Journal of Innovation in Psychology, Education and Didactics	Vol. 22, No. 2 2018	167 - 174

DEVELOPMENT AND VALIDATION OF A SHORT SCALE FOR MEASURING BIG FIVE PERSONALITY TRAITS: THE IPIP-BFM-25 INDONESIA

Hanif AKHTAR [a]*, Saifuddin AZWAR [a]

[a] Universitas Gadjah Mada, Yogyakarta, Indonesia

Abstract

The shifting trend of research involving more constructs made the use of short-scale increasingly popular. This research aimed to develop a short version of the IPIP-BFM-50 that measures Big Five personality traits. There were three studies in this research, including adaptation of the parental scale, short-scale development, and short-scale validation. The total subjects used in this study were 1,003 people with an age range of 14 - 46 years. Validity and reliability of the scale were verified through a series of studies. The result of content validity, factorial validity, and convergent validity showed satisfactory result. The entire scale had a reliability coefficient above .70 using internal consistency and test-retest approach. The correlation between IPIP-BFM-25 and BFI with external variables showed the same pattern. In general, IPIP-BFM-25 Indonesia was a psychometrically acceptable and practically useful short scale for measuring the Big Five personality dimensions.

Key words*:* Big Five, Indonesian adaptation, IPIP, short scale

1. Introduction

The big five personality model is currently one of the most recognized personality models and is widely used in the academic disciplines of Psychology. The popularity of the big five personality model implied on the development of scale based on this model. Several scales have been developed based on the Big Five personality model. The existing scales are the Big Five Inventory (BFI) consisting of 44 items (John & Srivastava, 1999), the NEO PI-R consisting of 240 items (Costa Jr & McCrae, 1995), and the Trait Descriptive Adjective (TDA) consisting of

* Corresponding Author. Lecturer Hanif Akhtar
E-mail address: hanifakhtar@umm.ac.id

100 adjectives (Goldberg, 1992). Several studies had been done to adapt the instruments to Indonesia, but unfortunately some instruments such as NEO-PI-R, were copyrighted and might not be used in whole or in part without written permission. This condition would discourage continual efforts on further test development and discourage comparative-validity studies (Goldberg et al., 2006). One of the great prospects in personality measurement is the existence of the International Personality Item Pool (IPIP). IPIP is an International collaboration to develop an inventory of personality that can be accessed by everyone through the website (Goldberg et al., 2006). All researchers are free to use existing items. There are currently 3,320 items composed of multiple scales and there are already 686 international publications using IPIP (http://ipip.ori.org/index.htm). The items in the IPIP are items that also measure the same construct with various commercial instruments, such as IPIP-NEO which measures the same construct as the NEO PI-R (Goldberg et al., 2006).

One of the most popular scales on the IPIP website is the IPIP Big Five Factor Marker (IPIP-BFM) used in the Goldberg (1992) study. This research will use 50-item IPIP version of the Big Five Markers which will be called IPIP-BFM-50. The IPIP-BFM-50 contained 50 items of short phrases that measure five big personality dimensions: Extraversion, Agreeableness, Conscientiousness, Emotional stability and Intellect (Strus, Cieciuch, & Rowiński, 2017). The popularity of IPIP-BFM-50 attracted many researchers around the world to adapt and conduct cross-cultural validation to different countries, such as in Croatia (Mlačić & Goldberg, 2007), Poland (Strus et al., 2017), Scotland (Gow, Whiteman, Pattie, & Deary, 2005), China (Zheng et al., 2008), and Portugal (Oliveira, 2017).

Another problem inhibiting the development of research in personality was the availability of an efficient scale. Most of the existing personality scales had stout items. In general, long-form instruments would have better psychometric properties than short-form (Gosling, Rentfrow, & Swann, 2003). However, researchers sometimes are faced with situations where it is irrelevant to use long-form instruments, such as on internet studies, multiple-variables studies, pre-screening studies, and longitudinal studies (Robins, Tracy, Trzesniewski, Potter, & Gosling, 2001). In general, short scales are less time-consuming, more cost-effective, allow for ease of test administration, and avoid respondents' boredom. (Herzberg & Brähler, 2006; Joseph, Linley, Harwood, Lewis, & McCollam, 2004). Therefore, some developers started to develop an instrument with fewer items but did not sacrifice the validity. In the psychological research, current trends suggest research questions to be more complex and involve more constructs supported by the used of sophisticated statistical methods (Ziegler, Kemper, & Kruyen, 2014).

The involvement of more constructs allowed researchers to see the phenomenon more comprehensively. But consequently, the research would take considerable time, cost and energy. In addition, respondents were also likely to be bored and reluctant to complete the scale. This situation could actually lead to error of measurement. Therefore, a short scale could be a practical alternative to overcome it. To meet the needs of a practical personality scale, some researchers made short versions of existing personality scales, such as NEO FFI (McCrae & Costa, 2004), BFI-10 (Rammstedt & John, 2007), Mini-IPIP (Donnellan, Oswald, Baird, & Lucas, 2006), and Ten Item Personality Inventory (TIPI) (Gosling et al., 2003). Unfortunately, some instruments such as TIPI still had some limitations. TIPI had low alpha reliability and the factor structure was not as expected (Carvalho, Nunes, Primi & Nunes, 2012; Gosling et al., 2003).

In Indonesia, there is currently no short version of the Big Five personality scale. The only measure of personality with the shortest items was the 44-items of BFI adapted by Ramdhani (2012). The major problem faced by personality researchers in Indonesia was related to the lack of a scale that was free to use and efficient but with acceptable psychometric properties. To answer the problem, this study aimed to develop a short version of the existing open-sourced scale. This study was similar to that of Donnellan et al. (2006) who developed the Mini-IPIP from the parent scale of IPIP-BFM-50. The selection of IPIP-BFM-50 was because it was frequently used by many researchers and it was publicly available to researchers on the IPIP website at no cost. Although some procedures replicated of what Donnellan and colleagues (2006) had done, but this study differed because items were selected based on empirical data of Indonesian samples. In addition, the target of selected items was 25 items, different from Mini-IPIP (Donnellan et al., 2006) which had 20 items. In fact, three of the five scales in Mini-IPIP had alpha reliability below .70 (Donnellan et al., 2006), whereas Nunnally (1978) recommended a minimum limit of the reliability coefficients for the research instrument to be .70. Therefore, the scale developed in this study had more items in the hope that the reliability coefficients obtained would also be better.

This research was divided into three studies beginning with the adaptation of the parent measure, IPIP-BFM-50. Furthermore, we developed the IPIP-BFM-25 into Indonesian and conducted a series of validity and reliability tests on short-scale development. Validity was tested by content validity, factorial validity, convergent validity, and Confirmatory Factor Analysis (CFA), while reliability was tested by internal consistency and test-retest approaches. Specifically, we wanted to know whether the short measure had adequate psychometric properties when compared to the parent measure or other big five personality measures.

2. Study 1: Adaptation of IPIP-BFM-50 into Indonesian

The purpose of study 1 was to adapt the parent measure (IPIP-BFM-50) into Indonesian. The procedures of the adaptation were taken under the guidance of the International Test Commission (ITC) (2016) with technical considerations from Beaton, Bombardier, Guillemin & Ferraz (2000) and Azwar (2017).

2.1. Method

Participants

The participants involved for validation of the adaptation scale were 502 people with 212 men (42.2%) and 290 women (57.8%). The subject age ranged from 14 to 46 years with an average of 20.65 years. All participants were domiciled in Yogyakarta.

Material

The material used in Study 1 was IPIP-BFM-50. IPIP-BFM-50 was the target scale to be adapted into Indonesian. IPIP-BFM-50 consisted of 50 items, each item contained a short phrase describing the participants. The participants were asked to assess themselves on a scale of 1 to 5. The original IPIP-BFM-50 had a fairly good alpha reliability ranging from .79 (Conscientiousness) to .87 (Extraversion).

Procedures

There were six steps taken in study 1. First, the two translators translated the original scale independently. Second, the two translators discussed the results of the their translation to produce an agreed translation. The discussions were facilited by a moderator. Third, the profesional translator translated back the agreed translation into English. The results of back-translation were analyzed by the developer of the scale to see whether it had similar meaning. Fourth, the experts, including language and personality experts, conducted a review of the translation of the scale to ensure that the scale had similar meaning to the original and was appropriate for Indonesian. Fifth, the scale was administered to small sample size and we interviewed them to ensure that they could understand the items and the instruction of the scale. Sixth, we administered the scale to large sample size and estimated the validity and reliability. Validity was estimated with content validity (Aiken's V) and factorial validity, whereas reliability was estimated with the internal consistency approach.

2.2. Results

Translation of the scale

After obtaining permission to adapt the scale from the developer, the two translators then translated the original scale into Indonesian independently. Both translators then discussed to select an agreed translation. The agreed translation was translated back into English and consulted to the developer. Developer stated that all items had equivalent meaning to the original items. The next step was a review from the experts. Based on the experts' considerations, minor revision was conducted on items number 23, 32, and 39 because it was not appropriate for Indonesian. The last step was to administer the scale to lay people and interview them. All of the five respondents said that the items and the instruction could already be understood.

Content Validity

Content validity was estimated with Aiken's V index. Items in IPIP-BFM-50 were presented to 12 Subject Expert Matters (SMEs) to assess the relevance of the items to the measured dimensions. Experts provided ratings on items with a score range 1-5. An item was relevant if it had Aiken's V index of at least 0.69 (Aiken, 1985). Of the 50 items available, the index V value ranged from .71 to .98. Thus, all items in the IPIP-BFM-50 translation met the criteria and were stated to be valid to measure the five dimensions of the big fives.

Reliability

Reliability was determined from the coefficient of alpha reliability. Reliability of the scales ranged from .76 (Agreeableness) to .86 (Emotional staibility). All scales had alpha reliability above the criteria of .70 (Nunnally, 1978), so each scale had satisfactory reliability. A summary of the scale properties can be seen in Table 1.

Tabel 1. Item-total correlation and reliability of IPIP-BFM-50 Indonesia (n=502)

Scale	Number of items	Item-total correlation	Alpha	SEM
Extraversion	10	.41 – .67	.83	2.45
Agreeableness	10	.29 – .59	.76	2.27
Conscientiousness	10	.43 – 057	.81	2.49
Emotional stability	10	.39 – .66	.86	2.42
Intellect	10	.26 – .64	.76	2.30

Note: SEM = Standard Error of Measurement

Factorial Validity

Factorial validity was estimated by factor analysis using Principal Component Analysis method forced into five factors. The analysis was done by varimax rotation method to see wheter the item distribution was still appropriate as the dimension measured. The Kaiser-Meyer-Olkin (KMO) result to estimate sample adequacy was .82 and The Barlett Test of Sphericity was 9136.419 with $p < .01$. The result of factor analysis showed that all items were distributed well according to the measured factors. All items had a high loading factor in the primary factor, ranging from .34 to .75, while loading factor on other factors (cross-loading) was relatively low. This indicates that the IPIP-BFM-50 had a satisfactory factorial validity. The distribution of loading factor of each item can be seen in Table 2.

Tabel 2. Rotated Factors of IPIP-BFM-50 (n=502)

Item	Factors				
	1	**2**	**3**	**4**	**5**
Memiliki suasana hati yang sering cepat berubah (ES39)	**.73**	.08	.12	-.03	.06
Mudah merasa jengkel (ES44)	**.73**	-.05	.04	.23	.10
Mudah khawatir (ES14)	**.69**	.14	-.07	-.05	-.02
Mudah merasa kesal (ES29)	**.68**	-.07	.10	.22	.11
Sering merasa sedih (ES49)	**.68**	.13	.09	-.01	-.07
Mudah merasa tertekan (ES4)	**.68**	.12	.02	-.00	-.06
Memiliki perasaan yang berubah-ubah (ES34)	**.67**	.02	.12	-.03	.07
Mudah merasa terganggu (ES24)	**.61**	.05	-.12	.21	.09
Jarang merasa sedih (ES19)	**.53**	.15	.11	-.01	-.07
Merasa tenang hampir setiap saat (ES9)	**.41**	-.05	.10	.10	.12
Berinteraksi dengan banyak orang dalam suatu acara (E31)	.10	**.75**	.11	.14	.04
Sedikit berkata (E26)	.03	**.69**	-.10	-.05	.02
Tidak banyak berbicara (E6)	.03	**.68**	-.08	-.09	.05
Menghidupkan suasana dalam suatu acara (E1)	.06	**.65**	.03	.11	.03
Memulai suatu percakapan (E21)	.05	**.64**	.12	.16	.06
Tidak suka menjadi pusat perhatian (E36)	.00	**.64**	-.05	-.05	.16
Tidak keberatan menjadi pusat perhatian (E41)	-.05	**.60**	.00	.05	.12
Lebih suka bekerja di belakang layar (E16)	.07	**.54**	-.07	-.02	-.02
Tidak banyak berbicara pada orang yang tidak dikenal (E46)	.09	**.53**	.01	.13	.06
Merasa nyaman berada di sekitar orang lain (E11)	.08	**.49**	.00	.30	-.05
Selalu mempersiapkan segala hal (C3)	.01	.01	**.70**	.06	.03
Menyukai keteraturan (C33)	-.14	-.03	**.69**	.03	.01
Melakukan aktivitas sesuai jadwal atau agenda (C43)	-.00	.01	**.68**	.08	.04
Segera mengerjakan tugas yang diberikan (C23)	.06	.03	**.65**	.10	-.02
Telaten dalam mengerjakan tugas (C48)	.10	-.09	**.63**	.17	.17
Memperhatikan hal-hal secara rinci (C13)	-.11	-.01	**.57**	.11	.26
Mengabaikan tugas-tugas saya (C38)	.21	.00	**.51**	.18	-.03
Meninggalkan barang pribadi di sembarang tempat (C8)	.23	-.04	**.47**	-.01	-.05
Sering lupa meletakkan barang kembali pada tempatnya (C28)	.35	-.01	**.47**	-.07	.11
Mengacaukan banyak hal (C18)	.31	.08	**.46**	.17	-.13
Peduli dengan orang lain (A7)	.00	.13	.11	**.70**	.01
Meluangkan waktu untuk orang lain (A37)	.02	.10	.09	**.65**	.08
Bersimpati dengan perasaan orang lain (A17)	-.02	.00	.14	**.64**	.10

Memahami perasaan orang lain (A42)	.03	.16	.15	**.64**	.12
Membuat orang lain merasa nyaman (A47)	.04	.20	.24	**.56**	.11
Tidak terlalu tertarik dengan kondisi orang lain (A32)	.09	.31	-.06	**.52**	.03
Bersikap kasar pada orang lain (A12)	.27	-.14	.24	**.47**	-.06
Tidak tertarik dengan masalah orang lain (A22)	.05	.20	-.18	**.43**	.00
Lemah lembut (A27)	.10	-.22	.25	**.41**	-.02
Tidak terlalu memedulikan orang lain (A2)	.03	-.09	.03	**.40**	.07
Memiliki banyak ide (I50)	.06	.08	.05	.15	**.74**
Memiliki imajinasi yang sangat kuat (I15)	-.15	.04	-.04	.17	**.67**
Tidak memiliki imajinasi yang baik (I30)	.19	.07	-.02	.15	**.61**
Memiliki ide-ide yang cemerlang (I25)	.06	.21	.12	.17	**.61**
Kesulitan memahami ide yang bersifat abstrak (I10)	.22	.00	-.03	-.07	**.55**
Tidak tertarik dengan ide-ide abstrak (I20)	.07	.00	-.03	.05	**.55**
Cepat dalam memahami sesuatu (I35)	.07	.09	.28	.13	**.51**
Menguasai banyak kosakata (I5)	-.10	.17	.13	-.10	**.49**
Menggunakan istilah-istilah yang sulit (I40)	-.13	.00	-.06	-.14	**.43**
Meluangkan waktu untuk merefleksikan berbagai hal (I45)	.06	-.08	.23	.23	**.34**

Based on the estimation of validity and reliability that had been done in study 1, the Indonesian adaptation of IPIP-BFM-50 had acceptable psychometric properties. Therefore, study 1 produced an output of a BIG FIVE personality traits measure that had been validated in Indonesian samples named IPIP-BFM-50 Indonesia.

3. Study 2: Development of short-scale IPIP-BFM-25 Indonesia

Study 2 was conducted to develop short-scale IPIP-BFM-25 Indonesia by selecting items that fit the criteria. In addition, study 2 was also conducted to perform initial validation of the scale that had been developed.

3.1. Method

Participants

The data used in study 2 was similar to the data used for estimating psychometric properties in study 1, namely data on a sample of 502 participants.

Materials

The material used in study 2 was IPIP-BFM-50 Indonesia. IPIP-BFM-50 Indonesia consisted of 50 items. Each item contains short phrases describing subjects with scale 1 - 5. IPIP-BFM-50 Indonesia had a fairly good alpha reliability ranging from .76 (Agreeableness) to . 86 (Emotional stability).

Procedures

Short-scale development would be done by selecting five items in each dimension. There were four stages in developing the IPIP-BFM-25. First, we selected items based on criteria. The criteria specified were: 1) The items had Aiken's V above .80. 2) The items had a high discrimination score. The discrimination score was derived from the loading factor minus the absolute average of cross-loading. 3) The selected items might produce a scale having reliability above .70. The next step after the item was selected: we estimated the overlap index between the short scale and the parent scale. In the last step, we estimated the short-scale alpha reliability and factorial validity.

3.2. Results

Item selection

The main criteria for selecting items was that the item must have Aiken's V above .80. The first stage selection showed that there were six items (I40. A27. C18. A32. C38. A22) having index V below .80. Thus, these items had no possibility of being selected. The further selection is based on discrimination score. The discrimination score was obtained from the factor analysis that was conducted in study 1 (see Table 2). The discrimination score was estimated by calculating the difference between its loading on its primary factor and the average of the absolute cross-loading. Suppose the item number ES39, "Have frequent mood swings", had a loading factor on the Emotional Stability dimension of .73 and the absolute average of cross-loading on the other four dimensions of .07. Thus the item had a discrimination score of .73 - .07 = .66. After we gained all of the discrimination score, five items in each dimension with the highest discrimination score were selected to be the items of IPIP-BFM-25.

Overlap between IPIP-BFM-50 and IPIP-BFM-25

To estimate the index of overlap between long-form and short-form of the scale. we correlated on total score of both scales using Pearson’s product moment correlation. The correlation matrix between IPIP-BFM-50 and IPIP-BFM-25 can be seen in Table 3.

Table 3. Correlation Matrix between IPIP-BFM-50 and IPIP-BFM-25 (n = 502)

	Extra_50	**Agree_50**	**Consc_50**	**Emot_50**	**Intel_50**
Extra_25	**.938****	.137**	.009	.140**	.181**
Agree_25	.245**	**.876****	.294**	.161**	.251**
Consc_25	-.015	.285**	**.890****	.110*	.158**
Emot _25	.157**	.169**	.224**	**.937****	.104*
Intel_25	.155**	.201**	.104*	.138**	**.896****

Note: Extra=Extraversion. Agree=Agreeableness. Cons=Conscientiousness. Emot=Emotional stability. Intel=Intellect

**Correlation is significant at the 0.01 level (2-tailed)

**Correlation is significant at the 0.05 level (2-tailed)

Table 3 showed that on the same dimension, the correlation coefficient ranged between .876 - .938. Thus, there was a very strong correlation between scores on IPIP-BFM-50 with IPIP-BFM-25 (Evans. 1996). The overlap between IPIP-BFM-50 and IPIP-BFM-25 was quite high, and it meant that IPIP-BFM-25 was able to represent IPIP-BFM-5.

We acknowledged that this correlation coefficient was overestimated, because in the total score of long-form there was also a score of the short-form. Thus the error or random variance of items on a short-form was also reproduced on a long-form. Nonetheless. this information was important enough to illustrate how alignment between the total scores on IPIP-BFM-50 and IPIP-BFM-25. A strong correlation indicated that IPIP-BFM-25 was capable of representing scores of the parent-scale IPIP-BFM-5.

Reliability of IPIP-BFM-25

Reliability was determined from the value of the coefficient of alpha reliability. Reliability of the scales ranges from .73 (Intellect) to .80 (Emotional staibility). All scales had alpha reliability above the criteria of .70 (Nunnally. 1978), so each scale had satisfactory reliability. A summary of the scale properties can be seen in Table 4.

Tabel 4. Item-total correlation and alpha reliability IPIP-BFM-25 Indonesia (n=502)

Scale	**Number of items**	**Item-total correlation**	**Alpha**	**SEM**
Extraversion	5	.456 – .654	.76	1.62
Agreeableness	5	.528 – .620	.78	1.24
Conscientiousness	5	.511 – .592	.77	1.51
Emotional stability	5	.550 – .677	.80	1.66
Intellect	5	.410 – .562	.73	1.54

Note: SEM = Standard Error of Measurement

Factorial Validity of IPIP-BFM-25

Factorial validity was estimated by factor analysis using the Principal Component Analysis method forced into five factors. The analysis was done by varimax rotation method to see wheter the item distribution was still appropriate as the dimension measured. The Kaiser-Meyer-Olkin (KMO) result to estimate sample adequacy was 0.798 and The Barlett Test of Sphericity was 3828.802 with $p < .01$.

The result of factor analysis showed that all items were distributed well according to the measured factors. All items had a high loading factor in the primary factor, ranging from .62 to .85. While loading factor on other factors (cross-loading) was relatively low, all of them below .3. This result indicated that the IPIP-BFM-25 had a satisfactory factorial validity. Based on the estimation of validity and reliability that had been done in study 2, the Indonesia short-form of IPIP-BFM-50 had an acceptable psychometric properties. Therefore, study 2 produced an output of a short measure of big five personality traits that had been validated in Indonesian samples named IPIP-BFM-25 Indonesia.

4. Study 3: Psychometric evaluation of IPIP-BFM-25 Indonesia on independent samples

Study 3 was performed to complement the psychometric property evaluation of IPIP-BFM-25 which was done in study 2 using different samples. This procedure was very important according to Smith, McCarthy & Anderson (2000).

4.1. Method

Participants

The participants used in study 3 were 501 people with 160 men (31.9%) and 341 women (68.1%). Subject ages ranged from 15 to 40 years with an average age of 19.17 years and a Standard Deviation of 3.25.

Materials

There were four scales used in this study. The first was IPIP-BFM-25. IPIP-BFM-25 which was generated from study 2 had its psychometric properties evaluated. Second was BFI. BFI was a big five personality scale that had been adapted into Indonesian by Ramdhani (2012). BFI was used for estimating convergent validity because IPIP-BFM and BFI measure the same construct. BFI consisted of 44 items with a scale of 1 - 5. The alpha reliability coefficient of BFI reliability

in Indonesian sample ranged between .70 - .79 (Ramdhani. 2012). To compare the pattern of external correlates, the scale of Positive Affect Negative Affect Schedule (PANAS) (Watson, Clark & Tellegen. 1988) and Satifaction with Life Scale (SWLS) (Diener et al.. 1985) were used. PANAS consisted of 20 items and was used to measure positive and negatif affect, while SWLS consisted of five items and was used to measure life satisfaction. PANAS and SWLS have been translated and validated earlier in the preliminary study. The dimensions of positive affect had Alpha reliability of .836; dimensions of negative affection had Alpha reliability of .846; and life satisfaction had Alpha reliability of .846.

Procedures

Each participant completed four scales, namely IPIP-BFM-25, BFI, PANAS and SWLS. Some of the participants completed the IPIP-BFM-25 again three weeks latter to estimate the test-retest reliability. There were four steps done in the process of data analysis. First, we estimated the reliability through two approaches, namely the internal consistency approach and test-retest approach. Second, we conducted a confirmatory factor analysis. Third, we estimated convergent validity by correlating IPIP-BFM-25 with BFI. Last, we compared the correlation pattern of IPIP-BFM-25 and BFI with external variables.

4.2. Results

Reliability of IPIP-BFM-25

Reliability of the scales ranges from 0.709 (Intellect) to 0.797 (Conscientiousness). All scales had alpha reliability above the criteria of 0.70 (Nunnally. 1978), so each scale had satisfactory reliability. The summary of the reliability estimates using internal consistency approach and test-retest approach is presented in Table 5.

Table 5. Item-total correlation and reliability of IPIP-BFM-25 Indonesia (n=501)

Scale	Number of items	Item-total correlation	Alpha Cronbach	SEM	Test-retest (3 weeks. n=119)
Extraversion	5	.483 – .704	.796	1.52	.768
Agreeableness	5	.443 – .591	.778	1.49	.714
Conscientiousness	5	.535 – .662	.797	1.42	.803
Emotional stability	5	.494 – .636	.788	1.63	.739
Intellect	5	.414 – .506	.709	1.51	.725

Note: SEM = Standard Error of Measurement

Confirmatory Factor Analysis (CFA)

Confirmatory factor analysis was conducted to find out whether the proposed model on a theoretical basis fits in size to the research data (Schumacker & Lomax. 2010). The model was analysed using the Maximum Likelihood method. The parameters used to test model fit were CFI and RMSEA. The expected CFI value was above 0.90 (Hooper. Coughlan. & Mullen. 2008). While the expected RMSEA value was below 0.08 (Hu & Bentler, 1999). A recent CFA of the IPIP-BFM-25 Indonesia did not show good model fit (chi-square / df = 924.022 / 265. CFI = 0.827. and RMSEA = 0.071). The RMSEA shows that the model fit with the data, while CFI indicated that the model did not fit. The results of this analysis were similar to some of the previous findings of the Mini-IPIP scale confirmatory factor analysis (Cooper, Smillie & Corr. 2010; Donnellan et al.. 2006; Topolewska, Skimina, Strus, Cieciuch & Rowiński. 2017; Wielkiewicz, 2015).

Convergent and Discriminant Validity

Convergent and discriminant validity was estimated by correlating IPIP-BFM-25 with BFI. Pearson correlation results between the total score of IPIP-BFM-25 and BFI in each dimension are shown in Table 6.

Table 6. Correlation between IPIP-BFM-25 and BFI

	BFI				
	Extra	Agree	Cons	Neu	Open
IPIP-BFM-25					
Extraversion	**.825****	.218**	.132**	-.280**	-.0049
Agreeableness		**.554****	.246**	-.0124**	.166**
Conscientiousness			**.767****	-.0102*	-.005
Emotional stability				**-.757****	-.054
Intellect					**.662****

Note: Extra = Extraversion. Agree = Agreeableness. Cons = Conscientiousness. Neu = Neuroticsm. Open = Openness
**Correlation is significant at the .01 level (2-tailed)
**Correlation is significant at the .05 level (2-tailed)

Convergent validity was shown on the diagonal of table 6 typed in bold. The correlations between IPIP-BFM-25 and BFI in each dimension were: Extraversion r = .825; Agreeableness r = .554; Conscientiousness r = .767; Emotional stability r = -.757 and Intellect r = .662. All these correlations were significant at the .01 level. As a note. the fourth factor of BFI was Neuroticism which was the opposite of Emotional stability, therefore the correlation was negative. The

absolute average of the convergent correlation was .713. These values fall into the strong correlation (Evans, 1996). Discriminant validity was indicated by correlation values between different dimensions that were not typed in bold. The value of the discriminant correlation coefficient ranges from .005 to .28. The absolute average of the discriminant correlation was .112. These values fall into very weak correlation (Evans. 1996). The convergent correlation coefficient was much stronger than the discriminant correlation coefficient. It demonstrated that IPIP-BFM-25 had a satisfactory construct validity.

Comparison of correlation pattern of IPIP-BFM-25 and BFI to external variables

Comparison of correlation pattern of IPIP-BFM-25 and BFI to external variables was conducted by correlating IPIP-BFM-25 and BFI with positive affect, negative affect, life satisfaction and age. This procedure was conducted to find out if the correlation coefficients of BFI and IPIP-BFM-25 were equivalent to the external variable. This pattern of external correlates is shown in Table 7.

Table 7. Correlation of IPIP-BFM-25 and BFI with external variables

Dimensions	Scale	Life satisfaction	Positive affect	Negative affect	Age
Extraversion	IPIP-BFM-25	.191**	.387**	-.193**	-.027
	BFI	.243**	.472**	-.248**	.026
Agreeableness	IPIP-BFM-25	.263**	.356**	-.096*	.163**
	BFI	.363**	.287**	-.379**	.105*
Conscientiousness	IPIP-BFM-25	.229**	.331**	-.156**	.126**
	BFI	.291**	.421**	-.302**	.160**
Emotional stability	IPIP-BFM-25	.204**	.197**	-.590**	.049
	BFI	-.318**	-.327**	.647**	-.074
Intellect	IPIP-BFM-25	.061	.216**	-.124**	.076
	BFI	.015	.224**	-.001	.102*

**Correlation is significant at the 0.01 level (2-tailed)
**Correlation is significant at the 0.05 level (2-tailed)

The results of the correlation analysis between IPIP-BFM-25 and BFI with positive affect, negative affect, life satisfaction and age showed that the correlation pattern accross instruments were similar. All statistical conclusions obtained using the IPIP-BFM-25 scale and the BFI were also the same, except on the Intellect dimension. In general, the patterns of external correlates of the IPIP-BFM-25 matched the patterns of external correlates of the BFI, albeit with a lower coefficient.

The analysis of correlations with external variables using IPIP-BFM-25 was also consistent with the findings of previous studies using longer instruments. The pattern of correlations between the five dimensions of the Big Five with positive affect, negative affect, life satisfaction and age had similarities with findings from previous studies (Allemand, Zimprich & Hendriks, 2008; González Gutiérrez, Jiménez, Hernández & Puente, 2005; Grant, Langan-Fox & Anglim, 2009; Lucas & Donnellan, 2009; Soto, John, Gosling & Potter, 2008; Steel, Schmidt & Shultz, 2008; Tanksale, 2015).

5. Discussion

This study aimed to develop a short scale from its parent scale IPIP-BFM-50. A series of studies conducted show that IPIP-BFM-25 possesses a psychometric property that is satisfactory and feasible to use. IPIP-BFM-25 has satisfactory reliability, both with the internal consistency approach and with the test-retest approach. IPIP-BFM-25 also has content validity, factorial validity, satisfying convergence validity. In addition, the patterns of external correlates of the IPIP-BFM-25 matched the patterns of external correlates of the BFI.

In general, the reliability coefficient of parent scale, i.e. IPIP-BFM-5 in this study is better when compared with the coefficient of reliability on another big five personality scale in Indonesian, namely BFI. In the Ramdhani (2012) study it was reported that the Alpha reliability of the scales ranged from .73 - .79, while the reliability of Alpha IPIP-BFM Indonesia on the sample of 502 people ranged from .762 to .862. According to subject comments when completing the scale, some items in BFI have terms that are less familiar for them, such as "asertif (assertive)", "estetik (esthetic)" and "supel (sociable)".

Although generally the procedure of short-scale development in this study replicates what has been done by Donnellan et al. (2006), there are two different things related to the selection of items and the number of items selected. In IPIP-BFM-25, item selection is based on index of Aiken's V and discrimination score. We expect that the selected item has a strong correlation with the dimensions measured and of independence with other dimensions. We also expect that the selected items is relevant and important for measuring the Big Five personality. In addition, the number of selected items in IPIP-BFM-25 is also more higher than the number of items in Mini-IPIP. As a consequence, the Alpha reliability coefficient of IPIP-BFM-25 is better when compared to Mini-IPIP, i.e. all scales have Alpha above 0.70.

Although each of the IPIP-BFM-25 scale has satisfactory reliability, which is above 0.70. but when compared to IPIP-BFM-50, the coefficient reliability on this short scale is relatively lower. This is a common thing and occurs in the majority of studies on the preparation of short-scale (Batinic, Wolff & Haupt, 2008; Donnellan et al., 2006; Fischbach & Moosbrugger, 2008; Topolewska et al., 2017). However, the decrease in reliability in this study is still within acceptable limits.

The correlation analysis with external variables using IPIP-BFM-25 is also consistent with the findings of previous studies using longer instruments. The results of this study indicate that Agreeableness and Conscientiousness correlate positively with age. It is in line with the findings of Allemand et al. (2008); Lucas dan Donnellan (2009); dan Soto et al. (2008). This study also shows that Extraversion has the strongest positive correlation with positive affect, whereas Emotional stability has the strongest negative correlation with negative affect. This is consistent with the findings of González Gutiérrez et al. (2005), Grant et al. (2009), Steel et al. (2008), Tanksale (2015). Meanwhile, Intellect dimension has the lowest correlation with the life satisfaction variable. This finding is also consistent with the findings of Grant et al. (2009), Steel et al. (2008), Tanksale (2015).

The confirmatory factor analysis between the proposed model and the data is not fit. These findings are not surprising since some of the previous studies on Big Five measurements have also been found to be misfit models (Cooper et al., 2010; Donnellan et al., 2006; Topolewska et al., 2017; Wielkiewicz, 2015). The personality model is something very complex and each item always has a relationship with at least two factors. Therefore, cross-loading restrictions to zero are absolutely unrealistic and make the model less fit (Cooper et al.. 2010). In this study, this is reflected in the modification indices suggested by the AMOS program, which is to connect some items with other dimensions that are not targets in order to obtain a fit model.

Finally, this study has as output a short-form of scale for measuring big five personality traits which has satisfactory validity and reliability and is well suited to the culture of Indonesian society. With the existence of an alternative short measure of big five personality trait in Indonesia, it is expected that research on personality may increase because it uses an instrument that is developed through standard procedures and has a satisfactory psychometric property. In addition, the characteristic of IPIP that is open-source also allows this scale to be used more easily, both for commercial and non-commercial purposes.

IPIP-BFM-25 can be a very useful instrument for academics. especially for researchers who want to conduct a study on the role of Big Five personality in relation to other psychological

aspects. IPIP-BFM-25 can provide an alternative instrument that is more efficient and can save time, costs and energy. In addition, IPIP-BFM-25 is also very useful for research involving many constructs. The existence of a scale with fewer items can also increase the motivation of respondents to engage in research in the field of psychology.

This research still has some limitations in terms of the quality of the scale produced as well as the sample. The participants of the study were selected only in the area of Yogyakarta. In addition, the IPIP-BFM-25 is only capable of measuring the main factors of personality, but is not capable of measuring a more narrow facet. Therefore, the IPIP-BFM-25 is only recommended for use under certain circumstances, such as: limited resources for research, research using personality as a control variable, preliminary studies or research through the Internet. It is advisable that users who want to asses a comprehensive personality diagnosis of individuals to use another instrument such as NEO-PI-R. But, other personality measures that have longer items, such as BFI (John & Srivastava, 1999) are also unable to measure the individual's personality at the facet level. In such a case, Smith et al. (2000) suggests conducting content analysis by correlating factor scores with facet scores to show that they represent a more narrow facet domain.

6. Conclusion

From a series of studies conducted, it appears that IPIP-BFM-25 has a satisfactory psychometric property. IPIP-BFM-25 has satisfactory validity and reliability. In addition, the patterns of external correlates of the IPIP-BFM-25 matched the patterns of external correlates of the BFI. Thus, IPIP-BFM-25 is feasible to be used for measuring the big five personality trait, especially for the purposes of psychological research. More specifically, researchers who involve the big five personality dimensions as the variables are suggested to use IPIP-BFM-25 when the resources is limited because it presents a more efficient and practical research instrument than any other existing scale.

Acknowledgement

Acknowledgment is directed to the Indonesian Endowment Fund for Education (LPDP) of the Ministry of Finance of the Republic of Indonesia who supported and funded this research.

References

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131–142, https://doi.org/10.1177/0013164485451012.

Allemand, M.. Zimprich, D., & Hendriks, A. A. J. (2008). Age differences in five personality domains across the life span. *Developmental Psychology*, 44(3), 758–770, https://doi.org/10.1037/0012-1649.44.3.758.

Azwar, S. (2017). *Metode Penelitian Psikologi Edisi 2*. Yogyakarta: Pustaka Pelajar.

Batinic, B., Wolff. H.-G., & Haupt, C. M. (2008). Construction and Factorial Structure of a Short Version of the Trendsetting Questionnaire (TDS-K). *European Journal of Psychological Assessment*, 24(2), 88–94, https://doi.org/10.1027/1015-5759.24.2.88.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, 25(24), 3186–3191.

Carvalho, L. de F., Nunes, M. F. O., Primi, R., & Nunes, C. H. S. da S. (2012). Unfavorable evidence for personality assessment with a 10-item instrument. *Paidéia* (Ribeirão Preto), 22(51), 63–71, https://doi.org/10.1590/S0103-863X2012000100008.

Cooper, A. J., Smillie, L. D., & Corr, P. J. (2010). A confirmatory factor analysis of the Mini-IPIP five-factor model personality scale. *Personality and Individual Differences*, 48(5), 688–691, https://doi.org/10.1016/j.paid.2010.01.004.

Costa Jr., P. T., & McCrae, R. R. (1995). Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. *Journal of Personality Assessment*, 64(1), 21–50.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49(1), 71–75.

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, https://doi.org/10.1037/1040-3590.18.2.192.

Evans, J. D. (1996). *Straightforward Statistics for the Behavioral Sciences*. Pacific Grove. CA: Brooks/Cole Publishing.

Fischbach, A., & Moosbrugger, H. (2008). 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, https://doi.org/10.1027/1015-5759.24.2.95.

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

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, https://doi.org/10.1016/j.jrp.2005.08.007.

González Gutiérrez, J. L., Jiménez, B. M., Hernández, E. G., & Puente, C. P. (2005). Personality and subjective well-being: big five correlates and demographic variables. *Personality and Individual Differences*, 38(7), 1561–1569, https://doi.org/10.1016/j.paid.2004.09.015.

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, https://doi.org/10.1016/S0092-6566(03)00046-1.

Grant, S., Langan-Fox, J., & Anglim, J. (2009). The Big Five Traits as Predictors of Subjective and Psychological Well-Being. *Psychological Reports*, 105(1), 205–231, https://doi.org/10.2466/PR0.105.1.205-231.

Herzberg, P. Y., & Brähler, E. (2006). Assessing the Big-Five Personality Domains via Short Forms. *European Journal of Psychological Assessment*, 22(3), 139–148, https://doi.org/10.1027/1015-5759.22.3.139.

International Test Commission (2016). *The ITC Guidelines for Translating and Adapting Tests* (Second edition). Retrieved from www.InTestCom.org.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, Measurement, and Theoretical Perspectives. in L. Pervin and O. P. John (Eds.), *Handbook of Personality: Theory and Research* (2nd ed.). New York: Guilford.

Joseph, S., Linley, P. A., Harwood, J., Lewis, C. A., & McCollam, P. (2004). Rapid assessment of well-being: The Short Depression-Happiness Scale (SDHS). *Psychology and Psychotherapy: Theory. Research and Practice*, 77(4), 463–478.

Lucas, R. E., & Donnellan, M. B. (2009). Age differences in personality: Evidence from a nationally representative Australian sample. *Developmental Psychology*, 45(5), 1353–1363, https://doi.org/10.1037/a0013914.

McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, 36(3), 587–596, https://doi.org/10.1016/S0191-8869(03)00118-1.

Mlačić, B., & Goldberg, L. R. (2007). An Analysis of a Cross-Cultural Personality Inventory: The IPIP Big-Five Factor Markers in Croatia. *Journal of Personality Assessment*, 88(2), 168–177, https://doi.org/10.1080/00223890701267993.

Nunnally, J. (1978). *Psychometric theory* (2nd Edition). New York: McGraw-Hill.

Oliveira, J. P. (2017). Psychometric Properties of the Portuguese Version of the Mini-IPIP five-Factor Model Personality Scale. *Current Psychology*, 1-8, https://doi.org/10.1007/s12144-017-9625-5.

Ramdhani, N. (2012). Adaptasi Bahasa dan Budaya dari Skala Kepribadian Big Five. *Jurnal Psikologi*, 39(2), 189–205.

Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1), 203–212, https://doi.org/10.1016/j.jrp.2006.02.001.

Robins, R. W., Tracy, J. L., Trzesniewski, K.., Potter, J., & Gosling, S. D. (2001). Personality Correlates of Self-Esteem. *Journal of Research in Personality*, 35(4), 463–482, https://doi.org/10.1006/jrpe.2001.2324.

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed). New York: Routledge.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, 12(1), 102–111, https://doi.org/10.1037//1040-3590.12.1.102.

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), 718–737, https://doi.org/10.1037/0022-3514.94.4.718.

Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin*, 134(1), 138–161, https://doi.org/10.1037/0033-2909.134.1.138.

Strus, W., Cieciuch, J., & Rowiński, T. (2017). The Polish adaptation of the IPIP-BFM-50 questionnaire for measuring five personality traits in the lexical approach. *Roczniki Psychologiczne/ Annals of Psychology*, 17(2), 347–366.

Tanksale, D. (2015). Big Five personality traits: Are they really important for the subjective well-being of Indians?: personality and subjective well-being. *International Journal of Psychology*, 50(1), 64–69, https://doi.org/10.1002/ijop.12060.

Topolewska, E., Skimina, E., Strus, W., Cieciuch, J., & Rowiński, T. (2017). The short IPIP-BFM-20 questionnaire for measuring the Big Five. *Roczniki Psychologiczne/ Annals of Psychology*, 17(2), 385–402.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070.

Wielkiewicz, R. M. (2015). *Confirmatory and exploratory factor analysis of the Mini-IPIP with a multi-institutional sample of first-year college students*. Retrieved from http://digitalcommons.csbsju.edu/psychology_pubs/93/.

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(7), 649–654. https://doi.org/10.1016/j.paid.2008.07.009.

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short Scales – Five Misunderstandings and Ways to Overcome Them. *Journal of Individual Differences*, 35(4), 185–189, https://doi.org/10.1027/1614-0001/a000148.