0.31	-0.05
26. has an assertive personality	4.07	1.02	-0.24	0.34	0.07	0.33	-0.25
40. likes to reflect, play with ideas"	4.04	0.97	-0.03	0.01	0.48	0.37	-0.07
25. is inventive	3.88	1.00	-0.19	0.15	0.47	0.32	-0.11
30. values artistic, aesthetic experiences"	3.42	1.26	0.01	-0.17	0.67	0.06	0.14
5. is original, comes up with new ideas"	3.75	1.06	-0.08	0.17	0.45	0.24	-0.11
15. is ingenious, a deep thinker"	3.96	0.99	0.11	-0.04	0.38	0.46	-0.14
20. has an active imagination	3.43	1.23	0.02	0.19	0.54	-0.05	-0.10
10. is curious about many different things	4.35	0.87	-0.11	0.23	0.42	0.25	-0.09
44. is sophisticated in art, music, or literature"	2.79	1.31	-0.04	-0.11	0.60	-0.08	0.10
41r. has few artistic interests	3.13	1.43	-0.12	-0.17	0.52	-0.06	0.16
35r. prefers work that is routine	2.93	1.33	-0.25	-0.01	0.20	-0.06	-0.10
3. does a thorough job	4.34	0.93	0.01	0.02	-0.01	0.65	-0.10
28. perseveres until the task is finished	4.36	0.89	0.00	-0.01	-0.03	0.71	0.02
18r. tends to be disorganized	3.63	1.28	-0.04	-0.11	-0.44	0.50	0.19
23r. tends to be lazy	4.18	1.05	-0.06	-0.04	-0.33	0.52	0.23
13. is a reliable worker	4.46	0.86	0.07	0.08	0.02	0.59	0.04
33. does things efficiently	4.32	0.82	-0.02	0.03	0.03	0.69	0.06
38. makes plans and follows through with them	4.11	0.94	-0.07	0.16	0.10	0.61	-0.07
43r. is easily distracted	3.32	1.21	-0.34	-0.14	-0.26	0.42	0.16
8r. can be somewhat careless	2.89	1.31	0.01	-0.15	-0.42	0.39	0.22
32. is considerate and kind to almost everyone	4.20	0.86	0.21	0.22	0.20	0.18	0.49
17. has a forgiving nature	4.23	0.98	0.10	0.17	0.24	0.07	0.45
7. is helpful and unselfish with others	4.30	0.96	0.14	0.07	0.18	0.20	0.23
12r. starts quarrels with others	4.32	0.96	-0.25	-0.13	-0.01	0.04	0.54
37r. is sometimes rude to others	3.75	1.21	-0.13	-0.15	-0.04	0.06	0.68
27r. can be cold and aloof	3.73	1.25	0.00	0.27	-0.08	-0.05	0.58
22. is generally trusting	3.73	1.12	0.04	0.22	0.25	-0.09	0.35
2r. tends to find fault with others	3.23	1.24	-0.16	-0.18	-0.11	-0.01	0.55
42. likes to cooperate with others	4.29	0.92	0.06	0.36	0.10	0.17	0.24

N = Neuroticism, E = Extraversion, O = Openness to Experience, C = Conscientiousness, A = Agreeableness

Items are ordered as in Appendix of Denissen, Geenen, van Aken, Gosling, & Potter, 2008.

Negatively framed items were reversed before analysis.

- 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.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417-440.
- Fieuws, S. & Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, 62: 424-431.
- Fieuws, S., Verbeke, G., Maes, B., et al. (2008). Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics*, 9: 419-431.
- Goldberg, L. R. (1990). An Alternative "Description of Personality": The Big-Five Factor Structure. *Journal of Personality and Social Psychology*, 59(6):1216-1229.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59(2), 93.
- Graham, J. W., & Hofer, S. M. (2000). Multiple imputation in multivariate research. In: T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, applied approaches, and specific examples*, p. 201-218. Hillsdale, NJ: Erlbaum.
- Hofstee, W. K. B., Kiers, H. A. L., de Raad, B., Goldberg, L. R., & Ostendorf, F. (1997). Comparison of Big-Five structures of personality traits in Dutch, English, and German. *European Journal of Personality*, 11, 15-31.
- John, O. P., Naumann, L. P., & Soto, C. C. (2008). Paradigm Shift to the Integrative Big Five Trait Taxonomy. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 114-158). New York: Guilford.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102-138). New York: Guilford.
- Laird, N. & Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics*, 38: 963-974.
- Lang, F. R., Lüdtke, O., & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica*, 47(3), 111-121.
- Leite, W., & Beretvas, S. N. (2010). The performance of multiple imputation for Likert-type items with missing data. *Journal of Modern Applied Statistical Methods*, 9(1), 8.
- Longford, N. T., & Muthén, B. O. (1992). Factor analysis for clustered observations. *Psychometrika*, 57(4), 581-597.
- McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52, 81-90.
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, 60(2), 175-215.
- Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York:

Springer.

Plaisant, O., Courtois, R., Réveillère, C., Mendelsohn, G. A., & John, O. P. (2010). Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R. In *Annales Médico-psychologiques, revue psychiatrique*, 168(2), 97-106

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology*, 94(4), 718.

Szirmák, Z., & De Raad, B. (1994). Taxonomy and structure of Hungarian personality traits. *European Journal of Personality*, 8(2), 95-117.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate analysis*. California State University Northridge: Harper Collins College Publishers.

Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of vectors of the mind*. Chicago: University of Chicago.

Widiger, T. A., & Trull, T. J. (1997). Assessment of the five-factor model of personality. *Journal of Personality Assessment*, 68(2), 228-250.