ROCZNIKI PSYCHOLOGICZNE / ANNALS OF PSYCHOLOGY
2014, XVII, 2, 385-402

EWA TOPOLEWSKA[a]
EWA SKIMINA[a]
WŁODZIMIERZ STRUS[a]
JAN CIECIUCH[bc]
TOMASZ ROWIŃSKI[a]

[a]Cardinal Stefan Wyszyński University in Warsaw
[b]The University of Finance and Management in Warsaw
[c]Universität Zürich

THE SHORT IPIP-BFM-20 QUESTIONNAIRE FOR MEASURING THE BIG FIVE

The most frequently used measures of personality consist of a large number of items. However, it is their short versions that have enjoyed popularity in recent years. This article is a presentation of the short form of Goldberg's questionnaire for measuring the Big Five personality traits. The questionnaire measures five traits (Extraversion, Agreeableness, Conscientiousness, Emotional Stability, Intellect), and consists of 20 items. It is a shortened version of the 50-item Big Five Markers questionnaire from the resources of the International Personality Item Pool, whose Polish version was prepared by Strus, Cieciuch, and Rowiński (2014b). In constructing the short version, we followed the procedure developed by Donnellan and colleagues (2006), aimed at maximizing the internal consistency and independence of scales. The research was carried out on a group of $N = 903$ people aged between 16 and 83 years. The validity (verified in confirmatory factor analysis and in the analysis of correlations between the questionnaire's scales and other measures of the five personality traits) and reliability (measured by Cronbach's alpha coefficient) are satisfactory and make the questionnaire fit for use in scientific research.

Keywords: the Big Five, short version of the questionnaire, personality, personality traits, International Personality Item Pool.

Corresponding author: EWA TOPOLEWSKA – Cardinal Stefan Wyszyński University in Warsaw, ul. Wóycickiego 1/3 budynek 14, 01-938 Warszawa; e-mail: topolewska.ewa@gmail.com

Włodzimierz Strus's research work was financed with the funds of the Polish National Science Center, granted by decision no. 2012/07/B/HS6/01374.

Jan Cieciuch's research work was financed with the funds of the Polish National Science Center, granted by decision no. DEC-2011/01/D/HS6/04077.

Short Measures of Personality Traits

In recent years, there has been a growing interest in short versions of questionnaires for measuring personality (Baldasaro, Shanahan, & Bauer, 2013; McCrae & Costa, 2007; Thalmayer, Saucier, & Eigenhuis, 2011). Several shortened measures of personality published in English and German, showing satisfactory psychometric properties, have already gained popularity among researchers. These are, for example: *the Questionnaire Big Six Scale* (24QB6; Thalmayer et al., 2011), being a 24-item questionnaire measuring six personality traits, as well as 10-item questionnaires measuring five personality traits: *the 10-Item Big Five Inventory* (BFI-10; Rammstedt, 2007); *Ten-Item Personality Inventory* (TIPI; Gosling, Rentfrow, & Swann, 2003); *Mini-Markers* (Saucier, 1994), being a 40-adjective version of Goldberg's (1992) list of the best lexical markers of the Big Five, originally comprising 100 items (*Big Five Factor Markers*); or the 20-item Mini-IPIP inventory, being a short version of the 50-item questionnaire for measuring the Big Five personality traits (Donnellan, Oswald, Baird, & Lucas, 2006) from the resources of the *International Personality Item Pool* (IPIP).

The advantages of shortened versions of questionnaires include low cost and the short time that it takes to fill them in (Herzberg & Brähler, 2006). This makes it possible to include personality measurement in studies whose time is limited, such as those conducted online (Gosling et al., 2003), or in studies with repeated measurement (e.g., longitudinal studies). Shortened versions are much less tiresome for respondents and therefore reduce the risk of errors resulting from accidental indication of answers (Thalmayer et al., 2011). Their application is also a good solution in the case of respondents who may have difficulties in reading (McCrae & Costa, 2007).

Reise and Henson (2007) demonstrated that most of the variability in traits measured by the NEO-PI-R questionnaire on 8-item scales can be measured at a comparable quality level using 4-item scales. They used the CAT procedure (computerized adaptive testing). In this procedure, in order to maximize the precision of measurement, the computer systematically selects test items for each particular person based on information following from the person's earlier responses (Weiss, 2004).

Nevertheless, the construction and use of short versions of questionnaires involves certain risks. Reducing the number of items may result in lower reliability of scales (McCrae & Costa, 2007) or increase the risk of error in conclusions concerning the relationships between personality traits and other constructs measured. The items selected for the shortened version of the questionnaire,

measuring personality traits, usually represent them to a limited extent, which may make it difficult to demonstrate the relationships between the traits and other variables (Credé, Harms, Niehorster, & Gaye-Valentine, 2012). For example, in the model proposed by McCrae and Costa (2007) each of the five basic personality traits consists of six facets, and therefore a scale consisting of only three items cannot fully convey the specificity of a given trait. Using a scale thus shortened for explaining the variance of some other variable may lead to an underestimation of the prognostic validity of the personality trait that the scale measures. What is more, if some other construct is used for explaining part of the remaining variance of that variable, the predictive value of that construct may be overestimated. This would be a result of underestimating the initial variance explained by the personality trait measured using the short scale and of underestimating the correlation of that trait with this construct (Credé et al., 2012).

However, it is possible to reduce the above risks. Careful selection of items for short versions of scales makes it possible to achieve satisfactory reliability (Fischbach & Moosbrugger, 2007). Numerous studies have also confirmed that the use of an appropriate method of constructing short versions of questionnaires can result in the short versions having only slightly lower criterion validity than their full-length prototypes (e.g., Credé et al., 2012; Frazier, Naugle, & Haggerty, 2012; Thalmayer et al., 2011). A good solution is to shorten scales to four items. McCrae and Costa (2007) showed that the mean from this number of items reduces random error and systematic error and that using a balanced key, with half of the items reverse-coded, enables controlling the tendency to agree.

Selected Ways of Shortening Questionnaires

Various authors have used different techniques to select the optimal items from the basic version of a measure that would make up its shortened version. A few typical approaches found in the literature will be briefly described below.

When creating the NEO-PI-3 shortened questionnaire for investigating personality traits, McCrae and Costa (2007) used regression analysis. The dependent variable was the score on an 8-item scale and the independent variables were the items. The criterion of selecting items for the shortened version was the amount of variance explained by a given item in the regression analysis.

Batinic, Wolff, and Haupt (2007) used a different method when constructing a shortened version of the *Trendsetting Questionnaire* (TDS-K). The procedure had two stages. The first one was selecting those items from each scale whose

correlation with the overall score on that scale was higher than .5. In the second stage, the authors selected items with the highest face validity.

Another popular approach is based on factor analysis. It consists in selecting those items for the shortened version that have high factor loadings on the respective factors and low cross-loadings (Samson & Huber, 2010).

When creating shortened versions of questionnaires, some authors combine several criteria. For example, when shortening Eysenck's Personality Inventory, Fischbach and Moosbrugger (2007) sought to meet the criteria connected with both the orthogonality of factors and the reliability of scales.

Still another procedure was applied by Donnellan and colleagues (2006) in creating the 20-item version of the originally 50-item questionnaire measuring five personality traits. This procedure will be described in greater detail because we used it in our research.

The Procedure of Shortening the Questionnaire Proposed by Donnellan and Colleagues

The process through which Donnellan et al. (2006) selected the items to be included in the shortened version comprised several stages: (1) By means of exploratory factor analysis (EFA) with Varimax rotation, factor loadings were computed for each of the 50 items; also computed were their cross-loadings on the remaining factors. (2) Next, the mean of absolute values of all the cross-loadings was computed for each item. (3) The obtained mean was subtracted from the absolute value of factor loading of a given item on the respective factor. The discrimination score was thus obtained, being an indicator of the extent to which a given item was a good measure of a particular trait independently of the influence of other factors. This procedure was meant to guarantee the construction of scales that were both internally consistent and independent of one another. (4) Finally, two positively keyed items with the highest discrimination scores were selected for each scale, and two negatively keyed items with the highest discrimination scores were selected likewise, in accordance with the recommendations of Saucier and Goldberg (2002), so that the scales were balanced with regard to the number of positively and negatively keyed items.

Following the above procedure, the authors computed the discrimination score for each item. Item selection on its basis was impossible in a few cases because of the criterion described in point 4, according to which every shortened scale was to consist of two positively keyed and two negatively keyed items. The problem was that the Emotional Stability scale in the questionnaire being short-

ened had only two positively keyed items and the Intellect scale had only three negatively keyed ones (Donnellan et al., 2006).

Donnellan and colleagues (2006) verified the selected items in EFA procedure on a different sample. The hypothesized model was in principle confirmed, but a few items were found not to have acceptable factor loadings. For this reason, they were replaced with items that took the next places in the first EFA in terms of discrimination score (in the case of the Conscientiousness scale) or with items that better conveyed the theoretical meaning of the construct (in the case of the Intellect scale; Donnellan et al., 2006).

Thus, the procedure proposed by Donnellan et al. (2006) was not applied mechanically. It rather constituted a set of guidelines concerning item selection, but it was slightly modified where appropriate, if the modifications increased the chance of selecting such items that would make up a better-quality measure.

The 20-Item Questionnaire for Measuring the Big Five

The aim of the presented research was to develop a short (20-item) version of the Polish adaptation of the IPIP-BFM-50 questionnaire for measuring five personality traits (Strus, Cieciuch, & Rowiński, 2014b) using the procedure proposed by Donnellan and colleagues (2006).

The 50-item questionnaire, from IPIP resources, was named IPIP-FFM by Donnellan et al (2006), where FFM stands for *Five-Factor Model*. It is worth noting, however, that this is a questionnaire measuring the five basic personality traits as identified in the lexical approach (Goldberg, 1990, 1992). The differences between the lexical Big Five and the Five Factor Model derived from the psychometric approach are not fundamental and the two terms are often used interchangeably (John & Srivastava, 1999; De Raad & Perugini, 2002). Yet, from the point of view of IPIP resources, the difference is important, since those resources include questionnaires for measuring both the lexical Big Five and the traits from the Five-Factor Model of Personality (Saucier & Goldberg, 2002). The 50-item questionnaire, whose shortened version was developed by Donnellan and colleagues (2006), is the questionnaire equivalent of Big Five Factor Markers (BFM), Goldberg's (1992) list of 100 adjectives that turned out to be the best lexical markers of the Big Five in the English language. Because IPIP resources also include questionnaires for measuring five personality traits in the psychometric tradition (Five-Factor Model, FFM), we precisely distinguish between IPIP-BFM (the lexical approach) and IPIP-FFM (or IPIP-NEO, the questionnaire ap-

proach) measures. In the light of the above, the name used by Donnellan and colleagues is not accurate. For this reason, we will be abbreviating the questionnaire's name to IPIP-BFM-50 (cf. Strus et al., 2014b). Donnellan and colleagues (2006) refer to the 20-item version of the measure as Mini-IPIP. We will be referring to our 20-item version as the IPIP-BFM-20.

The use of the IPIP-BFM-50 in analyses is justified by several reasons. Firstly, the questionnaire serves to measure personality in the commonly used Big Five Model. Secondly, it is a Polish adaptation of the IPIP-BFM-50 questionnaire, which was also the point of departure for the shortened version developed by Donnellan and colleagues (2006). In this respect, our study is a replication of the research carried out by Donnellan et al. (2006). Previous analyses show that Donnellan's Mini-IPIP is nearly equal in psychometric terms to its 50-item prototype (Credé et al., 2012; Donnellan et al., 2006; Thalmayer et al., 2011). Thirdly, the IPIP-BFM-50 is available free of charge from the public domain at www.ipip.ori.org (Goldberg et al., 2006), whose Polish version – www.ipip.edu.pl – has also been developed (Strus, Cieciuch, & Rowiński, 2011, 2013). Fourthly, the Polish adaptation of the IPIP-BFM-50 is available as adapted by Strus et al. (2014b). Taking into account Thalmayer at al.'s (2011) argumentation, we decided to repeat the procedure used by Donnellan et al. (2006) rather than adopt their short version. The procedure of item selection applied by Donnellan et al. (2006) may bring somewhat different results with different data. This is of particular importance when data has been collected in a different country and by means of a different language version, as in our case. In Poland, the IPIP-BFM-50 questionnaire is an adaptation of that measure from the English language. It can therefore be expected that some items may have different properties in the analyses carried out on data from the Polish sample.

The Present Study

We applied Donnellan's procedure on Polish data collected using IPIP-BFM-50 (Strus et al., 2014b). However, we made a significant modification to the procedure by adding another stage, verifying the quality of measurement. We verified the effects of item selection in confirmatory factor analysis (CFA), carried out on a different sample. During item selection we also used Cronbach's alpha as an additional criterion.

Moreover, we verified the validity of the measure by analyzing its relationships with four other questionnaires for measuring the Big Five.

METHOD

Participants

The study was carried out using the paper-and-pencil method. The participants were 903 people aged from 16 to 83 years (M_{age} = 30.97; SD_{age} = 13.82); 55% of them were women. The participants lived in various regions of Poland, but the vast majority of them (73.1%) came from central Poland. A considerable percentage of the participants (43.3%) lived in cities with over 500,000 inhabitants, 9.7% lived in towns with a population between 100,000 and 500,000 people, and 29.7% lived in towns with less than 100,000 inhabitants; 17.2% of participants lived in the countryside. About 31% of participants had higher education and just as many were still university students. Incomplete higher education was declared by 6.5% of the participants, post-secondary education – by 2.5%, secondary – by 21%, vocational – by 3.8%, and elementary – by 3.6%. The research was conducted by trained students; each of them examined a few people. Participation was voluntary and anonymous.

Measure

The IPIP-BFM-50

In our study we used items included in the IPIP-BFM-50 (Goldberg, 1999; Goldberg et al., 2006) as adapted into Polish by Strus et al. (2014b), based on which we constructed the IPIP-BFM-20.

The IPIP-BFM-50 questionnaire comprises 50 items, 10 per each scale. Participants indicated their answers on a 5-point scale from 1 (*very inaccurate*) to 5 (*very accurate*). Both the IPIP-BFM-50 and the IPIP-BFM-20 measure five personality traits in the lexical tradition. Table 1 presents their brief descriptions as proposed by Strus et al. (2014b).

Table 1
Description of the Five IPIP-BFM-50 Scales (Strus et al., 2014b)

Scale	Object of measurement	Individuals who score high may be described as:	Individuals who score low may be described as:
Extraversion	The level of activity, energy, as well as sociability and social confidence (assertiveness).	active, energetic, extraverted, talkative, bold, and assertive.	introverted, reserved, quiet, and socially inhibited.
Agreeableness	Positive (vs. negative) attitude towards people.	trustful, kind, considerate and warm as well as cooperative and helpful.	distrustful, selfish, unkind, rude, and emotionally cold towards other people.
Conscientiousness	The level of organization, diligence in pursuing goals and performing tasks as well as proneness to order and dutifulness.	organized, diligent, thorough and efficient in what they do as well as systematic and dutiful.	unsystematic and inconsistent, unconcerned with order and planning, negligent, careless, and undependable.
Emotional Stability	The level of reactivity and emotional stability, emotional resistance and tolerance to frustration.	imperturbable, calm, relaxed, not prone to negative emotional states.	anxious, nervous, moody, prone to worry and oversensitive as well as envious, touchy, prone to anger and irritation.
Intellect	Intellectual openness, creativity, and imagination.	intellectually active and cognitively open, creative, introspective, having a vivid imagination and a wide range of interests.	unintellectual, noninquisitive, unimaginative, simple, unsophisticated, unreflective and uncreative.

Other questionnaires measuring the Big Five traits

Personality traits were also measured using four other questionnaires. These were: (1) the NEO-PI-R questionnaire by Costa and McCrae (1992) adapted into Polish by Siuta (2006); (2) the IPIP-NEO-PI-R questionnaire, measuring the same personality traits as the NEO-PI-R but derived from IPIP resources. The Polish version of the IPIP-NEO-PI-R was prepared by Rowiński, Strus, Cieciuch, and Wieman. (3) *Big Five Aspects Scales* (BFAS; DeYoung, Quilty, & Peterson, 2007), measuring five personality traits and their ten aspects. The Polish version of the questionnaire was prepared by Strus, Cieciuch, and Rowiński (2012). (4) The IPIP-45AB5C questionnaire, measuring 45 variables in the AB5C model (*Abridged Big Five Dimensional Circumplex*) developed by Hofstee, De Raad, and Goldberg (1992). This measure was adapted into Polish by Strus, Cie-

ciuch, and Rowiński (2014a). Measurements using the above questionnaires were carried out between two and six weeks after the measurement using the IPIP-BFM-50.

Analysis Plans

We carried out statistical analyses on two study groups, obtained in the cross--validation procedure, in which we divided the group of $N = 903$ participants randomly into two subgroups (Browne, 2000). We performed exploratory statistical analysis on one of the subgroups and confirmatory analysis on the second one.

Based on the results of EFA performed on the first group ($n = 467$, $M_{age} = 31.58$, $SD_{age} = 14.40$, women = 57%), we selected items to be included in the short version of the questionnaire (following the procedure described above) and computed the values of Cronbach's alpha for the short scales made up of the selected items. In the second group ($n = 436$, $M_{age} = 30.26$, $SD_{age} = 13.17$, women = 54%), we checked the psychometric properties of the 20-item questionnaire by performing CFA and reliability analysis.

Additionally, we performed a correlation analysis of many methods, entering the results of Big Five measurement using the IPIP-BFM-50, the IPIP-BFM-20, and four other questionnaires.

RESULTS

Exploratory Analyses

Analyses aimed at selecting items for the shortened version of the IPIP-BFM-50 questionnaire were performed on the first group ($n = 467$). We performed all the analyses presented below using SPSS 20. EFA results for the 50 items of IPIP-BFM-50 can be found in Table 2.

Table 2

Factor Loadings in EFA With Forced Five Factors After Varimax Rotation, the Mean Absolute Value of Cross-Loadings, and Discrimination Scores (the Difference Between the Absolute Value of the Factor Loading of a Given Item on the Appropriate Factor and the Mean Absolute Value of That Item's all Cross-Loadings) in Group 1

Number item	Item content	Factor					Mean cross-loading	Discrimination scores
		1	2	3	4	5		
39n*	Have frequent mood swings	**.73**	-.01	.03	.15	-.01	.05	.68
34n	Change my mood a lot	**.70**	.06	-.06	.12	-.01	.06	.64
14n*	Worry about things	**.70**	-.12	.08	-.04	.00	.06	.62
49n	Often feel blue	**.66**	-.21	-.10	.13	-.01	.11	.55
19n*	Seldom feel blue	**-.63**	.19	.05	-.05	.01	.08	.56
24n	Am easily disturbed	**.60**	-.11	.08	-.03	-.10	.08	.52
44n	Get irritated easily	**.58**	.03	-.12	.13	.06	.09	.49
29n	Get upset easily	**.57**	.04	-.13	.14	-.02	.09	.48
4n	Get stressed out easily	**.56**	-.19	.12	-.01	-.03	.09	.48
9n*	Am relaxed most of the time	**-.43**	.23	.08	.06	.20	.14	.28
45i	Spend time reflecting on things	.38	-.05	.23	.06	**.32**	.18	.14
46e*	Am quiet around strangers	.09	**-.77**	.00	.09	-.06	.06	.71
1e*	Am the life of the party	-.03	**.72**	.04	.05	.21	.08	.64
16e*	Keep in the background	.20	**-.70**	-.13	.07	-.09	.12	.57
11e	Feel comfortable around people	-.20	**.66**	.26	.01	.07	.13	.52
31e*	Talk to a lot of different people at parties	-.01	**.63**	.02	-.03	.19	.06	.57
6e	Don't talk a lot	.05	**-.63**	-.18	-.01	-.09	.08	.55
21e	Start conversations	-.09	**.62**	.19	-.03	.20	.13	.49
36e	Don't like to draw attention to myself	.04	**-.54**	.04	-.13	-.20	.10	.44
41e	Don't mind being the center of attention	-.17	**.48**	.09	.04	.28	.14	.34
2u*	Feel little concern for others	.03	-.11	**-.62**	.02	-.03	.05	.57
22a*	Am not interested in other people's problems	-.01	-.09	**-.59**	.09	-.04	.06	.53
32a	Am not really interested in others	.13	-.14	**-.58**	-.01	-.15	.11	.47
17a*	Sympathize with others' feelings	-.06	-.04	**.57**	-.07	-.05	.05	.52
7a	Am interested in people	.08	.22	**.57**	.02	.15	.12	.45
42a	Feel others' emotions	.22	.00	**.57**	-.03	.21	.12	.45
37a*	Take time out for others	.03	.09	**.52**	.01	.09	.05	.47
27a	Have a soft heart	-.05	.05	**.42**	-.14	.08	.08	.34
12a	Insult people	.26	.17	**-.40**	.24	.13	.20	.20
47a	Make people feel at ease	-.06	.20	**.31**	.03	.15	.11	.20
23c*	Get chores done right away	-.01	.06	-.02	**-.66**	-.07	.04	.62
8c*	Leave my belongings around	.04	.08	-.10	**.65**	.21	.11	.54
28c*	Often forget to put things back in their proper place	.07	.03	.00	**.61**	.15	.06	.54
43c*	Follow a schedule	-.03	.00	.04	**-.55**	.06	.03	.51
33c	Like order	.07	-.05	.17	**-.54**	-.05	.08	.45

Number item	Item content	Factor					Mean cross-loading	Discrimination scores
		1	2	3	4	5		
38c	Shirk my duties	.12	.04	-.05	**.53**	-.11	.08	.45
18c	Make a mess of things	.26	.00	.01	**.52**	-.05	.08	.44
48c	Am exacting in my work	-.06	.07	.07	**-.47**	.14	.08	.39
3c	Am always prepared	-.09	.03	-.06	**-.45**	.10	.07	.38
13c	Pay attention to details	.13	.02	.09	**-.29**	.26	.13	.17
30i*	Do not have a good imagination	-.09	-.13	-.30	-.01	**-.63**	.13	.50
50i*	Am full of ideas	-.07	.27	.13	.01	**.61**	.12	.49
25i	Have excellent ideas	-.18	.24	.10	.02	**.55**	.14	.41
15i	Have a vivid imagination	.22	.10	.22	.14	**.53**	.17	.36
5i*	Have a rich vocabulary	-.05	.13	-.01	-.14	**.48**	.09	.39
40i	Use difficult words	-.04	.13	-.07	.02	**.44**	.06	.37
20i	Am not interested in abstract ideas	-.08	-.04	-.06	-.09	**-.40**	.07	.33
35i	Am quick to understand things	-.17	.09	.08	-.14	**.38**	.12	.26
26e	Have little to say	.01	**-.26**	-.07	.03	-.33	.11	.22
10i*	Have difficulty understanding abstract ideas	.10	-.10	-.05	.17	**-.32**	.10	.21

Note. n – Emotional Stability items; e – Extraversion items; a – Agreeableness items; c – Conscientiousness items; i – Intellect items. * Items included in the final version of the IPIP-BFM-20.

According to the results obtained by applying Donnellan's procedure, the shortened version of the IPIP-BFM-50 should include the following items: 1, 16, 31, 46 (Extraversion); 2, 17, 22, 37 (Agreeableness); 8, 23, 28, 43 (Conscientiousness); 9, 19, 34, 39 (Emotional Stability); and 20, 25, 30, 50 (Intellect). However, after such selection very similar items are found within a factor. This is the case for the Emotional Stability and Intellect scales. In the case of Emotional Stability, the similar items are 34 ("Change my mood a lot") and 39 ("Have frequent mood swings"). As can be seen in Table 2, the discrimination score of Item 14 ("Worry about things") is lower than that of Item 34 by only 0.02. An analogous situation occurred in the case of the Intellect scale, the similar items being 50 ("Am full of ideas") and 25 ("Have excellent ideas"). As Table 2 shows, Item 5 ("I have a rich vocabulary") has a lower discrimination score that Item 25 by only 0.02. In order to semantically differentiate the items making up the shortened version of the IPIP-BFM-50, in the final version of this measure we replaced Item 34 with Item 14 in the Emotional Stability scale and Item 25 with Item 5 in the Intellect scale.

For scales thus constructed, we computed Cronbach's alpha coefficients. The items we entered in the analysis were those obtained after applying Donnellan's

procedure and after replacing two items for semantic reasons: 34 with 14 and 25 with 5. Cronbach's alpha coefficient for the 4-item Intellect scale was found to be too low (α = .57). According to Donnellan's procedure (Donnellan et al., 2006), in the shortened version of the questionnaire there should be two items measuring a given trait directly and two reverse-scored ones. That particular scale in the IPIP-BFM-50 includes three negatively keyed items: 30 ("Do not have a good imagination"), 10 („Have difficulty understanding abstract ideas"), and 20 („Am not interested in abstract ideas"). In order to obtain an acceptable value of Cronbach's alpha reliability coefficient for the Intellect scale and at the same time implement the recommendation to include two negatively keyed items in the shortened version of the measure, we replaced one of the negatively keyed Intellect items. Of the two items that negatively load the Intellect scale, selected in accordance with Donnellan's procedure, Item 20 has a lower discrimination score (.33) than Item 30 (.50). Based on the discrimination scores, we replaced Item 20 with Item 10. Thus constructed, the Intellect scale has a reliability of α = .61.

Eventually, the IPIP-BFM-20 comprises the following items: 1, 16, 31, 46 (Extraversion); 2, 17, 22, 37 (Agreeableness); 8, 23, 28, 43 (Conscientiousness); 9, 14, 19, 39 (Emotional Stability); and 5, 10, 30, 50 (Intellect).

Confirmatory Analyses

The items selected for the short version of the questionnaire were entered in CFA, performed in AMOS 20. The assessment of the model's goodness-of-fit to data was based on RMSEA, CFI, and SRMR indices. The threshold values of model acceptability were RMSEA and SRMR values below .08 and CFI value above .9 (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004).

The following statistical fit indices were obtained for the proposed model (Figure 1): CFI = .899, RMSEA = .05, SRMR = .06. However, according to the modification indices, it is possible to correlate one pair of errors in the model: Item 23 error with Item 43 error (both items being part of the Conscientiousness scale). After the correlation of these errors, the model has the following goodness-of-fit to data: CFI = .911, RMSEA = .05, SRMR = .06.

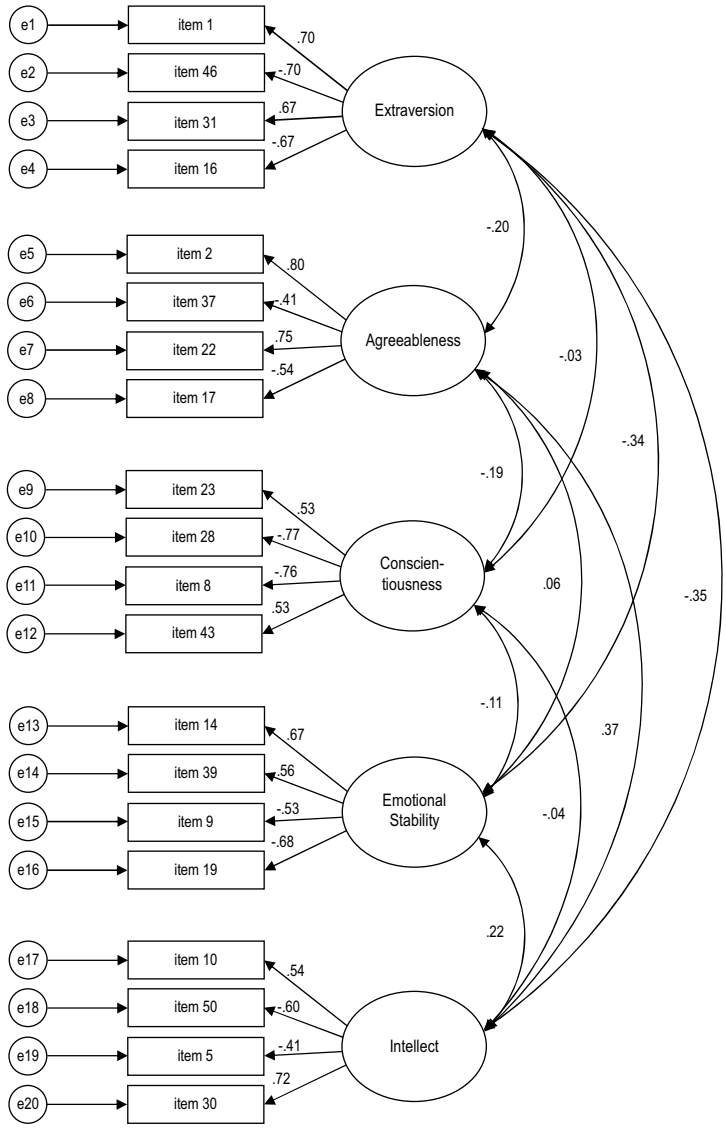

Figure 1. The five-factor model of the short questionnaire for measuring the Big Five in Group 2 (n = 436).

We checked the reliability of the final versions of the questionnaire's short scales using Cronbach's alpha. The values of that coefficient reached the acceptable level for all the scales. The results obtained are presented in Table 3.

Table 3
Cronbach's α Coefficients in Each Group

	IPIP-BFM-50 (50 items)	IPIP-BFM-20 (20 items)	
	The entire sample $n = 903$	Group 1 $n = 467$	Group 2 $n = 436$
Extraversion	.86	.82	.78
Agreeableness	.79	.69	.71
Conscientiousness	.76	.72	.75
Emotional Stability	.86	.73	.70
Intellect	.80	.61	.65

The IPIP-BFM-20 and Other Measures of Five Personality Traits

In order to check the validity of IPIP-BFM-20 scales, the scores obtained using this questionnaire were correlated with those obtained using other questionnaires for measuring five personality traits. Table 4 presents the correlation matrix of five personality traits measured using both the IPIP-BFM-20 and the IPIP-BFM-50 with four other personality measures.

Analyses of correlations confirmed the validity of the IPIP-BFM-20 questionnaire as regards the relations between its scores and those of other scales measuring the same (or very similar) traits. As Table 4 shows, the values of the correlation coefficient between the scales of IPIP-BFM-20 and the corresponding scales of other measures of five personality traits are higher than correlations with scales measuring other traits. The highest correlation coefficients were obtained between scales measuring Neuroticism or its opposite, Emotional Stability (.67 or higher). Coefficients of correlation between the corresponding scales of Extraversion, Conscientiousness, and Intellect (or Openness to Experience) were equal to or higher than .54. Correlations were the weakest between scales measuring Agreeableness, their coefficient values ranging from .37 to .74. A similar result was explained by Strus et al. (2014b) as an effect of differences in the definition of traits between the lexical and questionnaire approaches. Warmth, being a component of Agreeableness in the lexical tradition, falls into the scope of Extraversion in the questionnaire tradition. The definition of Agreeableness in the questionnaire approach encompasses more aspects of morality and humility than the corresponding definition in the lexical approach (cf. Ashton & Lee, 2005).

Table 4

Correlations (Pearson's r*) Between IPIP-BFM-20 and IPIP-BFM-50 Scales and NEO-PI-R (*n *= 883), IPIP-NEO-PI-R (*n *= 368), BFAS (*n *= 297), and IPIP-45AB5C Scales (*N *= 903) as Well as Intercorrelations Among IPIP-BFM-20 Scales (*N *= 903)*

		Emotional Stability		Extraversion		Conscientio-usness		Agreeableness		Intellect	
		IPIP-BFM-20	IPIP-BFM-50	IPIP-BFM-20	IPIP-BFM-50	IPIP-BFM-20	IPIP-BFM-50	IPIP-BFM-20	IPIP-BFM-50	IPIP-BFM-20	IPIP-BFM-50
IPIP-BFM-20	Emotional Stability	1	**.90****	.24**	.26**	.08*	.09**	.04	.09**	.14**	.06
	Extraversion	.24**	.21**	1	**.92****	-.02	.02	.15**	.18**	.32**	.29**
	Conscientio-usness	.08*	.10**	-.02	-.05	1	**.89****	.11**	.12**	-.01	-.11**
	Agreeableness	.04	.07*	.15**	.21**	.11**	.18**	1	**.88****	.25**	.24**
	Intellect	.14**	.16**	.32**	.41**	-.01	.12**	.25**	.32**	1	**.88****
NEO-PI-R	Neuroticism	**-.70****	**-.72****	-.25**	-.27**	-.14**	-.18**	-.03	-.07*	-.21**	-.11**
	Extraversion	.23**	.20**	**.59****	**.65****	-.12**	-.04	.26**	.31**	.39**	.39**
	Conscientio-usness	.16**	.19**	.00	.03	**.55****	**.67****	.14**	.18**	.13**	.08*
	Agreeableness	-.03	.00	-.13**	-.16**	.14**	.12**	**.37****	**.46****	-.09**	-.13**
	Openness	-.06	-.04	.22**	.30**	-.13**	-.03	.30*	.36**	**.54****	**.64****
IPIP-NEO-PI-R	Neuroticism	**-.67****	**-.68****	-.27**	-.29**	-.11*	-.16**	-.03	-.06	-.18**	-.10
	Extraversion	.27**	.23**	**.55****	**.63****	-.16**	-.16**	.21**	.25**	.31**	.32**
	Conscientio-usness	.15**	.17**	.02	.01	**.56****	**.67****	.18**	.20**	.15**	.04
	Agreeableness	-.09	-.04	-.08	-.09	.23**	.26**	**.47****	**.56****	.03	-.04
	Openness	-.09	-.07	.23**	.31**	-.09	.01	.30**	.33**	**.58****	**.65****
BFAS	Neuroticism	**-.69****	**-.76****	-.20**	-.21**	-.10	-.14*	-.07	-.11	-.17**	-.09
	Extraversion	.32**	.32**	**.61****	**.67****	-.06	.07	.28**	.31**	.44**	.43**
	Conscientio-usness	.16**	.17**	.02	.01	**.56****	**.62**	.10	.16**	-.02	-.09
	Agreeableness	-.04	.02	-.13*	-.09	.13*	.17**	**.54****	**.63****	.09	.05
	Openness	.10	.13*	.18**	.27**	.06	.16**	.28**	.37**	**.57****	**.61****
IPIP-45AB5C	Emotional Stability	**.77****	**.87****	.07*	.06	.18**	.19**	.05	.10**	.10**	-.02
	Extraversion	.33**	.28**	**.80****	**.89****	-.08*	-.00	.22**	.28**	.44**	.43**
	Conscientio-usness	.17**	.21**	-.02	-.03	**.76****	**.88****	.13**	.17**	.13**	.03
	Agreeableness	-.05	-.01	.05	.08*	.20**	.25**	**.74****	**.85****	.21**	.18**
	Intellect	.03	.05	.24**	.33**	.01	.19**	.31**	.37**	**.77****	**.87****

Note. * $p < .05$; ** $p < .01$.

Moreover, the differences between correlation coefficients of corresponding IPIP-BFM-20 and IPIP-BFM-50 scales were found not to be higher than .12. Given that IPIP-BFM-50 scales consist of 10 items while in the shortened version there are four items, the decrease of correlation coefficients by .12 may be regarded as small.

*

Analyses based on the procedure proposed by Donnellan and colleagues (2006) made it possible to construct the Polish 20-item IPIP-BFM-20 questionnaire, being a shortened version of the IPIP-BFM-50, which serves to measure the Big Five in Goldberg's lexical model. The IPIP-BFM-20 questionnaire is characterized by good validity (satisfactory results in CFA fit indices and correlation coefficients with other Big Five measures) as well as reliability (acceptable Cronbach's α values).

The results of the analyses carried out so far suggest that the Polish version of the IPIP-BFM-20 is at least as good an instrument for measuring the Big Five as its English-language counterpart. It may therefore be useful in all kinds of research where reducing the number of items in the personality questionnaires is advisable.

REFERENCES

Ashton, M. C., & Lee, K. (2005). Honesty-humility, the Big Five, and the Five-Factor Model. *Journal of Personality, 73*(5), 1321-1353.

Baldasaro, R. E., Shanahan, M. J., & Bauer, D. J. (2013). Psychometric properties of the Mini-IPIP in a large, nationally representative sample of young adults. *Journal of Personality Assessment, 95*(1), 74-84.

Batinic, B., Wolff, H. G., & Haupt, Ch. M. (2007). Construction and factorial structure of a short version of the Trendsetting Questionnaire (TDS-K): A cross-validation using multigroup confirmatory factor analyses. *European Journal of Psychological Assessment, 24*(2), 88-94.

Browne, M. (2000). Cross-validation methods. *Journal of Mathematical Psychology, 44*(1), 108-132.

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.

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.

De Raad, B., & Perugini, M. (2002). Big Five factor assessment: Introduction. In B. De Raad & M. Perugini (Eds.), *Big Five assessment* (pp. 1-26). Seattle–Toronto–Bern–Göttingen: Hogrefe and Huber Publishers.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*(5), 880-896.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The Mini-IPIP scales: Tiny-yet-effective measures of the Big Five factors of personality. *Psychological Assessment*, *18*(2), 192-203.

Fischbach, A., & Moosbrugger, H. (2007). The German EPP-D: How to shorten and improve the reliability and validity of Eysenck's Personality Profiler (EPP). *European Journal of Psychological Assessment, 24*(2), 95-100.

Frazier, T. W., Naugle, R. I., & Haggerty, K. A. (2006). Psychometric adequacy and comparability of the short and full forms of the personality assessment inventory. *Psychological Assessment, 18*(3), 324-333.

Goldberg, L. R. (1990). An alternative "description of personality": The Big Five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216-1229.

Goldberg, L. R. (1992). The development of markers of the Big Five factor structure. *Psychological Assessment, 4*(1), 26-42.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (vol. 7, pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*(1), 84-96.

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.

Herzberg, P. Y., & Brähler, E. (2006). Assessing the Big-Five personality domains via short forms: A cautionary note and a proposal. *European Journal of Psychological Assessment, 22*(3), 139-148.

Hofstee, W. K. B., de Raad, B., & Goldberg, L. R. (1992). Integration of the Big Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology, 63*(1), 146-163.

Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102-138). New York: Guilford Press.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cut-off values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320-341.

McCrae, R. R., & Costa, P. T. (2007). Brief versions of the NEO-PI-3. *Journal of Individual Differences, 28*(3), 116-128.

Rammstedt, B. (2007). The 10-Item Big Five Inventory: Norm values and investigation of sociodemographic effects based on a German population representative sample. *European Journal of Psychological Assessment, 23*(3), 193-201.

Reise, S. P., & Henson, J. M. (2000). Computerization and adaptive administration of the NEO-PI-R. *Assessment, 7*(4), 347-364.

Samson, A. C., & Huber, O. W. (2010). Short German versions of Empathizing and Systemizing Self-Assessment Scales. *Swiss Journal of Psychology, 69*(4), 239-244.

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment, 63*(3), 506-516.

Saucier, G., & Goldberg, L. R. (2002). Assessing the Big Five: Applications of 10 psychometric criteria to the development of marker scales. In B. de Raad & M. Perugini (Eds.), *Big Five assessment* (pp. 30-54). Ashland, OH: Hogrefe & Huber Publishers.

Strus, W., Cieciuch, J., & Rowiński, T. (2011). Kołowy model struktury cech osobowości w ujęciu Lewisa Goldberga. *Studia Psychologica, 11*(2)*,* 65-93.

Strus, W., Cieciuch, J., & Rowiński, T. (2012). *Polish version of Big Five Aspects Scales from International Personality Item Pool*. Poster for the 16th European Conference on Personality, Triest, July 10-14, 2012.

Strus, W., Cieciuch, J., & Rowiński, T. (2014a). Circumplex structure of personality traits measured with the IPIP-45AB5C questionnaire in Poland*. Personality and Individual Differences, 71*, 77-82.

Strus, W., Cieciuch, J., & Rowiński, T. (2014b). The Polish adaptation of the IPIP-BFM-50 questionnaire for measuring five personality traits in the lexical approach. *Roczniki Psychologiczne, 17*(2), 347-366.

Thalmayer, A. G., Saucier, G., & Eigenhuis, A. (2011). Comparative validity of brief to medium-length Big Five and Big Six personality questionnaires. *Psychological Assessment, 23*(4), 995-1009.

Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development, 37*(2), 70-84.