

# Psychometric Properties of the Arabic Big Five Personality Inventory (ABFPI)



**Ahmed M Abdel Khalek\***

*Department of Psychology – Faculty of Arts, Alexandria University, Egypt*

**Submission:** April 23, 2019; **Published:** May 10, 2019

**\*Corresponding author:** Ahmed M Abdel Khalek, Department of Psychology – Faculty of Arts, Alexandria University, Egypt

## Abstract

The Big Five personality factors have received much attention during the last few decades. The objectives of the current study on the Arabic Big Five Personality Inventory (ABFPI) were (a) to compute its alpha reliability, (b) to estimate its test-retest reliability, (c) to compute its criterion-related validity, (d) to report the inventory's descriptive statistics, (e) to examine the sex-related differences, and (f) to develop an English version of the scale. A sample of Egyptian undergraduates ( $N = 225$ ) took part in this study. Results indicated that Cronbach's alpha and test-retest reliabilities ranged between .70 and .83, indicating acceptable to good internal consistency and temporal stability. The criterion-related validity ranged from .49 to .86 ( $p < 0.1$  and above) against the NEO-FFI, indicating good validity. The lowest mean score was for neuroticism, whereas the highest was for agreeableness in men and women. The only statistically significant sex-related difference was for neuroticism in favor of women. It was concluded that the ABFPI has good psychometric properties. However, there is a need for further studies, particularly regarding its factorial structure.

**Keywords:** Arabic Big Five Personality Inventory; Neuroticism; Extraversion; Agreeableness, Openness; Conscientiousness; Reliability; Validity

## Introduction

Over the past few decades, the Big Five (BF) personality model emerged as the dominant framework for measuring personality traits. BF consist of neuroticism (N), extraversion (E), agreeableness (A), openness to experience (O), and conscientiousness (C). Consensus around the BF structure has grown steadily since the early 1990's. Goldberg [1] stated that the Five-Factor Model (FFM) reflects an emerging consensus about the general framework of a taxonomic representation of personality traits. Many personality psychologists agree that the BF captures the most important and basic individual differences in personality traits.

The exploration of the BF personality model was performed on the basis of five methods. They are as follows: (a) lexical studies [1-4], (b) self-rating and peer-rating scales, (c) correlations and factor analyses of personality questionnaires, (d) the observation of the actual behavior of study participants, and (e) the free self-description of personality [5]. McCrae & Costa [6] stated that trait-based research on personality premised on four assumptions about human nature: (a) personality traits exist and are measurable, (b) these traits vary across individuals, (c) the causes of human behavior are rooted within the individual (e.g., personality traits affect individual behavior), and (d) people can understand themselves and others (p. 161). McCrae & Costa [7]

proposed that "the traits of the Five-Factor Model are best viewed as explanations of an intermediate category of characteristic adaptations, which in turn provide explanations for behavior" (p. 247).

Personality traits are relatively stable across time and situations, though some allowance is made for change [8]. The BF personality structure has been replicated in a variety of languages and contexts [9,10]. McCrae et al. [11] analyzed personality peer-reports made by members of 50 cultures using translations of the NEO-PI-R. They found similar factor structures, age and gender differences across most cultures. They concluded that these results support the hypothesis indicating that features of personality traits are common to all human groups. As such, what would be the meaning and components of the BF?

Neuroticism refers to a lack of positive adjustment and emotional stability. It is associated with a wide array of negative emotions. Extraversion is associated with being active, energetic, gregarious, and assertive, as well as a tendency to experience positive emotionality. Conscientiousness is characterized by tendencies toward goal-directed, persistent, self-disciplined, careful, deliberate and organized behavior [12]. The agreeableness factor consists of the following sub-traits or facets: trust,

straightforwardness, altruism, compliance, modesty, and tender-mindedness. The sub-traits of the openness to experience factor are as follows: fantasy, aesthetics, feelings, actions, ideas, and values [13].

There is accumulating evidence that the BF influences a variety of different behaviors and life outcomes from academic achievement, motivation, political attitudes, work-family conflict, vocational selection, to psychopathology and health, among other variables. Therefore, the BF personality model is a fruitful avenue for research and applications.

The relationship between the BF and health is a prominent example for further research. Higher conscientiousness and lower neuroticism have been associated with better physical health [14]. Conscientiousness is a key determinant of better health behaviors, i.e., more exercise, a healthier diet, less substance use, and safer sex behaviors, better perceived health, and reduced mortality. On the other hand, high neuroticism is associated with negative health behaviors, i.e., less exercise, poor diet, greater substance abuse, worse perceived health, greater functional impairment, and multimorbidity [15].

### The Present Study

A wealth of inventories and scales to assess the BF personality factors is available [13,16-18], among other sources, particularly on the internet. In the present researcher's view, there are four problems in some of these questionnaires: (a) some are very long, (b) some are very short, (c) their long statements, and (d) their negatively-worded items.

First, concerning the long scales, some of the BF inventories contain a large number of items, up to 240 for example. A number of psychometrically oriented papers have been published investigating the length of scales. Merrens & Richards [19] studied the association between the length of personality inventories and their interpretation of generalized personality. They concluded that the short form was more favorably evaluated psychometrically. In three studies, Burisch [20] maintained that short scales are as valid, on the average, as long scales, even though some of the short scales are merely subsets of the long ones. Burisch [21] concluded that lengthening a scale beyond some point can actually weaken its validity. Furthermore, the present researcher has noticed a sharp difference between participants in the 1970's and those from the generation of the social media in the new millennium. The former participants were more patient, enthusiastic, zealous, intrinsically motivated, and willing to co-operate in responding to psychological tests and questionnaires. Since participants nowadays have low motivation, it is more suitable to keep away from long scales to avoid boredom, which leads to random responses.

Second, the very short scales (such as two items for each factor) may suffer from low reliability and weak representative sampling of any given personality trait. Third, some statements or questions in certain questionnaires are too long. Some participants have

limited memory to remember the whole statement or question. Therefore, it is preferable to use short items. Fourth, regarding the negatively worded items, it is well-known, based on actual observation in testing sessions, that a large portion of participants tend to have problems in understanding double negatives. Carver & Scheier [22] stated that "negatively worded items often turn out to be harder to understand or more complicated to answer than positively worded items" (p.47). Similarly, Schriesheim & Hill [23] concluded that negatively worded items impair response accuracy.

As a remedy to the problem of understanding double negatives, some researchers use negatively worded items (e.g., "I feel blue" in happiness scales) and then recode the response. Based on Baumeister et al.'s [24] paper entitled: *Bad is Stronger than Good*, this procedure is problematic as there is evidence that items describing negative emotions tend to evoke much stronger responses than items describing positive ones. People tend to underestimate the frequency of positive as opposed to negative affect. The previously mentioned authors concluded that "bad emotions generally produce more cognitive processing and have other effects on behavior that are stronger than positive emotions" (p.334). Further, reversely scored items may elicit response biases. For these reasons, the use of negatively worded items was minimized to two items in the new scale (ABFPI). Meanwhile, this scale is neither very long nor very short (i.e., 6 items for each factor) and the items are short (3 to 7 words).

In a previous paper, Abdel-Khalek [25] explained the first stage in developing the Arabic Big Five Personality Inventory (ABFPI). The final items in this inventory were based on a large item pool (455 item). One sample took part in the stage of selecting the items ( $N= 1,161$ ). Another sample ( $N = 450$ ) was used in the computation of the criterion-related validity of each item. The alpha reliabilities of the ABFPI were acceptable to good (.74 to .92) except for the agreeableness factor ( $r_{11} = .63$ ). Therefore, another cycle was carried out to change some items in this factor. A few items in the agreeableness factor have been changed to improve its internal consistency in the present study.

The aims of the present study regarding the ABFPI were (a) to compute its Cronbach's alpha reliability, (b) to estimate its test-retest reliability, (c) to compute its criterion-related validity at the level of the total items of each factor, (d) to report some preliminary results (descriptive statistics), (e) to examine the sex-related differences of the ABFPI, and (f) to develop an English version of the scale.

### Material and Method

#### Participants

A convenience sample of volunteer undergraduates ( $N= 225$ ) from Alexandria University, Egypt took part in this study (114 men; 111 women). Their ages ranged from 18 to 40 years. They were students from different departments and colleges. A sample of 150 students was used to compute alpha reliability, another of

35 students to estimate the test-retest reliability, and a third of 190 students to compute the criterion-related validity.

### The Arabic Big Five Personality Inventory (ABFPI)

The ABFPI consists of 30 short statements (six items for each factor). These items were selected from a large item pool (455 items). The selection of items was based on different steps: (a) the highest significant correlation coefficients with the rest of the items in the same factor, (b) an item-remainder correlation between .3 and .7, and (c) the highest correlations between each item of the present scale and the total score on the relevant factor in the NEO-FFI by Costa & McCrae [13].

Each item of the ABFPI is answered on a four-point Likert type scale: 1 (No), 2 (Some), 3 (Much), and 4 (Always). The total score in each factor could range from 6 to 24, with higher scores on the factor indicating a higher trait standing. Two items only must be recoded (No. 6 and 26). The ABFPI was intended to be used as a trait and not a state scale. Thus, the instructions include the term “in general”.

### Procedure

The last version of the ABFPI was administered anonymously to students in group testing sessions in their classrooms during university hours in the first semester of the academic year 2018-

2019. Students had volunteered for the study after the researchers briefly explained its purpose and assured them of anonymity. Two students in the Master of Psychology Program in Alexandria University carried out the testing. SPSS [26] was used for the statistical analysis of the data.

### Results

Table 1 presents Cronbach’s alpha, 7-day test-retest reliability, and the criterion-related validity of the inventory. As can be seen in Table 1, the alpha and test-retest reliabilities ranged from .70 to .83, indicating acceptable to good reliability. As for the criterion-related validity, the coefficients ranged from .49 to .86 ( $p < .01$  and above) against the NEO-FFI [13]. Table 2 sets out the descriptive statistics and the sex-related differences of the ABFPI.

**Table 1:** Alpha reliability (N = 150), test-retest reliability (N = 35), and the criterion-related validity (CRV; N = 190) of the ABFPI.

Factor	Alpha	Retest	CRV*
Neuroticism (N)	.78	.83	.86
Extraversion (E)	.79	.72	.83
Agreeableness (A)	.80	.70	.55
Openness (O)	.75	.82	.49
Conscientiousness (C)	.82	.83	.70

\* $p < .01$  and above.

**Table 2:** Mean (M), standard deviation (SD), t value, and effect size (d) of the ABFPI in men (n = 114) and women (n = 111).

Scale	Men		Women		t	p	d
	M	SD	M	SD			
Neuroticism	11.89	3.89	13.23	3.76	2.63	.009	.35*
Extraversion	16.32	4.24	16.50	3.70	0.36	-	-
Agreeableness	19.17	3.56	19.44	3.43	0.59	-	-
Openness	17.63	3.49	16.91	3.91	1.46	-	-
Conscientiousness	17.75	4.07	17.35	3.89	0.74	-	-

\*Small effect size.

The only statistically significant difference between men and women, as seen in Table 2, is in the neuroticism factor in favor of women and the effect size is small. However, there were no significant differences on the other factors. The ABFPI was developed at first in Arabic. Then, this version was translated into English. Several cycles of translation and back-translation were carried out with the help of bilingual psychologists. Professor David Lester checked and edited the English form of the inventory.

### Discussion

The Big Five personality factors have received much attention during the last few decades around the globe, including Arab countries. The general aim of the current research was to develop the Arabic Big-Five personality inventory (ABFPI) with the following purposes: (a) to avoid adding too many items, (b) to avoid using too few items, (c) to use short statements, and (d) to use the minimum number of negatively-worded items. The present

study successfully fulfilled these general aims, as well as other specific objectives, i.e., to estimate the psychometric parameters and to explore the sex-related differences in the ABFPI.

For the Cronbach’s alpha, the coefficients ranged between .75 and .82, indicating acceptable to good internal consistency of the five factors of the ABFPI. Regarding the test-retest reliability of this inventory, the coefficients ranged from .70 to .83, indicating acceptable to good temporal stability. Both types of reliability ranged between .70 and .83. Kline [27] and Nunnally [28] had suggested that reliabilities of .70 or higher are acceptable to validate research results. More recently, Furr [29] proposed that a reliability of .70 is desirable for research purposes. A scale with a reliability below .70 would mean that more than 30% of the scores contain random errors, which do not correlate with other scores on the same scale. Therefore, it is safe to conclude that the ABFPI enjoys acceptable to good reliabilities.

In the first stage of developing the ABFPI [25], the 20 items with the highest correlations with the remaining items in the same factor were retained. These 20 items were correlated (item by item) with the total score on the same factor of the NEO-FFI. Thus, the six items with the highest correlations were retained. In the present study, the correlations between the total score on each factor of the ABFPI and the total score on the same factor in the NEO-FFI were computed to estimate the criterion-related validity of the new inventory.

The criterion-related validity of the ABFPI against a gold-standard, i.e., the NEO-FFI [13], ranged from .49 to .86. All these validity coefficients were statistically significant ( $p < .01$  and above), indicating good validity of the five factors in the new inventory. The descriptive statistics were computed for the ABFPI. Since the five factors contain the same number of items (six), the comparison between the mean scores was feasible. The lowest mean score was for neuroticism, whereas the highest was for agreeableness in men and women. The only statistically significant difference between the sexes was on neuroticism in favor of women and the effect size was small. This finding is congruent with many other studies [30-35].

## Conclusion

The ABFPI has acceptable to good psychometric properties, i.e., internal consistency, test-retest reliability, and criterion-related validity against a gold standard, namely the NEO-FFI. In addition, this inventory has specific advantages: (a) its length is neither very short nor very long (i.e., 30 items) usually written on one page, (b) short-phrase items were used (from 3 to 7 words), and (c) the minimum use of opposite keying directions (2 items out of 30). The lowest mean score on the ABFPI was for the neuroticism factor, whereas the highest mean score was for agreeableness, among a sample of male and female undergraduates. The only statistically significant sex-related difference was on neuroticism in favor of women.

## Limitations

Notwithstanding the acceptable to good psychometric characteristics of the ABFPI, there are some limitations. Foremost among them is the sample. It was taken from one university, so the results cannot be generalized to a larger population in Egypt. Furthermore, this was a convenience sample not a probability one. A next step would be to recruit a probability sample with different age groups. There is also a need to explore the factorial structure of the inventory. In addition, the English version of the ABFPI merits a test on an English-speaking sample. These are projects for future studies.

## Acknowledgment

I would like to thank Professor David Lester, Richard Stockton University, New Jersey in the US, for his editing of the English version of the scale, Yomna Kamal, and Yosra Kamal

for their assistance in administering the scale, and Aya Ahmed Hasan for her assistance in the statistical analysis of the data along with Yosra and Yomna.

## References

1. Goldberg LR (1993) The structure of phenotypic personality traits. *Am Psychol* 48: 26-34.
2. Allport GW, Odbert HS (1936) Trait names: A psycho-lexical study. *Psychological Monograph* 47: 211.
3. Galton F (1884) Measurement of character. *Fortnightly Review* 36: 179-185.
4. John OP, Angleitner A, Ostendorf F (1988) The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality* 2: 171-203.
5. Digman JM (1990) Personality structure: Emergence of the five-factor model. *Annual Review of Psychology* 41: 417-440.
6. McCrae RR, Costa PT (2008) The Five-Factor theory of personality. In: OP John, RW Robins & LA Pervin (Eds.), *Handbook of personality: Theory and research*, Guilford: New York, United States, pp. 159-181.
7. McCrae RR, Costa PT (1995) Trait explanations in personality psychology. *European Journal of Personality* 9: 231-252.
8. Caspi A, Roberts BW, Shiner RL (2005) Personality development: Stability and change. *Annu Rev Psycho* 56: 453-484.
9. John OP, Naumann LP, Soto CJ (2008) Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In: OP John, RW Robins, LA Pervin (Eds.) *Handbook of personality: Theory and research*, Guilford: New York, United States, pp. 114-158.
10. McCrae RR, Allik J (2002) The five-factor model of personality across cultures. In: Springer Science + Business Media, New York, United States.
11. McCrae RR, Terracciano A (2005) Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology* 88: 547-561.
12. Brown S D, Hirschi A (2013) Personality, career development and occupational attainment. In: SD Brown, RW Lent (Eds.), *Career development and counseling: Putting theory and research to work* (2<sup>nd</sup> edn), Wiley, New Jersey, United States, pp. 299-328
13. Costa P T, McCrae R R (1992) Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO FFI): Professional Manual. In: Psychological Assessment Resources, Odessa, Florida.
14. Strickhouser JE, Zell E, Krizan Z (2017) Does personality predict health and well-being? A meta-analysis. *Health Psychol* 36: 797-810.
15. Rochefort C, Hoerger M, Turiano N A, Duberstein P (2018) Big Five personality and health in adults with and without cancer. *Journal of Health Psychology*
16. Buchanan T, Johnson JA, Goldberg LR (2005) Implementing a five-factor personality inventory for use on the internet. *European Journal of Psychological Assessment* 21: 116-128.
17. De Raad B, Perugini M (2002) Big Five assessment. In: Hogrefe & Huber, Seattle, Washington, USA.
18. Donnellan MB, Oswald FL, Baird BM, Lucas RE (2006) The mini scales: Tiny yet-effective measures of the big five factors of personality. *Psychol Assess* 18: 192-203.
19. Murrrens MR, Richards WS (1973) Length of personality inventory and the evaluation of a generalized personality interpretation. *J Pers Assess* 37: 83-85.



20. Burisch M (1984) You don't always get what you pay for: Measuring depression with short and simple versus long and sophisticated scales. *Journal of Research in Personality* 18: 81-98.
21. Burisch M (1997) Test length and validity revisited. *European Journal of Personality* 11: 303-315.
22. Carver CS, Scheier MF (2000) *Perspectives on personality*. In: (4<sup>th</sup> edn), Allyn & Baco, Boston, Massachusetts, United States.
23. Schriesheim CA, Hill KD (1981) Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement* 41: 1101-1114.
24. Baumeister RF, Bratslavsky E, Finkenauer C, Vohs K D (2001) Bad is stronger than good. *Review of General Psychology* 5: 323-370.
25. Abdel-Khalek AM (2018b) The Arabic Big Five Personality Inventory (ABFPI): Setting the stage. *Psychology and Behavioral Science International Journal* 9(4).
26. SPSS Inc. (2009) *SPSS: Statistical data analysis: Base 18.0, Users guide*. SPSS Inc: Chicago, Illinois, United States.
27. Kline P (1998) *The new psychometrics: Science, psychology, and psychometrics*. In: Routledge, London, United Kingdom.
28. Nunnally JC (1978) *Psychometric theory*. In: (2<sup>nd</sup> edn), Jossey-Bass, San Francisco, United States.
29. Furr RM (2011) *Scale construction and psychometrics for social and personality psychology*. Sage, Thousand Oaks: California, USA.
30. Abdel-Khalek A M (2013) Constructions of anxiety and dimensional personality model among college students. *Psychol Rep* 112: 992-1004.
31. Abdel-Khalek AM (2018) Sex differences in personality dimensions in an Egyptian sample. *Mankind Quarterly* 58: 588-598.
32. Abdel Khalek A, Eysenck SBG (1983) A cross-cultural study of personality: Egypt and England. *Research in Behavior and Personality* 3: 215-226.
33. Escorial S, Navas MJ (2007) Analysis of the gender variable in the Eysenck Personality Questionnaire-Revised Scales using differential item functioning techniques. *Educational and Psychological Measurement* 67: 990-1001.
34. Eysenck HJ, Eysenck SBG (1975) *Manual of the Eysenck Personality Questionnaire*. In: Hodder & Stoughton Educational, London, United Kingdom.
35. Lynn R, Martin T (1997) Gender differences in extraversion, neuroticism, and psychoticism in 37 nations. *J Soc Psychol* 137: 369-373.



This work is licensed under Creative Commons Attribution 4.0 License  
DOI: [10.19080/PBSIJ.2019.11.555820](https://doi.org/10.19080/PBSIJ.2019.11.555820)

### Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats  
( Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission

<https://juniperpublishers.com/online-submission.php>