

The Big Five Inventory (BFI)

Reliability and Validity of its Italian Translation in Three Independent Nonclinical Samples

Andrea Fossati, Serena Borroni, Donatella Marchione, and Cesare Maffei

School of Psychology, San Raffaele Vita-Salute University, Milano, Italy

Abstract. The internal consistency reliability, factor structure, and convergent-discriminant validity of the Italian translation of the Big Five Inventory (BFI) were assessed in two independent samples of nonclinical adult volunteers (Sample 1: $N = 500$; Sample 2: $N = 316$) and in one sample of adolescent volunteers (Sample 3: $N = 223$). Two adult subsamples ($n = 70$, and $n = 141$, respectively) also provided 2-month retest reliability data. The internal consistency reliabilities were adequate for all five BFI scales (mean α values were .77, .78, and .81 for Sample 1, Sample 2, and Sample 3, respectively); all test-retest correlations were greater than .75 in both adult participant subsamples. Principal component analyses showed that only the first five components of the BFI item correlation matrix could be reproduced safely across the three samples. The BFI scales showed adequate convergent-discriminant validity coefficients in all three samples. These findings suggest that the BFI is a succinct measure of the Big Five personality traits and it provides satisfactory reliability and validity data.

Keywords: personality, Big Five, reliability, validity

Introduction

After decades of activity, the field of personality research has widely adopted a general taxonomy that describes individual differences in human personality according to five higher-order factors (Matthews, Deary, & Whiteman, 2003) – the so-called “Big Five” personality dimensions.

The availability of this taxonomy has prompted personality research in a wide variety of different contexts, ranging from academic to clinical settings. In turn, these flourishing research efforts have promoted the development of short and even super-short Big Five measures (for a review, see Denissen, Geenen, van Aken, Gosling, & Potter, 2008), which could be used when assessment time is limited.

Specifically designed in the late 1980s (John, Donahue, & Kentle, 1991) as a short instrument to measure the Big Five dimensions, the Big Five Inventory (BFI) consists of 44 short-phrase items that can be answered in about 5 minutes; it represents a convenient tradeoff between domain extent of coverage and shortness. Indeed, although the BFI scales included only 8 to 10 items, they showed good psychometric properties. The BFI showed adequate reliability in adult nonclinical subjects (John & Srivastava, 1999) and yielded sound validity evidence on the basis of substantial convergent and divergent relations with other Big Five instruments and of peer ratings (John & Srivastava, 1999). Currently, reliability and validity data are available for several translations of the BFI

(Benet-Martínez & John, 1998; Denissen et al., 2008; Lang, Ludtke, & Asendorpf, 2001; Plaisant, Srivastava, Mendelsohn, Debray, & John, 2005). Interestingly, these findings closely replicated those that had been reported for the original version of the scale.

Recently, the BFI was translated into Italian by one of the authors (A.F.) in order, first, to provide a short, easy-to-administer, widely used instrument for the assessment of the Big Five traits: and, second, to allow comparison between Italian and European/North American findings deriving from the Five Factor model. Thus, the present study aimed to test reliability, factor validity, and convergent-discriminant validity of the BFI in two independent samples of Italian adult volunteers as well as in a sample of Italian high-school students. In two adult participant subsamples, we were also able to assess the 2-month retest reliability of the BFI scales.

Considering the BFI predictive validity, in Sample 2 (adults) we tried to test the hypothesis that suggests that a high score in Neuroticism characterizes Attachment anxiety and a low score in Extraversion characterizes Attachment avoidance. This hypothesis is based on recent behavior genetic data linking the two attachment styles to personality systems of approach (i.e., Extraversion) and avoidance (i.e., Neuroticism) (Donnellan, Burt, Levendovsky, & Klump, 2008). As further evidence of the BFI predictive validity, in line with consistent data on the predictive role of Big Five traits on work achievement (Barrick

& Mount, 1991; Hurtz & Donovan, 2000) and academic performance (Chamorro-Premuzic, Furnham, & Ackerman, 2006), we tested the hypothesis that in the high-school student sample the BFI Conscientiousness scores were positively correlated with the average school grade. Considering Chamorro-Premuzic et al.'s (2006) seminal paper, we expected a significant correlation between the BFI Openness to Experience scale and literary knowledge.

Method

Participants

Sample 1

The first sample was composed of 500 participants from the general population who had responded to advertisements requesting volunteers for psychological studies. Participants were administered the BFI and NEO-IPIP questionnaires after signing a written informed consent form to take part in the study. The questionnaires were filled in anonymously. 309 participants (61.8%) were female and 191 (38.2%) were male; their mean age was 34.4 years ($SD = 11.46$ years; range: 18–87 years). 11 participants (2.2%) had an elementary-school diploma, 72 (14.4%) had a middle-school diploma, 237 (47.4%) had a high-school diploma, and 180 (36.0%) had a university or master's degree. None of the participants was paid either directly or indirectly in order to participate in the study.

A subsample of 70 participants agreed to participate in the 2-month retest study: 27 participants (38.6%) were male and 43 (61.4%) were female; their mean age was 34.4 years ($SD = 11.5$ years).

Sample 2

The second sample was composed of 318 nonclinical adult volunteers who agreed to participate in the study after it was described in detail. Participants were administered the BFI and NEO-IPIP questionnaires after signing a written informed consent form to participate in the study. The questionnaires were filled in anonymously. In addition, Sample 2 participants were administered the Attachment Style Questionnaire (ASQ; Feeney, Noller, & Hanrahan, 1994). 213 participants (67.1%) were female, and 105 (32.9%) were male; participants' mean age was 31.4 years ($SD = 14.1$ years; range: 18–87 years). Three participants (0.9%) had an elementary-school diploma, 47 (14.8%) had a middle-school diploma, 211 (66.3%) had a high-school diploma, and 57 (18.0%) had a university or master's degree. Overall, Sample 2 participants were slightly younger and showed a lower frequency of university/master degree than Sample 1 participants. A subsample of 141 participants agreed to take part in the 2-month retest reliability study:

100 participants (70.9%) were female and 41 (29.1%) were male; their mean age was 30.5 years ($SD = 15.5$ years). None of the participants was paid either directly or indirectly in order to participate in the study.

Sample 3

The third sample was composed of 223 high-school students: 150 participants (67.3%) were female and 73 (32.7%) were male; participants' mean age was 17.0 years ($SD = 1.4$ years; range: 14–19 years). Student's average grades were 6.94 ($SD = 0.85$) for literary subjects and 6.69 ($SD = 1.14$) for scientific subjects; the overall mean grade, averaged over literary and scientific subjects, was 6.86 ($SD = 0.83$). All participants signed a written informed consent in order to take part in the study; in the case of minors, parents or caregivers were also asked to sign the written informed consent form. None of the participants was paid either directly or indirectly in order to participate in the study. The participants were administered the BFI and the NEO-IPIP in classroom sessions at the end of the school year. The questionnaires were filled in and scored anonymously. However, the participants were given an identification code that corresponded to their rank in the class register, and their questionnaires were identified by that code. Trained undergraduate psychology students administered the questionnaires; they knew both the participants' identities and their grades averaged across literary and scientific subjects of the current school year, but not the participants' questionnaire scores. On the other hand, the psychologists who scored the questionnaires and matched the codes were not given the participants' identity data. Obviously, teachers had no access to participants' questionnaire data.

Measures

Big Five Inventory

Although it stems from the lexical approach to the Big Five trait assessment, the BFI (John & Srivastava, 1999) consists of 44 short-phrase items, rated on a 5-point scale from *disagree strongly* to *agree strongly*. The items were selected using both consensual expert judgment and empirical item analyses to represent the core (i.e., most prototypical) traits that define each Big Five domain. The BFI items are assigned to five scales measuring Extraversion (E; 8 items), Agreeableness (A; 9 items), Conscientiousness (C; 9 items), Neuroticism (N; 8 items), and Openness to experience (O; 10 items).

Following Denissen and colleagues' (2008) suggestions, equivalence with the original meaning of the items was the guiding principle in the translation process. First, the BFI was translated into Italian by one of the authors (A.F.). The guiding principle was to respect the original meaning of the

items as per Denissen and colleagues' (2008) suggestions. Then the second author (S.B.) reviewed the translation independently. After reaching a consensus, we let an English mother-tongue professional translator translate the Italian version back into English, and this English back-translation was sent to the author of the BFI. If the latest version differed from the English original, the first author, the second author, and the professional translator come to an agreement on the definitive Italian translation.

NEO-IPIP

The NEO-IPIP questionnaire was developed on the basis of the International Personality Item Pool (IPIP; Goldberg, 1999; Johnson, 2005) to provide a representation of the domain constructs of the Five Factor Model, as expressed in Costa and McCrae's (1992) revised NEO personality inventory (NEO-PI-R). The NEO-IPIP items are brief and use familiar words and simple grammar; each item is rated on a 5-point scale. Unlike the BFI items, the NEO-IPIP items represent samples of domains of behavior rather than samples of domains of lexical terms. This warranted the absence of item overlap between the two NEO-IPIP and the BFI scales, which could spuriously inflate the scale correlations. The NEO-IPIP was designed to match the original NEO-PI-R structure in which five personality dimensions are described by 30 facets (Costa & McCrae, 1992). In the present study, we used the 120-item version of the NEO-IPIP, which provides 4 items for each facet and 24 items for each Big Five trait. The 120-item NEO-IPIP questionnaire was translated independently by the third author and by a clinical psychologist fluent in English. When they reached a consensus translation, it was checked through back-translation by an English mother-tongue professional translator.

Attachment Style Questionnaire (ASQ, Feeney, Noller, & Hanrahan, 1994)

The ASQ is a 40-item Likert-type, self-administered questionnaire designed to measure five dimensions of adult attachment: Confidence in self and others (8 items), Discomfort with closeness (10 items), Relationships as secondary (7 items), Need for approval (7 items), and Preoccupation with relationships (8 items). The ASQ also provides two second-order scales, which are called Attachment anxiety and Attachment avoidance. Both English (Feeney et al., 1994) and Italian (Fossati et al., 2003) versions of the ASQ showed adequate reliability and validity data. Cronbach's α in the current study ranged from .70 (Confidence and Preoccupation with relationships) to .84 (Relationships as secondary to achievement). The Cronbach's α values for Attachment anxiety and Attachment avoidance second-order scales were .81 and .83, respectively.

Results

Descriptive Statistics and Reliability Analyses

In Sample 1, weak, albeit significant, correlations with age were only observed for BFI A ($r = .14, p < .01$), C ($r = .16, p < .001$), and O ($r = -.15, p < .001$) scales; however, this finding was not replicated in Sample 2 and in the adolescent sample (Sample 3).

After Bonferroni correction of the nominal significance level, in Sample 1 only the BFI O scale was significantly associated with school level, $F(3, 496) = 10.11, p < .001, \eta^2 = .06$. In particular, Bonferroni posthoc contrasts showed that participants with a university or master's degree had significantly higher average scores on O than participants with elementary-school diplomas (Bonferroni $t = 4.70, p < .001$) and participants with middle-school diplomas (Bonferroni $t = 3.29, p < .002$). However, these associations between the BFI O scale and participants' school level was not replicated in Sample 2, in which no significant association between school level and BFI scale scores was observed.

Mean values for the BFI E, A, C, N, and O scales were, respectively: 26.41 ($SD = 5.79$), 34.99 ($SD = 4.19$), 35.82 ($SD = 5.47$), 23.29 ($SD = 5.99$), and 35.33 ($SD = 7.15$) in Sample 1; 25.12 ($SD = 5.88$), 31.43 ($SD = 5.91$), 33.36 ($SD = 6.71$), 25.12 ($SD = 5.50$), and 27.57 ($SD = 5.44$) in Sample 2; 27.78 ($SD = 6.73$), 33.23 ($SD = 6.02$), 30.94 ($SD = 7.09$), 24.82 ($SD = 6.19$), and 37.03 ($SD = 6.35$) in Sample 3. Reliability estimates were adequate and consistently replicated across the three samples. Mean Cronbach's α values were .77 (min. (A) = .71, max. (N and O) = .80) in Sample 1, .78 (min. (A) = .73, max. (C) = .85) in Sample 2, and .81 (min. (A) = .76, max. (E) = .86) in Sample 3.

The 2-month retest reliability estimates were adequate and similar in the two independent subsamples that took part in the retest reliability study. Despite the high attrition rate, no systematic differences were observed between test and retest subsamples in terms of BFI values at time 1 (T1). According to MANOVA results, subsample 2 did not show any significant difference on the five BFI scales from test subsample 2, Wilks $\Lambda = .98$, Pillai $V = .02, p > .40$; similar findings were observed for subsample 1.

In retest subsample 1 ($n = 70$), means at T1 ranged from 22.39 ($SD = 5.07$) for E to 36.09 ($SD = 5.24$) for C, and from T2 mean values for E, A, C, N, and O scales; means at T2 ranged from 22.09 ($SD = 5.19$) for E to 36.89 ($SD = 5.33$) for C. In the retest subsample 1 ($n = 70$), test-retest r values were .84, .87, .83, .82, and .86 (all $ps < .001$) for the BFI E, A, C, N, and O scales, respectively.

In retest subsample 2 ($n = 141$), means at T1 ranged from 21.39 ($SD = 5.05$) for E to 33.01 ($SD = 7.34$) for C, and from 21.78 ($SD = 4.85$) for E to 33.42 ($SD = 6.74$) for C at T2. In the retest subsample 2 ($n = 141$), test-retest r values were .85, .76, .90, .75, and .81 (all $ps < .001$) for the same five scales.

Table 1. Varimax-rotated Big Five Inventory item factor loadings in Sample 1 ($N = 500$)

BFI item (Scale)	Principal component loadings					h^2
	O	N	C	E	A	
1. Talkative (E)	.18	.03	-.04	.67	.08	.49
2. Find faults with others (A; R)	-.10	-.29	.07	-.07	.42	.29
3. Does a thorough job (C)	.04	.08	.67	.04	.18	.49
4. Depressed, blue (N)	-.14	.55	-.10	-.24	-.03	.39
6. Reserved (E)	-.08	-.06	-.20	.56	-.08	.38
7. Helpful and unselfish (A)	.27	.10	.17	.04	.47	.33
8. Careless (C; R)	-.09	-.20	.35*	-.25	.35	.35
9. Relaxed (N; R)	-.15	.62	-.14	-.12	-.09	.45
10. Curious (O)	.57	-.12	.10	.16	.09	.38
11. Full of energy (E)	.29	-.16	.32	.40	.08	.38
12. Starts quarrels (A; R)	-.04	-.37	.20	-.15	.45	.40
13. Reliable (C)	.07	.03	.61	-.06	.17	.41
14. Tense (N)	-.07	.66	-.02	-.12	.04	.45
15. Ingenious (O)	.70	-.01	.17	.08	-.12	.54
16. Generates enthusiasm (E)	.46	-.07	.10	.51	.15	.51
17. Forgiving nature (A)	.03	-.03	.01	.16	.60	.38
18. Disorganized (C; R)	-.12	-.23	.59	-.12	-.04	.43
19. Worries a lot (N)	-.13	.56	.07	-.08	.27	.41
20. Active imagination (O)	.66	.03	-.05	.06	.03	.45
21. Quite (E; R)	-.02	-.11	.03	.72	.07	.54
22. Trusting (A)	.01	.03	-.09	.09	.40	.18
23. Lazy (C; R)	-.09	-.23	.52	.26	-.14	.42
24. Emotionally stable (N; R)	-.09	.57	-.17	-.02	-.09	.38
25. Inventive (O)	.71	-.14	.15	.22	-.10	.61
26. Assertive (E)	.42	-.11	.30	.50	-.02	.53
27. Cold and aloof (A; R)	-.07	-.10	-.05	.30	.34	.22
28. Perseveres (C)	.28	.02	.64	.03	.07	.50
29. Moody (N)	.06	.65	-.06	-.14	.03	.45
30. Values artistic experiences (O)	.60	.12	-.06	-.07	.27	.45
31. Shy, inhibited (E; R)	-.05	-.37	.00	.57	-.13	.48
33. Does things efficiently (C)	.10	-.07	.69	.06	.17	.52
34. Calm (N; R)	-.23	.51	-.13	.03	-.24	.39
35. Prefers routine (O; R)	.32	-.12	.01	.10	-.20	.16
36. Outgoing, sociable (E)	.32	-.01	.13	.67	.21	.61
38. Makes plans (C)	.21	-.07	.62	.09	-.01	.44
39. Nervous (N)	-.10	.73	.04	.06	-.27	.61
40. Likes to reflect (O)	.62	-.08	.09	-.01	.09	.40
41. Few artistic interests (O; R)	.50	-.10	-.06	-.16	.07	.29
42. Likes cooperate (A; R)	.33	.02	.19	.26	.49	.45
43. Distracted (C; R)	-.17	-.28	.53	-.12	-.01	.40
44. Sophisticated in arts (O)	.65	-.12	-.01	-.03	.21	.48
Eigenvalue	7.26	3.82	3.17	2.37	2.20	

Note. Principal components are listed in order of extraction; BFI items are abridged. *: item with its highest loading on an unexpected factor; R: Reverse item; h^2 : Communality.

BFI Item Principal Component Analyses

Overall, the Kaiser-Meyer-Olkin's measure of sampling adequacy (MSA) values were .85, .84, and .79 for Sample 1, Sample 2, and Sample 3, respectively. In Sample 1, the median MSA value for the individual BFI items was .83, and 95.5% ($n = 42$) of the BFI items had MSA values equal to, or greater than .70; the lowest MSA value was .60. Similar MSA values were observed in Sample 2 (median = .83, percentage of MSA values $\geq .70 = 93.2$, min. = .64) and in Sample 3 (median = .78, percentage of MSA values $\geq .70 = 93.2$, min. = .66). This suggested that PCA could be conducted without having to remove any unsuitable items.

In all samples, the minimum average partial statistic (Zwick & Velicer, 1986) reached its minimum value when six PC were extracted, thus suggesting a 6-factor solution for the BFI item correlation matrices. Similar indications also came from quasi-inferential parallel analyses (Buja & Eyuboglu, 1992) and scree plot evaluations¹. When dimensionality analyses were carried out on ipsatized data in order to control for acquiescence effect (Soto, John, Gosling, & Potter, 2008), all methods suggested a 6-factor solution.

Where the taxonomic view of factor analysis is concerned, there are considerable differences between demonstrating that a data set contains factors and that we have extracted them from the data accurately and reliably. In this case, it is important to show that the same factor procedures applied to different sets of respondents provide factors that are stable across the different sets of respondents (Everett, 1983). The coefficient of comparability provides a suitable factor stability test (Everett, 1983). The coefficient of comparability is obtained by subjecting each set of respondents to identical factor analyses; the two sets of factor score coefficients are used to calculate two sets of factor scores for each respondent; and the cross-correlation coefficients between these duplicate factor scores provides the comparability coefficient. Comparability and congruence coefficients tend to show close agreement. Only factors with comparability coefficients greater than .90 should be retained for further analysis (Everett, 1983; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996).

When we attempted a six-component solution, the sixth component could not be replicated across Sample 1 and Sample 2 (comparability coefficient = -.60), across Sample 1 and Sample 3 (comparability coefficient = .62), and across Sample 2 and Sample 3 (comparability coefficient = .59). In other words, the sixth component seemed to capitalize on sample characteristics rather than representing a substantive component. On the other hand, the five-component solution of the BFI item correlation

matrix yielded varimax-rotated PCs that closely matched across the three samples, as indicated by comparability coefficients ranging from .91 to .97 for the comparisons between Sample 1 and Sample 2, from .92 to .94 for the comparisons between Sample 1 and Sample 3, and from .90 to .95 for the comparisons between Sample 2 solution and Sample 3 solution. Thus, only the first five components represented a reliable and accurate representation of the data. Although it did not represent an acquiescence effect, the sixth principal component proved to be a difficulty factor (i.e., a method factor) that could not be replicated.

Thus, both empirical and theoretical issues suggested that no more than five principal components should be retained for further analyses in the three samples. The first five PC explained 42.8%, 45.4%, and 47.1% of the BFI item variance in Sample 1, Sample 2, and Sample 3, respectively.

Varimax-rotated principal component loadings are in Table 1. Since the findings were almost identical in terms of ease of presentation in all three samples, only Sample 1 findings are displayed². In Sample 1, 97.7% of the BFI items showed their largest loadings on their expected component; this result was replicated in Sample 2 and in Sample 3. When the BFI item factor structure obtained in Sample 1 was compared to that of Sample 3 (i.e., the adolescent sample), factor CC values ranged from .91 (Agreeableness) to .94 (Conscientiousness and Neuroticism). Similar findings were observed for the comparisons between Sample 2 and Sample 3 (factor CC values ranged from .90 in the case of Openness to Experience to .95 in the case of Extraversion). As expected, the BFI item factor structure was largely replicated across Samples 1 and 2 (i.e., the adult samples), with CC values ranging from .91 (Agreeableness) to .97 (Extraversion).

On average, the factor structure of the Italian version of the BFI, obtained in our three independent samples, closely matched the factor structure of the Dutch version (Denissen et al., 2008) of the BFI, as indicated by median CC values of .90 (min. [Sample 3] = .89, max. [Sample 1] = .94), .90 (min. [Sample 1] = .86, max. [Sample 2 and Sample 3] = .90), .91 (min. [Sample 3] = .91, max. [Sample 1] = .94), .93 (min. [Sample 2] = .90, max. [Sample 3] = .94), and .91 (min. [Sample 2 and Sample 3] = .91, max. [Sample 1] = .92), for E, A, C, N, and O scales, respectively. The factor structure of the Italian version of the BFI also showed an adequate correspondence to the Spanish version of the BFI (Benet-Martínez & John, 1998), at least for the scales E (median CC = .94), A (median CC = .86), C (median CC = .90), and N (median CC = .91). Consistent with the findings of Denissen and colleagues (2008), the O scales showed a lower replicability than the other four scales (median CC = .78). Similar

¹ A summary table of MAP statistics and quasi-inferential parallel analysis results is available upon request from the authors.

² Sample 2 and Sample 3 tables are available upon request from the authors.

Table 2. Convergent and discriminant correlations of the Big Five Inventory and NEO-IPIP scales

Sample 1 (<i>N</i> = 500)	1	2	3	4	5	6	7	8	9	10
1. E-BFI										
2. A-BFI	.20									
3. C-BFI	.14	0.31								
4. N-BFI	-.31	-.26	-.27							
5. O-BFI	.39	.19	.13	-.24						
6. E-NEO	.71	.14	.18	-.32	.44					
7. A-NEO	.03	.51	.11	-.04	.15	.04				
8. C-NEO	.03	.25	.71	-.26	.15	.13	.20			
9. N-NEO	-.29	-.18	-.25	.70	-.17	0.28	-.01	-.29		
10. O-NEO	.30	.16	-.05	-.14	.67	.40	.28	.03	-.10	
Sample 2 (<i>N</i> = 318)	1	2	3	4	5	6	7	8	9	10
1. E-BFI										
2. A-BFI	.20									
3. C-BFI	.09	.33								
4. N-BFI	-.16	-.04	-.01							
5. O-BFI	.35	.37	.33	-.18						
6. E-NEO	.71	.32	.29	-.25	.54					
7. A-NEO	.07	.71	.39	.05	.33	.23				
8. C-NEO	.06	.38	.82	-.07	.32	.25	0.45			
9. N-NEO	-.20	-.13	-.17	.73	-.16	0.28	-.01	-.27		
10. O-NEO	.33	.31	.17	-.13	.73	.48	.37	.18	-.13	
Sample 3 (<i>N</i> = 223)	1	2	3	4	5	6	7	8	9	10
1. E-BFI										
2. A-BFI	.02									
3. C-BFI	-.06	.13								
4. N-BFI	-.20	-.26	-.07							
5. O-BFI	.04	.06	-.02	.02						
6. E-NEO	.56	.06	.12	-.22	.17					
7. A-NEO	-.06	.74	.04	-.05	.09	.06				
8. C-NEO	-.10	.14	.81	-.04	0.06	-.02	.11			
9. N-NEO	-.21	-.20	-.20	.77	.00	-.31	-.02	-.19		
10. O-NEO	.04	.14	-.08	.13	.69	.11	.28	-.06	.15	

Note. **Bold** highlights convergent validity correlations; correlation coefficients greater in absolute value than .16, .20, and .24 are significant at Bonferroni corrected *p*-level (i.e., *p* < .00037) in Sample 1, Sample 2, and Sample 3, respectively.

findings were obtained for the comparison between the factor structures of the Italian and English versions of the BFI, as indicated by median CC values computed across the three Italian samples of .92, .84, .91, .92, and .75 for the scales E, A, C, N, and O, respectively.

Convergent and Discriminant Validity Analyses

In the three samples, the variance-covariance matrices of the BFI and NEO-IPIP scales were not significantly different in female and male subsamples, as indicated by Box's *M* values ranging from 60.71 (Sample 1) to 73.51 (Sample

2), all *p* values > .05. The correlation matrices of the BFI and NEO-IPIP scales are listed in Table 2. All convergent validity coefficients were substantial and significant, and exceeded all the discriminant validity coefficients. This seemed to suggest adequate convergent and discriminant validity of the BFI scale, at least with respect to the NEO-IPIP scales.

BFI Predictive Validity

In line with our hypotheses, when Sample 2 participants' age and gender effects were held constant, ASQ Avoidance scale showed its largest, negative correlation to the BFI Agreeableness scale, partial *r* = -.54, *p* < .001; attach-

ment Avoidance was also negatively correlated with the BFI E (partial $r = -.38, p < .001$), O (partial $r = -.30, p < .001$), and C (partial $r = -.18, p < .005$) scales. As expected, Anxiety attachment dimension was positively correlated to the BFI N scale ($r = .38, p < .001$) and negatively correlated to both C ($r = -.24, p < .001$), and E ($r = -.22, p < .001$) scales.

In Sample 3, the literary and scientific subjects of the current school year were averaged for each participant. Means for the average grades in literary, scientific, and general school subjects were 6.9 ($SD = 0.84$), 6.7 ($SD = 1.14$), and 6.9 ($SD = 0.82$), respectively. As expected, when participants' age and gender effects were held constant, BFI C scores were consistently associated with school performance. In particular, BFI C scale scores were significantly associated with overall average grades (partial $r = .44, p < .001$), scientific subject average grades (partial $r = .43, p < .001$), and literary subject average grades (partial $r = .33, p < .001$). Literary subject average grades were also significantly correlated to BFI O scale scores (partial $r = .18, p < .01$).

Discussion

Overall, our results strongly support the reliability and construct validity of the BFI in its Italian translation. Our results seem to suggest that the BFI may be useful in assessing basic personality traits in both adult and adolescent participants.

Although they were slightly lower than those previously reported in the literature (Benet-Martínez & John, 1998; Denissen et al., 2008; John & Srivastava, 1999), internal consistency reliability values were also adequate for the Italian translation of the BFI, particularly if we consider that the BFI scales are composed of a small number of items (ranging from 8 to 10 in the different scales). Despite the high attrition rate of the retest samples, test-retest correlations were also large, positive, and significant, suggesting at least a satisfactory 2-month temporal stability of the BFI scale scores. Moreover, these values were similar to those reported by John and Srivastava (1999) for the English version of the BFI.

Dimensionality and principal component analyses of the BFI item correlation matrices strongly supported previous cross-cultural observations suggesting that BFI item correlations are adequately explained by the first five factors, with any extra factors condensing only spurious variance (Soto et al., 2008; Worrell & Cross, 2004). Although Soto and colleagues (2008) suggested that it may be difficult to replicate the factor structure of the BFI among preadolescents and adolescents because of acquiescence effects, in this study the five-factor structure of the BFI items was consistently replicated not only in the two adult participant samples, but also in the adolescent participant sample, with CC values suggesting

closely matching factors. This result was consistent with Denissen and colleagues' (2008) findings on the stability of the BFI factor structure across different age groups. However, it should be observed that in our study, adolescent participants were in the 14–19-year age range, for which the CC values reported in Soto and colleagues' (2008) study were greater than .90 for all the BFI scales and very close to the values observed in our study. The amount of the BFI item variance explained by the five-factor solution was highly similar in the three samples that took part in this study and consistent with previously reported values (Worrell & Cross, 2004).

The factor structure of the Italian version of the BFI was highly similar to the factor structure of the Dutch version of the questionnaire (Denissen et al., 2008), with average CC values approximating or equaling the .90 benchmark indicating close correspondence. Thus, the comparisons between the factor structure of Italian translation of the BFI and published results for the English original and Spanish translation closely replicated the findings of Denissen and colleagues (2008). Furthermore, in the case of the Italian version of the BFI, CC values suggested overall similarity in factor structure with both the English (Benet-Martínez & John, 1998) and Spanish (Benet-Martínez & John, 1998) versions of the BFI. Consistent with Denissen and colleagues' (2008) findings, in our study only Openness to Experience showed some dissimilarity.

Consistent with previous findings on other translations of the BFI (Benet-Martínez & John, 1998; Denissen et al., 2008), the vast majority of the items in the Italian version of the BFI behaved as expected according to the BFI empirical model of item assignment to scales.

Convergent-discriminant validity results confirmed and extended John and Srivastava's (1999) observations on the construct validity of the BFI. In our study, the scales of the Italian translation of the BFI showed adequate convergent and discriminant validity both in adult and adolescent participant samples, at least with respect to the NEO-IPIP scales. Although the E and O dimensions showed moderate within-questionnaire (e.g., NEO-IPIP E with NEO-IPIP O, and BFI E with BFI O) and between-questionnaire correlations (e.g., NEO-IPIP E with BFI O, and BFI E with NEO-IPIP O), the convergent validity coefficients of the corresponding BFI scales were always substantially larger than these correlations. Overall, these findings seem to support the construct validity of the BFI in its Italian translation.

External validity for the Italian translation of the BFI data was somewhat limited, but in our opinion still interesting. The negative correlation that was observed in our study between attachment Avoidance and E, and the positive association between N and attachment Anxiety were highly consistent with Donnellan and colleagues' (2008) findings. Confirming and extending Nofle and Shaver's (2006) findings, our data showed that attachment Avoidance was negatively correlated to Agreeableness, Extra-

version, and Conscientiousness in adult subjects. Our results were also highly consistent with recent observations suggesting that attachment Anxiety may be positively associated with the BFI N and negatively associated with the BFI C scale (Nofle & Shaver, 2006). The negative correlation between the BFI E scale and attachment Anxiety closely replicated previous findings (Griffin & Bartholomew, 1994; Shaver et al., 1996).

Finally, our findings on the relationship between the BFI scales and school performance in Sample 3 extended the considerations on the role of Big Five traits on work (Barrick & Mount, 1991; Hurtz & Donovan, 2000) and academic (Chamorro-Premuzic et al., 2006) achievement based on adult data to high-school students. In our study, the BFI C scale showed significant positive correlations to both scientific and literary average grades of the high-school students, at least with respect to the current year grades. Consistent with Chamorro-Premuzic et al.'s (2006) findings, literary interests and abilities reflected by average literary grades were significantly associated with O personality dimension. Although it is possible that school performance may influence the responses to the BFI items, these results seem to suggest that the BFI may be a useful instrument for planning vocational and career development interventions with high-school students.

Obviously, the results of this study should be considered in the light of several limitations. Retest subsamples were plagued by very high attrition rates, which limit the relevance of our findings. In our opinion, the high attrition rates could be explained by two factors: the participants' unwillingness to have to be identified in order for the BFI to be administered a second time, and the difficulty in contacting more than once participants who were active community workers. With the exception of school grades, all data came from the same sources (i.e., self-report questionnaires), and we were unable to obtain independent measures of behavior. The three samples of participants who took part in this study were of moderate size and included exclusively nonclinical volunteers. Thus, they represented convenient study groups rather than samples, greatly limiting the generalizability of our findings. The convergent and discriminant validity data were based only on the NEO-IPIP: Further studies based on multiple instruments that imply really different methods (say, peer evaluations, observer data, etc.) of assessment of the Big Five traits are needed before our findings can be accepted. External validity data were based only on a very limited number of external constructs: Future studies will have to address the criterion validity of the BFI considering a wider range of variables (for instance, maladaptive personality aspects) than those taken into account in the present study.

Notwithstanding these limitations, the results of the present study seem to support the reliability and construct validity of the BFI as a measure of the Big Five traits in its Italian translation.

References

- Barrick, M.R., & Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.
- Benet-Martínez, V., & John, O.P. (1998). Los cinco grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75, 729–750.
- Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, 27, 509–540.
- Chamorro-Premuzic, T., Furnham, A., & Ackerman, P.L. (2006). Incremental validity of the typical intellectual engagement scale as predictor of different academic performance measures. *Journal of Personality Assessment*, 87, 261–268.
- 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.
- Denissen, J.J., Geenen, R., van Aken, M.A., Gosling, S.D., & Potter, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment*, 90, 152–157.
- Donnellan, M.B., Burt, S.A., Levendosky, A.A., & Klump, K.L. (2008). Genes, personality, and attachment in adults: A multivariate behavioral genetic analysis. *Personality and Social Psychology Bulletin*, 34, 3–16.
- Everett, J.E. (1983). Factor comparability as a means of determining the number of factors and their rotation. *Multivariate Behavioral Research*, 18, 197–218.
- Feeney, J.A., Noller, P., & Hanrahan, M. (1994). Assessing adult attachment. In M.B. Sperling & W.H. Berman (Eds.), *Attachment in adults: Clinical and developmental perspective* (pp. 128–152). New York: Guilford.
- Fossati, A., Feeney, J.A., Donati, D., Donini, M., Novella, L., Bagnato, M. et al. (2003). On the dimensionality of the Attachment Style Questionnaire in Italian clinical and nonclinical subjects. *Journal of Social and Personal Relationships*, 20, 55–79.
- Goldberg, L.R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.
- Griffin, D., & Bartholomew, K. (1994). Models of self and other: Fundamental dimensions underlying measures of adult attachment. *Journal of Personality and Social Psychology*, 52, 511–524.
- Hurtz, G.M., & Donovan, J.J. (2000). Personality and job performance: The Big Five revisited. *Journal of Applied Psychology*, 85, 869–879.
- John, O.P., Donahue, E.M., & Kentle, R.L. (1991). *The Big Five Inventory – Version 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.
- 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.
- Johnson, J.A. (2005). Ascertaining the validity of individual pro-

- protocols from web-based personality inventories. *European Journal of Personality*, 39, 103–129.
- Lang, F.R., Ludtke, O., & Asendorpf, J.B. (2001). Validity and psychometric equivalence of the German Version of the Big Five Inventory in young, middle-aged and old adults. *Diagnostica*, 47, 111–121.
- Matthews, G., Deary, I.J., & Whiteman, M.C. (2003). *Personality traits*. Cambridge: University Press.
- McCrae, R.R., Zonderman, A.B., Costa, P.T., Bond, M.H., & Paunonen, S.V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, 70, 552–566.
- Noftle, E.E., & Shaver, P.R. (2006). Attachment dimensions and the big five personality traits: Associations and comparative ability to predict relationship quality. *Journal of Research in Personality*, 40, 179–208.
- Plaisant, O., Srivastava, S., Mendelsohn, G.A., Debray, Q., & John, O.P. (2005). Relations between the French version of the Big Five Inventory and the DSM classification in a French clinical sample of psychiatric disorders. *Annales Medico-Psychologiques*, 163, 161–167.
- Shaver, P.R., Papalia, D., Clark, C.L., Koski, L.R., Tidwell, M.C., & Nalbone, D. (1996). Androgyny and attachment security: Two related models of optimal personality. *Personality and Social Psychology Bulletin*, 22, 582–597.
- Soto, J.C., 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, 718–737.
- Worrell, F.C., & Cross Jr., W.E. (2004). The reliability and validity of big five inventory scores with African American college students. *Journal of Multicultural Counseling and Development*, 32, 18–32.
- Zwick, W.R., & Velicer, W.F. (1986). A comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432–442.

Andrea Fossati

Servizio di Psicologia Clinica e Psicoterapia
San Raffaele Turro
via Stamira d'Ancona 20
I-20127 Milano
Italy
Tel. +39 02 26433241
Fax +39 02 26433408
E-mail fossati.andrea@hsr.it