ROCZNIKI PSYCHOLOGICZNE / ANNALS OF PSYCHOLOGY
2014, XVII, 2, 347-366

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 POLISH ADAPTATION OF THE IPIP-BFM-50 QUESTIONNAIRE FOR MEASURING FIVE PERSONALITY TRAITS IN THE LEXICAL APPROACH

The article presents the Polish adaptation of Goldberg's IPIP-BFM-50 questionnaire for measuring the five personality traits in the lexical tradition (Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellect). The adaptation procedure was carried out as a series of eight studies. Analyses were performed on a total of N = 7015 people aged from 10 to 83 (their mean age was 29 years). Reliability was assessed using Cronbach's alpha coefficient. Factorial validity was verified in confirmatory factor analysis. In multigroup confirmatory factor analysis, measurement invariance between various research situations was verified. External validity was assessed by comparing the scores obtained using the IPIP-BFM-50 with NEO-FFI and NEO-PI-R scores. The results support the conclusion that the IPIP-BFM-50 is a measure with satisfactory psychometric properties, fit for use in scientific research.

Keywords: personality, personality traits, the Big Five, International Personality Item Pool.

Corresponding author: WŁODZIMIERZ STRUS – Cardinal Stefan Wyszyński University in Warsaw, ul. Wóycickiego 1/3, budynek 14, 01-938 Warszawa; e-mail: w.strus@uksw.edu.pl

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.

The model of Big Five personality factors is currently the most widely known and the most frequently used taxonomy of personality traits. The model is usually referred to as the Big Five or the Five-Factor Model. Although these names are sometimes used interchangeably, they have their roots in two different research traditions. One of these is the lexical tradition and the other one is the tradition of questionnaire research (De Raad & Perugini, 2002; John & Srivastava, 1999; cf. Siuta, 2009).

The questionnaire tradition gave rise to commonly known and widely used questionnaires for measuring five personality traits: the NEO-PI-R (NEO Personality Inventory Revised) and the NEO-FFI (NEO Five-Factor Inventory), developed by Costa and McCrae (1992). The NEO-PI-R was adapted into Polish by Siuta (2009), and the Polish adaptation of the NEO-FFI was prepared by Zawadzki, Strelau, Szczepaniak and Śliwińska (1998). The lexical tradition also developed measures known in the literature worldwide, but they have not been commonly used in Poland so far. Filling this gap is the primary aim of the present article, which is a presentation of the Polish adaptation of Goldberg's (1992) IPIP-BFM-50 (i.e., the 50-item Big Five Markers questionnaire from the resources of the International Personality Item Pool). It consists of sentences (just like questionnaires in the psychometric tradition) but serves to measure the Big Five as identified in lexical research.

A majority of measures in the questionnaire tradition are available today in the form of licensed commercial questionnaires. They are used not only in scientific research but also in individual assessment. It is partly because of the diagnostic character of these measures that they are under special protection (e.g., it is forbidden to publish or modify test items, and the questionnaires may be used almost exclusively by authorized assessment psychologists), one of the consequences being that these questionnaires are available at a charge.

In Poland, there is a lack of noncommercial instruments measuring five personality traits that would have good psychometric properties while having been created for research purposes (not for individual assessment), even though there are many such instruments in the literature worldwide. Filling this gap is the second aim of the present paper. The IPIP-BFM-50 questionnaire is part of the resources of the International Personality Item Pool (IPIP; Goldberg, 1999; Goldberg, Johnson, Eber, et al., 2006), being a collection of test items and questionnaires available for researchers free of charge, without usage restrictions typical for licensed commercial measures. A consequence of the questionnaires from IPIP resources being designed for research is the lack of norms that usually accompany licensed measures serving also for diagnostic examinations. This is

also the case for the IPIP-BFM-50 questionnaire, which we have adapted in the research presented below.

It is worth adding that this questionnaire is referred to in the literature by various names: IPIP Big Five (Zheng, Goldberg, Zheng et al., 2008), Goldberg's IPIP 50 (Guenole & Chernyshenko, 2005), or IPIP Five-Factor Model (Donnellan, Oswald, Baird, & Lucas, 2006). Since there is no single established name, we propose IPIP-BFM-50, referring to the 50-item Big-Five Markers from the resources of the International Personality Item Pool. The name we propose appears to be the most precise one because it points to three key properties of this questionnaire: 1) its origination in the IPIP project, 2) its direct link with the lexical tradition, and 3) the version of the measure, a 50-item version in this case, whereas IPIP resources contain a 100-item version of the BFM questionnaire as well (Strus, Cieciuch, Davidov et al., 2013).

Five Personality Traits in the Lexical Tradition

The Five-Factor Model was originally developed and then verified within the lexical research tradition. Partly independently and partly on the basis of lexical research, the psychometric approach in research on the five factors of personality came into being, in which the model was theoretically elaborated and embedded in a broader theory of personality.

The key to the psycholexical approach is the so-called lexical hypothesis. It assumes that the most important individual personality differences have been encoded in the form of individual terms in some or in all of the world's languages (cf. Goldberg, 1981, 1990). This idea inspired a number of studies, initially conducted mainly on the English lexicon, which led to the identification and multiple replication of the structure of five big personality factors. Initiated by Allport and Odbert as well as by Cattell and developed by Fiske, Tupes, and Christal as well as by Norman (John & Srivastava, 1999), this research current was continued in the 1980s and 1990s by Goldberg. Goldberg carried out a series of studies (Goldberg, 1981, 1990, 1992) that made him the leading figure of the lexical approach. It was also Goldberg (1981) that introduced the term Big Five itself. The five personality traits distinguished in this tradition are the following: Extraversion or Surgency (Factor I), Agreeableness (Factor II), Conscientiousness (Factor III), Emotional Stability (Factor IV), and Intellect or Imagination (Factor V).

Adjective-Based Measurement of the Five Traits in the Lexical Tradition

The main aim of the lexical tradition was to describe the structure of personality encoded in language, understood as a set of independent factors. Various lists of adjectives were used, though they served as material for lexical research rather than as instruments for measuring any particular constructs (cf. Goldberg, 1990, 1999). However, when that aim had been achieved and personality factors had been identified, instruments for measuring them also began to appear.

These instruments traditionally consisted of adjectives used in self-report and observer-rating studies. The most frequently used measures in this tradition include two versions of Goldberg's (1992) adjectival Big Five Factor Markers (BFM) and Saucier's (1994) Mini-Markers. The first version of BFM consisted of 100 adjectives, rated on a 9-point scale. The second version of BFM consisted on 50 pairs of opposing adjectives. Saucier's Mini-Markers (1994) is a 40-adjective version of the 100-adjective BFM measure.

All these measures were developed on the basis of the English language, in which the Big Five was identified. At the same time, lexical research was carried out in languages other than English, too (cf. e.g., De Raad, Perugini, Hrebickova, & Szarota, 1998; Gorbaniuk, Budzińska, Owczarek, Bożek, & Juros, 2013). Sometimes it also resulted in measures being developed. In Poland, Szarota (1995) created the Polish Adjective List for measuring five personality traits identified in Polish lexical studies. Still, these measures were only used locally because they were designed for measuring personality traits identified in a particular local language and within local culture.

Sentence-Based Measurement of the Five Traits in the Lexical Tradition

In the lexical tradition, measures had the form of adjective lists. By contrast, from the very beginning, measures in the questionnaire tradition had the form of sentence sets. Either form involves considerable problems. Adjectives as items in measures of personality represent behaviors on a high level of abstraction. They are very general, imprecise, and often ambiguous; they do not take into account the context or the motivational aspect; they also make up a finite set (Jarmuż, 1994; Saucier & Goldberg, 2002).

Sentences are more semantically specific and embedded in a context and may take motivation into account, but in the case of commonly used questionnaires such as the NEO-PI-R or the NEO-FFI they are often rather long and

complicated. As a result, their understanding is also often ambiguous and depends on participants' verbal skills. For this reason, in the IPIP project, of which the IPIP-BFM-50 is part, Goldberg (1999) adopted a solution that overcomes – at least partly – the shortcomings of both adjectives and sentences as items in personality measures.

The IPIP-BFM questionnaire measures the five personality factors identified in the lexical tradition (Goldberg, 1992), but its items are sentences. At the same time, the form of these sentences differs from that of the sentences usually making up measures in the questionnaire tradition. This is because IPIP items follow the format developed by Hendriks, Hofstee, and De Raad (1999). Its essence lies in the items being short sentences formulated in behavioral terms.

The IPIP-BFM was developed on the basis of a study in which participants rated themselves using, among others, the 100-adjective version of BFM and responded to the pool of 1252 sentence items from IPIP resources (Goldberg, 1999). The items for the sentence-based version of BFM were selected on the basis of correlations with factors obtained in a study using the adjective-based version of BFM. This manner of selecting items allowed to avoid arbitrary choice of sentences and at the same time made it possible to avoid ambiguity, characteristic of adjective-based scales.

Two versions of the IPIP-BFM were thus developed: the 100-item version (IPIP-BFM-100) and the 50-item version (IPIP-BFM-50). All the items of the IPIP-BFM-50 are present in the IPIP-BFM-100, and correlations between the scales of the two versions ranged from .94 to .96 (Saucier & Goldberg, 2002).

Five Personality Factors in the Lexical and Questionnaire Traditions

Despite differences in the conceptualization of the five personality factors between the lexical and questionnaire traditions, a far-reaching correspondence exists between the two models – both in the theoretical meaning of the factors and in the empirical research conducted (cf. Biderman, Nguyen, Cunningham et al., 2011; Goldberg, 1992; John & Srivastava, 1999; McCrae & John, 1992).

The three most important differences between Goldberg's (1992) Five-Factor Model, which is the most widely known model in the lexical tradition, and Costa's and McCrae's model (1992), which is the most widely known model in the questionnaire tradition, are the following: (1) the meaning of Factor V: Intellect in the lexical tradition comprises a more narrow range of personality properties than Openness to Experience in the Five-Factor Model in the questionnaire tradi-

tion; (2) the placing of the warmth trait – in the Five-Factor Model in the questionnaire tradition it is a facet of Extraversion, whereas in the lexical model it falls into Agreeableness (John & Srivastava, 1999); (3) the name of Factor IV: the factor is understood in similar ways in the two models, but it is named after the Emotional Stability pole in the lexical approach and after the Neuroticism pole in the questionnaire tradition.

The IPIP-BFM-50 is an instrument for measuring the Big Five in Goldberg's (1992) lexical approach, corresponding to the five factors of personality in the questionnaire tradition. In the lexical approach, factors are traditionally described by the adjectives that have the highest loadings on them. In Table 1 we propose descriptions of variables measured by the IPIP-BFM-50 questionnaire. Those descriptions are based on Goldberg's (1992) list of 100 best lexical markers of the Big Five, namely on the Big Five Factor Markers.

Table 1
Description of the Five IPIP-BFM-50 Scales

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 self-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.

Hypotheses in the Present Study

When adapting the measure into Polish, we formulated the following expectations:

1. We expected the scales to have satisfactory reliability. We verified this expectation by analyzing the values of Cronbach's alpha.

2. We expected the questionnaire to have a five-factor structure. We verified this expectation by performing confirmatory factor analysis.

3. We expected the IPIP-BFM-50 to be a measure unaffected by various research conditions. We verified this expectation by performing two measurement invariance tests. The first test verified the invariance between the arrangement of test items in the questionnaire and an arrangement mixed with a pool of other test items. The other test verified measurement invariance between two research conditions: paper-and-pencil study and online study.

4. We expected satisfactory external validity. In verifying this expectation, we assumed that Goldberg's (1992) lexical Big Five was similar to the five factors of personality distinguished by Costa and McCrae (1992). We performed a verification of expectations using two techniques. The first was the analysis of correlations with the measurements of five personality traits in the questionnaire tradition. The other was the comparison of results concerning the differentiation of personality traits by gender and age with the results reported in the literature, obtained using measures developed in the questionnaire tradition.

METHOD

Participants and Procedure

We conducted a series of eight studies, of which seven were carried out using the paper-and-pencil method and one in online conditions. Those studies were carried out as part of several research projects concerning various aspects of personality, its structure and development.

Participation was voluntary and anonymous for everyone. The first, second, fifth, and seventh studies were carried out by trained students, who volunteered to help. Each student carried out a study in a group of 5 to 10 people. The second and third studies were conducted on a group basis, at schools and universities, by trained research assistants. The sixth study was carried out using the paper-and-pencil method by Magdalena Leśniewska as part of her master's thesis research.

The eighth study was carried out online by the authors and by students as part of their master's thesis research. Participants were recruited via Facebook.

Data were collected from a total of 7127 people, but the analyses presented below were performed on a group of 7015 participants, 112 individuals (1.5%) having been excluded due to outliers or missing data. We adopted 10% of unanswered test items as the threshold. The scores of individuals with the number of missing data equal to or higher than this threshold were excluded from further analyses. Table 2 gives the number of individuals analyzed in each study, together with information about their gender and age.

Table 2
Age, Gender, and the Number of Participants in Study Groups

Group	*N* (% of women)	Age	
		M	*SD*
Study 1	936 (54.9)	30.78	13.68
Study 2	685 (48.2)	31.10	11.39
Study 3	304 (50.5)	18.00	0.13
Study 4	414 (34.6)	22.02	3.48
Study 5	679 (34.3)	27.20	12.30
Study 6	861 (60.4)	38.94	14.44
Study 7	789 (56.7)	29.65	12.26
Study 8	2347 (58.3)	27.23	9.85
Total	7015 (52.9)	28.93	12.16

Measure

IPIP-BFM-50 test items

The IPIP-BFM-50 questionnaire was developed as part of Goldberg's IPIP project (Goldberg, 1999; Goldberg, Johnson, Eber et al., 2006), which also includes other instruments for measuring personality traits. One of them is the IPIP-45AB5C questionnaire, measuring 45 variables of the AB5C model (Abridged Big Five Dimensional Circumplex), developed by Hofstee, De Raad, and Goldberg (1992). Strus, Cieciuch, and Rowiński (2014) adapted this measure

into the Polish language. The measure consists of 486 items, including 48 items of the IPIP-BFM-50. In the first study we used the IPIP-45AB5C questionnaire, extended by the missing two items of the IPIP-BFM-50. The study using the IPIP-45AB5C was carried out in two rounds with a two-week interval between them.

The IPIP-BFM-50 questionnaire

In all the successive studies, the IPIP-BFM-50 was used as a separate measure, consisting of 50 items. Each of the scales (Extraversion, Conscientiousness, Agreeableness, Emotional Stability, and Intellect) consists of 10 items. Participants respond to the items on a 5-point Likert scale (*very inaccurate*) to 5 (*very accurate*). The items were translated by the authors of the present paper. In the process of translation, efforts were made to ensure both linguistic fidelity to the original and theoretical correspondence between the items and the constructs measured.

The NEO-PI-R and the NEO-FFI

In three studies we also applied measures of personality traits developed in the questionnaire tradition. In a subgroup of $N = 685$ participants in the second study, the measure used was the NEO-PI-R by Costa and McCrae (1992) as adapted into Polish by Siuta (2006). This questionnaire was applied in a separate research session, carried out about two weeks after the study using the IPIP-BFM-50. In the seventh study, the measure applied was Costa's and McCrae's (1992) NEO-FFI as adapted into Polish by Zawadzki and colleguess (1998), which was administered to $N = 782$ individuals.

RESULTS AND DISCUSSION

Descriptive Statistics

Table 3 presents descriptive statistics for each of the five traits measured by the IPIP-BFM-50. The distribution of scores for each scale is close to normal distribution. Skewness and kurtosis range between -1 and +1. A little deviation was only found in the case of the Agreeableess scale in the fourth study.

Table 3
Reliability and Descriptive Statistics of the Questionnaire's Scales for Each Study Group

Scale	Study group	*M*	*SD*	Skewness	Kurtosis	Cronbach's α
Extraversion	Study 1	3.30	0.72	-0.041	-0.401	.86
	Study 2	3.36	0.71	-0.092	-0.191	.87
	Study 3	3.46	0.73	-0.199	-0.465	.86
	Study 4	3.35	0.70	0.027	-0.039	.82
	Study 5	3.35	0.75	-0.103	-0.442	.89
	Study 6	3.39	0.70	0.126	-0.333	.83
	Study 7	3.49	0.78	-0.211	-0.466	.91
	Study 8	3.31	0.81	-0.151	-0.390	.90
Agreeableness	Study 1	3.91	0.52	-0.505	0.169	.79
	Study 2	3.86	0.55	-0.283	-0.179	.82
	Study 3	3.70	0.57	-0.069	-0.631	.79
	Study 4	3.81	0.65	-0.096	-1.128	.79
	Study 5	3.93	0.54	-0.432	-0.003	.82
	Study 6	3.86	0.63	-0.250	-0.694	.82
	Study 7	3.98	0.50	-0.484	0.152	.79
	Study 8	3.84	0.63	-0.611	0.370	.84
Conscientiousness	Study 1	3.46	0.63	-0.145	-0.321	.80
	Study 2	3.58	0.63	-0.297	-0.137	.82
	Study 3	3.14	0.64	0.012	-0.184	.80
	Study 4	3.37	0.63	0.186	-0.423	.75
	Study 5	3.49	0.61	-0.100	-0.241	.79
	Study 6	3.62	0.61	-0.252	-0.061	.75
	Study 7	3.47	0.65	-0.086	-0.564	.83
	Study 8	3.43	0.68	-0.138	-0.298	.83
Emotional Stability	Study 1	2.98	0.76	-0.018	-0.365	.87
	Study 2	3.11	0.79	-0.252	-0.368	.90
	Study 3	3.02	0.80	-0.051	-0.547	.88
	Study 4	3.05	0.79	0.030	-0.258	.86
	Study 5	2.93	0.72	-0.121	-0.181	.87
	Study 6	3.15	0.76	0.180	-0.016	.86
	Study 7	3.19	0.80	-0.205	-0.551	.90
	Study 8	3.01	0.84	-0.048	-0.558	.90

Scale	Study group	*M*	*SD*	Skewness	Kurtosis	Cronbach's α
Intellect	Study 1	3.49	0.57	-0.099	-0.231	.77
	Study 2	3.46	0.55	0.092	-0.115	.78
	Study 3	3.59	0.58	-0.016	-0.450	.78
	Study 4	3.56	0.53	0.239	-0.334	.70
	Study 5	3.56	0.57	0.164	-0.147	.80
	Study 6	3.41	0.57	0.164	-0.258	.73
	Study 7	3.61	0.59	-0.087	-0.506	.82
	Study 8	3.73	0.58	-0.218	-0.155	.78

Measurement Reliability

The values of Cronbach's α coefficient obtained in each study for the five scales of the IPIP-BFM-50 are presented in Table 3. They range between .73 (Intellect in the sixth study) to .91 (Extraversion in the seventh study). Discriminating power ranges from .39 (Intellect in the seventh study) to .67 (Extraversion in the seventh study). The mean values of Cronbach's α, computed on the basis of all the eight studies, are the following: .87 for Extraversion; .81 for Agreeableness; .80 for Conscientiousness; .88 for Emotional Stability, and .77 for Intellect. Measurement reliability may therefore be called very high.

Factorial Validity

Factorial validity was verified in confirmatory factor analysis, performed using Amos 21 software. Because of the large number of test items measuring a given trait, we applied the parceling procedure. It consisted in group means instead of individual items being entered as observable variables (Little, Cunningham, Shahar, & Widaman, 2002). In our research, we divided items making up each scale into three parcels randomly. Figure 1 presents the tested model, with factor loadings and correlations between latent variables in the aggregate group from all studies ($N = 7015$).

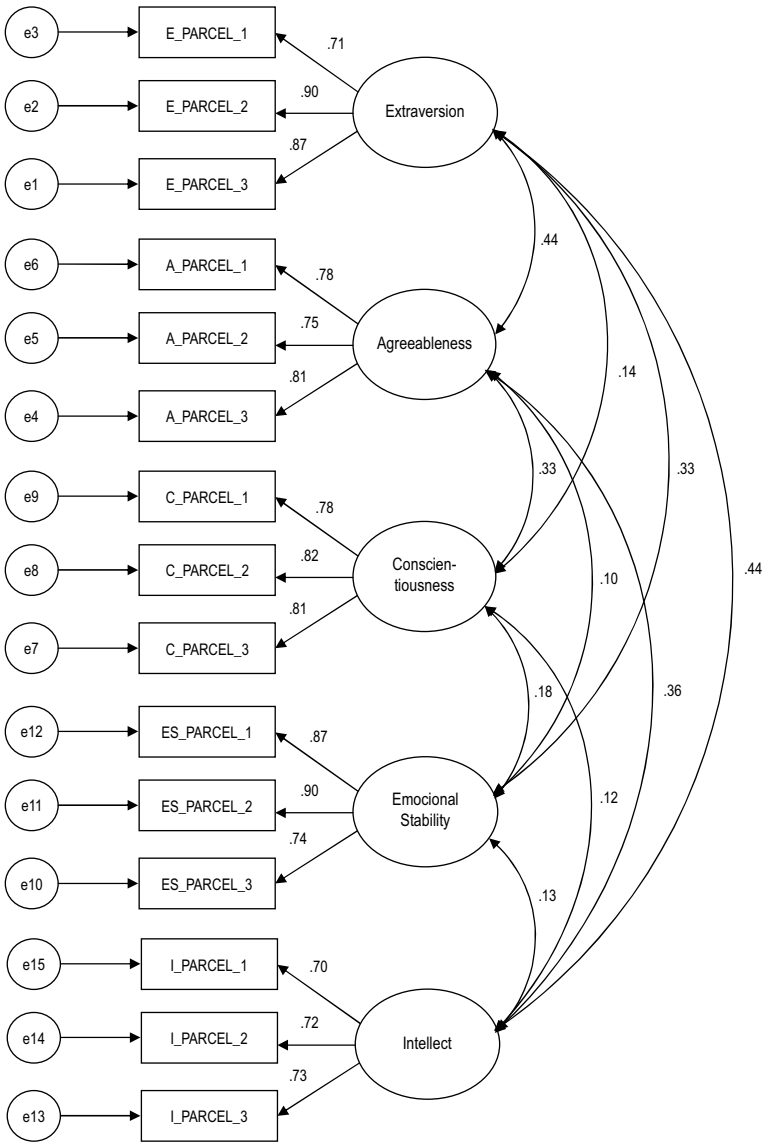

Figure 1. The measurement model of the IPIP-BFM-50 questionnaire with factor loadings and correlations between latent variables in the aggregate group from all studies ($N = 7015$).

The assessment of model fit to data was based on RMSEA, CFI, and SRMR indices. RMSEA and SRMR below .08 and CFI above .9 are adopted as thre-

shold values of model acceptability (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004).

Table 4 presents model fit indices for each of the eight studies and for all the data taken together. In the fourth study, the model was found to fit the data poorly. In the fifth study, RMSEA is higher than .08 and in the seventh study is equal to that value. SRMR falls within the acceptable range in all of the studies and CFI falls within that range in all except the fourth one. All the three goodness-of-fit indices obtained for the whole sample fall within the limits of acceptability according to the criteria given above. Taking into account the specificity of measurement and the problems, reported in the literature, with confirmatory factor analysis of questionnaires for measuring the Big Five as well as the values of fit indices obtained for the whole sample, the obtained results may be regarded as satisfactory verification of the five-factor structure of the IPIP-BFM-50 questionnaire.

Table 4
*Fit Indices of Confirmatory Factor Analysis for Each Study Group (*df = *80, χ^2 was significant in all studies)*[1]

Group	χ^2	CFI	RMSEA (90%)	SRMR
Study 1	396.96	.944	.065 [.059-.072]	.053
Study 2	377.09	.942	.074 [.066-.081]	.057
Study 3	217.38	.927	.075 [.063-.087]	.057
Study 4	390.99	.890	.097 [.088-.107]	.078
Study 5	446.09	.926	.082 [.075-.090]	.058
Study 6	505.18	.926	.079 [.072-.085]	.052
Study 7	488.06	.936	.080 [.074-.087]	.062
Study 8	1146.23	.943	.075 [.072-.079]	.056
The entire sample	3193.33	.938	.074 [.072-.077]	.054

Table 5 presents intercorrelations between latent variables from confirmatory factor analysis (below the diagonal) and observable variables, computed from the

[1] Information concerning standardized parameters of the model and the key to the test items making up each scale are available from the authors upon request.

key (above the diagonal). The results presented in Table 5 were obtained in the analysis of the entire sample.

Table 5
*Intercorrelations Between Scales of the IPIP-BFM-50 for the Entire Sample (*N = *7015). Below the diagonal, the table shows correlations between latent variables in confirmatory factor analysis, and above the diagonal – correlations between observable variables are shown (all the correlations are significant at* p < *.01)*

Scales	E	A	C	ES	I
E (Extraversion)	1	.34	.09	.27	.36
A (Agreeableness)	.44	1	.27	.10	.27
C (Conscientiousness)	.14	.33	1	.16	.07
ES (Emotional Stability)	.33	.10	.18	1	.09
I (Intellect)	.44	.36	.12	.13	1

Measurement Invariance Between Different Research Conditions

Two measurement invariance tests were applied in different research conditions. The first one tested measurement invariance between the following two situations: (1) the measurement of five traits using the IPIP-BFM-50 questionnaire in one research act and (2) the measurement of five traits by means of items from that questionnaire dispersed in a pool of 492 items for measuring 45 personality traits using the IPIP-45AB5C in two research acts, separated by a two-week interval. The second test of invariance was carried out between the traditional paper-and-pencil research method and online research.

Measurement invariance was verified in the procedure of multigroup confirmatory factor analysis (MGCFA; Cieciuch & Davidov, 2014; Vandenberg & Lance, 2000). Three levels of invariance were tested: configurational, metric, and scalar. According to Chen's (2007) proposal, metric measurement invariance in groups of $N > 300$ is considered acceptable when $\Delta CFI < .01$, $\Delta RMSEA < .015$, and $\Delta SRMR < .03$ between the configurational and metric levels. For scalar invariance, Chen (2007) proposes the following threshold values: $\Delta CFI < .01$, $\Delta RMSEA < .015$, and $\Delta SRMR < .01$ between the metric and scalar levels.

Table 6 presents model fit indices for each of the three levels of invariance in the two tests carried out. It turned out that in all conditions the measurement was invariant at the configurational, metric, and scalar levels, since CFI, RMSEA,

and SRMR delta values all fell within the acceptable range. This means that the IPIP-BFM-50 questionnaire is unaffected by various research conditions and yields results undistorted by the specificity of these conditions.

Table 6
Model Fit Indices in Multigroup Confirmatory Factor Analysis, Testing IPIP-BFM-50 Measurement Invariance in Different Research Conditions

Invariance level	χ^2	*df*	CFI	RMSEA	SRMR
The IPIP-BFM-50 questionnaire *versus* IPIP-BFM-50 questionnaire items in a pool of 492 items					
Configurational	2305.634	160	.936	.053 [.051-.055]	.053
Metric	2344.430	170	.935	.052 [.050-.054]	.053
Scalar	2502.617	180	.930	.052 [.050-.054]	.053
Paper-and-pencil *versus* online study					
Configurational	3363.859	160	.939	.053 [.051-.055]	.053
Metric	3407.622	170	.938	.052 [.050-.058]	.053
Scalar	3598.179	180	.935	.052 [.050-.058]	.054

Note. In testing invariance between different arrangements of IPIP-BFM-50 items, the groups consisted of $N = 3732$ and $N = 936$ participants, respectively; in testing invariance between the online and offline versions, the groups consisted of $N = 2347$ and $N = 4668$ participants, respectively.

External Validity

Correlations with other Big Five measures

Table 7 shows correlations of IPIP-BFM-50 scales with NEO-PI-R (the first and the second study) and NEO-FFI scales (the seventh study). The correlation coefficient values obtained confirm the theoretical validity of the IPIP-BFM-50. Correlations between corresponding scales were considerably higher than the remaining ones. The highest (negative) correlations were found between IPIP--BFM-50 Emotional Stability scales and the Neuroticism scales of the NEO-FFI and the NEO-PI-R. Rather unexpectedly, correlations were the lowest not in the case of Intellect and Openness to Experience scales but in the case Agreeableness scales. Especially the correlation between these scales in the IPIP-BFM and their counterparts in the NEO-PI-R was much lower than expected. This may be due to certain differences in the conceptualization of the Agreeableness factor between the lexical and questionnaire traditions. This is not only about the above-mentioned warmth, which is part of Agreeableness in the lexical tradition and an facet of Extraversion in the questionnaire tradition. Agreeableness in the ques-

tionnaire tradition contains more aspects connected with modesty and morality than the corresponding factor in the lexical approach (cf. Ashton & Lee, 2005).

Table 7
*Values of Pearson's r Coefficient of Correlation Between the IPIP-BFM-50 and the NEO-FFI (*N = *782) as Well as the NEO-PI-R (*N = *685)*

	Scales	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Intellect
NEO-FFI	Extraversion	**.67****	.32**	.01	.25**	.34**
	Agreeableness	.06	**.51****	.11*	.24**	-.09*
	Conscientiousness	.15**	.17**	**.69****	.26**	.09*
	Neuroticism	-.35**	-.17**	-.19**	**-.70****	-.26**
	Openness	.23**	.24**	-.04	.09*	**.56****
NEO-PI-R	Extraversion	**.60****	.28**	.01	.19**	.31**
	Agreeableness	-.12**	**.47****	.12**	.05	-.14
	Conscientiousness	.08*	.13**	**.61****	.18**	.07
	Neuroticism	-.25	-.04	-.22	**-.65**	-.14
	Openness	.22**	.31**	.02	.03	**.55****

Note. * $p < .01$; ** $p < .001$. The correlation coefficients between corresponding scales are in bold.

Differentiation of personality traits by gender

Comparisons of scores obtained by women and by men were performed on the entire sample, after the confirmation of scalar measurement invariance (Table 8). The greatest gender differences occurred in the case of Emotional Stability and Agreeableness. Women showed a considerably higher level of Neuroticism and Agreeableness than men. That result is fully consistent with the research conducted using the NEO-PI-R and the NEO-FFI both in Poland (Siuta, 2006; Zawadzki et al., 1998) and in the United States (Costa & McCrae, 1992). On the other hand, a higher level of Intellect was shown by men, which runs contrary to studies using NEO questionnaires. In those studies, it is women that showed higher Openness (Siuta, 2006; Zawadzki et al., 1998), although it must be said that this was mainly due to such aspects of this factor as aesthetics or feelings, which are more weakly represented in the lexical Intellect. When it comes to

gender differences in Conscientiousness and Extraversion, the results of studies using NEO questionnaires are not so clear. Research using the IPIP--BFM-50 showed a higher level of Conscientiousness in women, but no significant differences in Extraversion were found.

Table 8
*Differences in Means Between the Groups of Women (*N *= 3747) and Men (*N *= 3327)*

Scale	Women		Men		*t*(7072)
	M	*SD*	*M*	*SD*	
Extraversion	3.36	0.77	3.34	0.77	0.82
Agreeableness	3.90	0.59	3.80	0.62	7.29**
Conscientiousness	3.48	0.66	3.43	0.66	2.72**
Emotional Stability	2.97	0.80	3.14	0.79	-8.81**
Intellect	3.57	0.59	3.60	0.58	-2.44*

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

Relationships between personality traits and age

The following tendencies were found: Conscientiousness ($r = .26$), Emotional Stability ($r = .08$), and Agreeableness ($r = .02$) increase while Intellect ($r = -.18$) and Extraversion decrease with age ($r = -.06$). These findings are fully consistent with those obtained using NEO questionnaires in Poland and in the United States (Costa & McCrae, 1992; Siuta, 2006; Zawadzki et al., 1998), although values of the correlation coefficient are rather low.

*

Many instruments have been developed for measuring the basic five personality traits (cf. De Raad & Perugini, 2002). Apart from commercial inventories, such as Costa's and McCrae's (1992) NEO-PI-R and NEO-FFI (*NEO Five-Factor Inventory*), designed not only for scientific research but also for individual assessment, noncommercial questionnaires designed only for research purposes have been gaining popularity in recent years as well. One of them is Goldberg's (1992) IPIP-BFM-50 for measuring five personality traits as identified in the lexical tradition.

The paper presents the Polish adaptation of this questionnaire. The data subjected to analyses was collected from 7015 individuals in eight studies. The

IPIP-BFM-50 has good psychometric properties. Its reliability, verified in the analysis of Cronbach's alpha values, was found to be satisfactory. The questionnaire's five-factor structure may be regarded as satisfactorily confirmed in confirmatory factor analysis. The IPIP-BFM-50 turned out to be a measure unaffected by different research conditions, such as online versus offline study or studies with different arrangements of test items (arrangement of items as in the questionnaire versus questionnaire's statements mixed with other statements).

The IPIP-BFM-50 questionnaire is not subject to any usage restrictions in scientific research. As all the questionnaires from the IPIP project, it may be used free of charge in any form, paper or online. The advantages of the IPIP-BFM-50 presented above show that this measure can be used in the currently studied and discussed areas (cf. Strus & Cieciuch, 2014). This is due not only to good psychometric properties and no usage restrictions but also to the fact that it is a questionnaire measuring the lexical version of the Five-Factor Model. The IPIP-BFM-50 may also be said to be, in some sense, a kind of synthesis of the two traditions that have been instrumental in the emergence and development of this model.

REFERENCES

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

Biderman, M. D., Nguyen, N. T., Cunningham, C. J., & Ghorbani, N. (2011). The ubiquity of common method variance: The case of the Big Five. *Journal of Research in Personality, 45*, 417-429.

Brown, T. A. (2006). Confirmatory factor analysis for applied research. New York: Guilford Press.

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504.

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.

Cieciuch, J., & Davidov, E. (2014). Establishing measurement invariance across online and offline samples. A tutorial with the software packages Amos and Mplus. *Studia Psychologica.*

De Raad, B., & Perugini, M. (Eds.) (2002). *Big Five assessment*. Seattle–Toronto–Bern–Göttingen: Hogrefe and Huber Publishers.

De Raad, B., Perugini, M., Hrebickova, M., & Szarota, P. (1998). Lingua franca of personality: Taxonomies and structures based on the psycholexical approach. *Journal of Cross-Cultural Psychology, 29,* 212-232.

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.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (vol. 2, pp. 141-165). Beverly Hills, CA: Sage.

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: 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.

Gorbaniuk, O., Budzińska, A., Owczarek, M., Bożek, E., & Juros, K. (2013). The factor structure of Polish personality-descriptive adjectives: An alternative psycho-lexical study. *European Journal of Personality, 27*(3), 304-318.

Guenole, N., & Chernyshenko, O. S. (2005). The suitability of Goldberg's Big Five IPIP personality markers in New Zealand: A dimensionality, bias, and criterion validity evaluation. *New Zealand Journal of Psychology, 34,* 86-96.

Hendriks, A. A. J., Hofstee, W. K. B., & de Raad, B. (1999). The Five-Factor Personality Inventory (FFPI). *Personality and Individual Differences, 27*, 307-325.

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.

Jarmuż, S. (1994). Metodologiczne problemy związane z weryfikacją modelu „Wielkiej Piątki". *Przegląd Psychologiczny, 37*(1-2), 195-203.

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.

Little, T., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*(2), 151-173.

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. (2005). *Osobowość człowieka dorosłego*. Krakow: Wydawnictwo WAM.

McCrae, R. R., & John, O. P. (1992). An introduction to the Five-Factor Model and its applications. *Journal of Personality, 60*, 175-215.

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

Siuta, J. (2006). *Inwentarz Osobowości NEO-PI-R Paula T. Costy Jr i Roberta R. McCrae. Adaptacja polska. Podręcznik*. Warsaw: Pracownia Testów Psychologicznych PTP.

Siuta, J. (Ed.) (2009). *Diagnoza osobowości. Inwentarz NEO-PI-R w teorii i praktyce*. Warsaw: Pracownia Testów Psychologicznych PTP.

Strus, W., & Cieciuch, J. (2014, in press). Poza Wielką Piątkę – Nowe modele struktury osobowości. *Polskie Forum Psychologiczne*.

Strus, W., Cieciuch, J., Davidov, E., Vecchione, M., Rowiński, T., Gasiul, H., Stillwell, D., & Kosiński, M. (2013). *Factorial structure of personality as measured by the IPIP-100 in 33 countries.* Paper presented at the 12th European Conference on Psychological Assessment, San Sebastian, Spain, July 17-20, 2013.

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

Szarota, P. (1995). Polska Lista Przymiotnikowa (PLP): Narzędzie do diagnozy pięciu wielkich czynników osobowości. *Studia Psychologiczne, 33*, 227-255.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70.

Zawadzki, B., Strelau, J., Szczepaniak, P., & Śliwińska, M. (1998). *Inwentarz Osobowości NEO-FFI Costy i McCrae. Adaptacja polska.* Warsaw: Pracowania Testów Psychologicznych PTP.

Zheng, L., Goldberg, L. R, Zheng, Y., Zhao, Y., Tang, Y., & Liu, L. (2008). Reliability and concurrent validation of the IPIP Big-Five factor markers in China: Consistencies in factor structure between Internet-obtained heterosexual and homosexual samples. *Personality and Individual Differences, 45*, 649-654.