



PERGAMON

Personality and Individual Differences 36 (2004) 587–596

PERSONALITY AND
INDIVIDUAL DIFFERENCES

www.elsevier.com/locate/paid

A contemplated revision of the NEO Five-Factor Inventory

Robert R. McCrae*, Paul T. Costa Jr.

Department of Health and Human Services, National Institute on Aging, NIH, Baltimore, Maryland, USA

Received 23 April 2002; received in revised form 6 January 2003; accepted 21 February 2003

Abstract

Previous item factor analyses and readability analyses suggested that 14 of the 60 items in the NEO Five-Factor Inventory might usefully be replaced. New analyses in high school ($N=1959$) and adult ($N=1492$) samples led to the selection of new items from the remaining pool of Revised NEO Personality Inventory items. The resulting scales showed modest improvements in reliability and factor structure, and equivalent validity. These new scales should be appropriate for most respondents age 14 and up. However, continued use of the current instrument is also reasonable for most applications.

Published by Elsevier Ltd.

Keywords: Item analysis; Scale development; Factor analysis; Five-Factor Model; Adolescents

1. Introduction

The 60-item NEO Five-Factor Inventory (NEO-FFI) was developed to provide a concise measure of the five basic personality factors (Costa & McCrae, 1989). For each scale, 12 items were selected from the pool of 180 NEO Personality Inventory (NEO-PI) items, chiefly on the basis of their correlations with validimax factor scores (McCrae & Costa, 1989). The instrument uses a five-point Likert response format. Two-week retest reliability is uniformly high, ranging from 0.86 to 0.90 for the five scales (Robins, Fraley, Roberts, & Trzesniewski, 2001), and internal consistency ranges from 0.68 to 0.86 (Costa & McCrae, 1992). The NEO-FFI has been translated into several different languages and shown validity and utility in a number of different contexts; it is one of the most widely used measures of the Five-Factor Model (Pytlik Zillig, Hemenover, & Dienstbier, 2002).

* Corresponding author. Present address: Personality, Stress, Coping Section, Gerontology Research Center, 5600 Nathan Shock Drive, Baltimore, MD 21224-6825, USA. Fax: +1-410-558-8108.

E-mail address: jeffm@ipc.grc.nia.nih.gov (R.R. McCrae).

A recent article, however, suggested caution in its use. Egan, Deary, and Austin (2000) reported an item factor analysis of the NEO-FFI in a British sample composed of farmers, physicians, and clinical referrals ($N=1025$). They found clear support for Neuroticism (N), Agreeableness (A), and Conscientiousness (C) factors, but problems with the Extraversion (E) and Openness (O) factors. Specifically, five of the E items and three of the O items had loadings less than 0.30 on the intended factor, and several had higher loadings on other factors. Egan et al. (2000) concluded that “the NEO-FFI should be revised such that it is made more robust” (p. 918). In another item factor analysis, Parker and Stumpf (1998) found problems with some O and A items and concluded that “consideration should be given to replacing these problematic items ... with other items from the rich item pool of the NEO-PI-R” (p. 1021).

Criticisms of a scale based entirely on item analyses may not be appropriate. The five constructs measured by the NEO-FFI are extremely broad, and the necessary diversity of item content is likely to reduce internal consistency and make the recovery of item factors more difficult. The existing NEO-FFI has worked well in studies of heritability (Riemann, Angleitner, & Strelau, 1997), adult development (Robins et al., 2001), and the prediction of personality disorders (Brieger, Sommer, Blöink, & Marneros, 2000), and one ought perhaps to “leave well enough alone.” Ultimately, however, it is an empirical question whether revisions would improve the instrument. In this article we report efforts to improve the psychometric qualities of the NEO-FFI and to make it accessible to a somewhat larger population of respondents.

1.1. Identifying weaker items

Several item factor analyses have now been published, and an evaluation of the instrument should be based on all available evidence. In addition to the Egan et al. study, Holden and Fekken (1994) analyzed data from a sample of 243 Canadian college women. Parker and Stumpf (1998) administered the NEO-FFI to 598 gifted sixth grade students and obtained parent ratings on the observer rating form of the instrument for 147 of them. Item analyses of translations into German ($N=2112$; Borkenau & Ostendorf, 1993), and into Czech ($N=945$), Polish ($N=350$), and Slovak ($N=516$; Hřebíčková, Urbánek, Čermák, Szarota, Ficková, & Orlická, 2002) have also been reported.

In each of these studies, five factors were extracted and varimax factor loadings were reported. (Because Yoshimura, Ono, Nakamura, Nathan, & Suzuki, 2001, rotated eight factors, their Japanese data were not included in this analysis.) The five intended factors were clearly recovered in each case, although some items had less than optimal loadings. In general, the same items tended to be weak across different studies; the most consistently problematic items were from the O and A scales. To quantify item quality, items were scored as *good* (2; loading 0.40 or above), *fair* (1; loading 0.30–0.39), or *poor* (0; loading <0.30). The sum across the eight studies had a coefficient alpha of 0.91 and ranged from 1 to 16, with a mean of 13.6. Seven items had scores <8, suggesting that they were, on average, worse than fair. These items (3, 8, 27, 28, 29, 34, and 38) are clearly candidates for replacement.

It is noteworthy that only one of these items is from the E Scale. In general, E items worked well across these many different samples. In fact, two of the items that failed to work in the Egan et al. study loaded above 0.40 on the E factor in all seven other analyses. Some peculiarity of the Egan et al. sample may account for this.

In reviewing items for possible revision, special consideration was given to items in the A scale. The NEO-FFI items had been selected from the items in the original NEO-PI (Costa & McCrae, 1985), where only a global, 18-item A Scale had been included. The 48 A items of the Revised NEO-PI (NEO-PI-R; Costa & McCrae, 1992) provide a much larger pool from which to choose, and two additional A items (9 and 19) which scored relatively low in item quality across the eight studies were also selected for replacement.

A different consideration is the reading level and intelligibility of the items. In a study of 1973 high school students who completed the full, 240-item NEO-PI-R, students were instructed to leave blank any item they did not understand or did not believe to be relevant. Analyses of missing data identified 30 items that appeared to be difficult for at least 2% and up to 10% of the sample (McCrae et al., 2002). Half of these items had previously been identified as among the most difficult for Belgian adolescents (De Fruyt, Mervielde, Hoekstra, & Rolland, 2000). Five of these 30 items (6, 12, 15, 24, and 42) were used in the NEO-FFI; replacing them would presumably make the instrument applicable to a wider range of respondents.

Together, these analyses suggest that it would be worthwhile to find substitutes for 14 of the 60 NEO-FFI items (23.3%). Although new items might be written, a simpler approach is to consider the remaining items in the NEO-PI-R (cf. Parker & Stumpf, 1998). Of the 180 items not used in the NEO-FFI, 25 were among those identified as difficult for adolescent responders (McCrae et al., 2002). New NEO-FFI items can be selected from the remaining pool of 155 items.

Both item response theory methods (Reise & Henson, 2000) and confirmatory factor analysis (Egan et al., 2000) have been used to evaluate NEO-PI-R items. Most item analyses, however, have used exploratory factor analysis, and that method will be employed here.

2. Method

Two samples were used to select new items. The first consisted of high school students ($N = 1959$) with complete personality data (McCrae et al., 2002). These students were enrolled in psychology courses and volunteered to take the NEO-PI-R. They ranged in age from 14 to 18 ($M = 16.5$, $S.D. = 1.0$ years); approximately two-thirds were girls. The second sample included adults from the Baltimore Longitudinal Study of Aging (BLSA; Shock et al., 1984). BLSA participants are generally healthy and well-educated men and women who have volunteered to return to the Gerontology Research Center for periodic medical and psychosocial testing. Between 1991 and 2002, 1492 participants (695 men and 797 women) aged 19–93 ($M = 56.2$, $S.D. = 17.0$ years) completed the NEO-PI-R by computer administration; 65.1% of the sample was White, 27.6% Black, and 7.3% other race. The use of these two samples provides a kind of cross-validation and ensures that the final instrument will be appropriate across the full age range.

Three criteria were used to select new NEO-FFI items. First was the requirement that not more than two thirds of the items be keyed in the same direction, to minimize the effects of acquiescent responding. Next, items were correlated with their associated NEO-PI-R factor scores in both samples, and items with higher correlations were preferred. Finally, an effort was made to diversify item content by selecting from facets underrepresented in the remaining items of the scale. By these criteria, NEO-FFI items 6, 12, 27, 42, 3, 8, 28, 38, 9, 19, 24, 29, 34, and 15 were replaced by NEO-PI-R items 186, 127, 7, 32, 123, 48, 213, 133, 189, 169, 84, 139, 184, and 95, respectively.

Keying changed for several items. The result can be identified as the Revised NEO Five-Factor Inventory (NEO-FFI-R).

Validity of the old and new scales was assessed by correlations with archival measures of the FFM available for the BLSA sample. Self-reports and observer ratings by from one to four peers on a set of 80 bipolar adjective scales were obtained in 1983 (McCrae & Costa, 1987). Five orthogonal factors were derived from each source and interpreted as the lexical Big Five factors. Self-reports on the NEO-PI-R were obtained in 1986 and 1990 (Costa & McCrae, 1992).

3. Results

The 60 NEO-FFI items and the 60 NEO-FFI-R items were factored in both samples. In each analysis, the expected five-factor solution was obtained, confirming that the same structure can be found in adolescents and adults. Table 1 therefore presents factor loadings in the combined sample. For the NEO-FFI analysis, the first seven eigenvalues were 8.61, 4.24, 3.35, 3.16, 2.75, 1.44 and 1.40; for the NEO-FFI-R, they were 8.17, 4.76, 3.46, 3.18, 2.71, 1.58, and 1.40. In both cases, a scree test supports the retention of five factors. NEO-FFI communalities ranged from 0.15 to 0.54; NEO-FFI-R communalities ranged from 0.16 to 0.53. By the criteria used earlier, Table 1 shows that in this sample, the NEO-FFI has 52 good items, 5 fair items, and 3 poor items. The revision appeared to be successful: In the NEO-FFI-R there are 57 good items and 3 fair items, two of which just miss the 0.40 cut-off. In 13 of the 14 substitutions, the new item has a higher factor loading than the item it replaced. The exception is item 6, which is an excellent item for respondents who understand it, but which causes problems for a small minority of adolescents. The best replacement, subject to constraints on keying and diversification of content, is a good but not excellent definer of the N factor.

Table 2 presents psychometric information on the two versions in the two samples. Correlations between the two versions are high; even the A Scale, in which 5 of 12 items were changed, shows near-equivalence. The short scales are highly correlated with the NEO-PI-R factor scores, although the new version is not appreciably better in this regard than the old except in the case of A. Internal consistency estimates are all acceptable; in the case of O and A Scales, the revision appears to have strengthened reliability somewhat.

Table 3 reports means and standard deviations. A comparison of the NEO-FFI values with the norms shows that the High School sample is somewhat lower in A and C than college age norms, probably because A and C begin to increase in the years after high school (Robins et al., 2001). The adult sample is lower in N and higher in O than adult norms. These differences are not large—about one-half standard deviation—and suggest that the NEO-FFI-R values in Table 3 could be used provisionally as norms.

Table 4 reports correlations of adjective factors and NEO-PI-R domain scales with the NEO-FFI and NEO-FFI-R Scales. These data provide strong evidence for the validity of the FFI scales. In particular, correlations with peer adjective factors range from 0.29 to 0.43, all $P < 0.001$, despite differences in observer and instrument, and an interval of at least eight years between administrations. Correlations with self-report measures are understandably higher, ranging from 0.44 to 0.80. Correlations for the revised scales are higher than those for the original scales in

Table 1

Factor loadings for original and revised NEO Five-Factor Inventory items in the combined high school and Baltimore Longitudinal Study of Aging samples

Item	NEO-FFI factor					NEO-FFI-R factor				
	N	E	O	A	C	N	E	O	A	C
1	0.53	-0.10	0.02	0.04	0.07	0.55	-0.09	-0.01	0.03	0.07
6 ^{a,b}	0.64	-0.10	-0.11	0.06	-0.15	0.40	-0.20	0.13	-0.24	-0.01
11	0.71	0.05	-0.01	-0.05	-0.07	0.72	0.06	-0.01	-0.06	-0.09
16	0.63	-0.19	0.08	-0.07	-0.13	0.64	-0.18	0.06	-0.08	-0.15
21	0.65	-0.04	-0.03	-0.14	-0.08	0.66	-0.04	-0.04	-0.14	-0.10
26	0.70	-0.08	-0.05	-0.08	-0.19	0.68	-0.07	-0.06	-0.09	-0.21
31	0.61	-0.01	0.01	0.10	-0.07	0.63	-0.01	0.00	0.09	-0.08
36	0.56	-0.04	-0.05	-0.31	-0.07	0.55	-0.03	-0.06	-0.31	-0.08
41	0.64	-0.06	-0.16	-0.01	-0.29	0.62	-0.05	-0.15	-0.03	-0.31
46	0.61	-0.20	0.07	-0.04	-0.14	0.62	-0.17	0.05	-0.04	-0.15
51	0.54	-0.02	-0.10	-0.06	-0.27	0.52	-0.03	-0.10	-0.07	-0.29
56	0.60	0.02	-0.05	-0.05	-0.15	0.58	0.00	-0.03	-0.08	-0.17
2	0.10	0.65	-0.12	-0.09	-0.14	0.08	0.69	-0.12	-0.06	-0.14
7	0.02	0.55	0.09	0.07	-0.01	0.02	0.52	0.13	0.08	-0.01
12 ^{a,b}	-0.08	0.37	0.06	0.17	-0.09	-0.09	0.50	-0.05	0.19	-0.07
17	-0.01	0.61	0.11	0.17	0.06	-0.01	0.64	0.14	0.19	0.05
22	0.06	0.57	0.03	-0.23	-0.05	0.06	0.59	0.05	-0.22	-0.06
27 ^a	-0.11	0.44	-0.08	0.17	-0.12	-0.22	0.57	-0.05	0.00	-0.02
32	-0.02	0.59	0.11	-0.11	0.05	-0.03	0.56	0.13	-0.08	0.06
37	-0.21	0.68	0.01	0.12	0.13	-0.21	0.63	0.05	0.13	0.14
42 ^{a,b}	-0.31	0.51	0.06	0.25	0.05	-0.02	0.52	0.09	0.17	-0.01
47	0.12	0.42	0.12	-0.27	0.15	0.14	0.41	0.12	-0.25	0.17
52	-0.13	0.54	0.05	-0.14	0.29	-0.12	0.53	0.08	-0.12	0.30
57	-0.19	0.41	-0.00	-0.09	0.10	-0.18	0.43	-0.02	-0.05	0.11
3 ^a	0.27	0.16	0.34	-0.01	-0.33	0.30	0.23	0.36	-0.13	-0.22
8 ^a	-0.19	-0.12	0.16	-0.10	-0.30	-0.03	0.20	0.39	0.08	0.16
13	-0.01	0.00	0.63	0.17	0.08	-0.03	-0.03	0.63	0.15	0.07
18	-0.08	0.01	0.42	0.05	0.00	-0.07	0.01	0.40	0.02	0.02
23	0.08	0.11	0.56	0.17	-0.03	0.07	0.10	0.56	0.16	-0.04
28 ^a	-0.12	0.19	0.36	-0.04	0.06	0.09	0.05	0.40	0.01	-0.25
33	-0.05	0.07	0.50	0.08	0.09	-0.04	0.04	0.51	0.07	0.09
38 ^a	-0.07	-0.15	0.29	-0.08	-0.17	0.36	0.29	0.39	-0.02	0.07
43	0.08	0.07	0.61	0.11	-0.01	0.06	0.04	0.61	0.10	-0.01
48	-0.10	-0.05	0.64	0.07	0.03	-0.13	-0.08	0.63	0.05	0.01
53	-0.04	0.14	0.61	-0.10	0.17	-0.06	0.10	0.61	-0.10	0.16
58	-0.02	0.01	0.57	-0.23	-0.01	-0.07	-0.00	0.56	-0.22	-0.03
4	0.05	0.25	0.13	0.43	0.25	0.05	0.23	0.16	0.42	0.23
9 ^a	-0.41	-0.08	0.08	0.39	0.22	-0.17	-0.11	0.07	0.55	0.16
14	-0.10	0.07	-0.07	0.52	0.08	-0.09	0.07	-0.05	0.51	0.08
19 ^a	0.13	0.00	0.04	0.44	0.00	-0.12	-0.16	0.07	0.53	0.05
24 ^{a,b}	-0.39	0.15	0.00	0.45	-0.02	0.19	0.00	-0.08	0.48	-0.05
29 ^a	-0.38	-0.05	0.20	0.36	-0.01	-0.07	0.12	0.04	0.42	0.00
34 ^a	-0.11	0.39	0.11	0.21	0.19	-0.21	0.24	0.10	0.40	0.11

(continued on next page)

Table 1 (continued)

Item	NEO-FFI factor					NEO-FFI-R factor				
	N	E	O	A	C	N	E	O	A	C
39	−0.03	0.29	0.08	<i>0.61</i>	0.05	−0.02	0.25	0.11	<i>0.59</i>	0.06
44	−0.20	−0.08	−0.02	<i>0.53</i>	−0.09	−0.20	−0.08	−0.02	<i>0.52</i>	−0.09
49	0.09	0.24	0.20	<i>0.42</i>	0.31	0.09	0.20	0.24	<i>0.40</i>	0.28
54	−0.03	−0.10	0.09	<i>0.58</i>	0.09	−0.03	−0.13	0.11	<i>0.58</i>	0.08
59	−0.12	−0.12	−0.03	<i>0.56</i>	0.14	−0.10	−0.12	−0.03	<i>0.60</i>	0.14
5	−0.06	−0.05	−0.11	0.11	<i>0.56</i>	−0.04	−0.04	−0.11	0.11	<i>0.56</i>
10	−0.22	−0.05	−0.10	0.11	<i>0.62</i>	−0.19	−0.05	−0.10	0.11	<i>0.64</i>
15 ^{a,b}	−0.13	−0.12	0.13	−0.02	<i>0.46</i>	−0.30	−0.15	0.06	0.09	<i>0.47</i>
20	−0.09	−0.02	0.07	0.18	<i>0.57</i>	−0.07	−0.03	0.06	0.17	<i>0.58</i>
25	−0.10	0.15	−0.01	−0.12	<i>0.63</i>	−0.08	0.15	−0.00	−0.09	<i>0.63</i>
30	−0.32	−0.12	−0.04	0.13	<i>0.51</i>	−0.28	−0.10	−0.06	0.14	<i>0.54</i>
35	−0.06	0.24	0.13	0.01	<i>0.66</i>	−0.03	0.22	0.15	0.03	<i>0.67</i>
40	−0.14	0.10	0.16	0.11	<i>0.57</i>	−0.11	0.09	0.17	0.12	<i>0.57</i>
45	−0.28	−0.09	−0.02	0.12	<i>0.46</i>	−0.26	−0.07	−0.04	0.12	<i>0.48</i>
50	−0.11	0.13	0.04	−0.04	<i>0.66</i>	−0.08	0.12	0.06	−0.03	<i>0.67</i>
55	−0.21	−0.05	−0.01	0.11	<i>0.59</i>	−0.18	−0.03	−0.02	0.11	<i>0.60</i>
60	−0.02	0.20	0.13	−0.13	<i>0.59</i>	−0.00	0.19	0.14	−0.09	<i>0.59</i>
% ^c	9.89	7.20	5.60	5.76	8.42	8.90	7.48	5.97	6.37	8.41

These are Varimax-rotated principal components. $N = 3451$. Loadings greater than 0.40 in absolute magnitude are given in italics.

^a Item changed in revision.

^b Item replaced because of readability.

^c Percent of total variance explained by the rotated factor.

eight instances, but the differences are trivial in magnitude. It appears that the two sets of scales are equally valid.

Finally, Table 5 presents intercorrelations among the two sets of scales, because some authors have expressed concern about the orthogonality of the five scales (Block, 1995). For the NEO-FFI, absolute discriminant correlations ranged from 0.01 to 0.42, with a median of 0.18; for the NEO-FFI-R, they ranged from 0.02 to 0.43 with a median of 0.15. Curiously, there was a notable difference in the correlation between N and C in the two samples: Among adults, the correlation was a fairly substantial -0.43 , whereas among students it was only -0.21 . Overall, however, the five scales show clear evidence of discriminant validity.

4. Discussion

The NEO-FFI was not intended to provide definitive measurement of the five personality factors. Instead, it was designed as a brief instrument that would yield reasonable estimates of the factors, perhaps chiefly of use in exploratory research. In over a decade of use, it has shown itself to be reliable, valid, and useful in a variety of contexts and cultures.

Table 2
Psychometric properties of original and revised NEO Five-Factor Inventory scales in two samples

Scale	Version	R^a	r_{FFI}		r_{Factor}		α	
			HS	BLSA	HS	BLSA	HS	BLSA
Neuroticism	FFI	4			0.88	0.83	0.82	0.86
	FFI-R	4	0.99	0.99	0.88	0.83	0.82	0.84
Extraversion	FFI	4			0.86	0.83	0.78	0.80
	FFI-R	4	0.94	0.94	0.87	0.83	0.80	0.81
Openness	FFI	7			0.88	0.91	0.70	0.75
	FFI-R	5	0.91	0.91	0.88	0.89	0.76	0.77
Agreeableness	FFI	8			0.83	0.76	0.72	0.69
	FFI-R	8	0.88	0.87	0.88	0.81	0.75	0.73
Conscientiousness	FFI	4			0.87	0.86	0.81	0.79
	FFI-R	4	0.99	0.99	0.87	0.85	0.81	0.79

r_{FFI} = correlation with original FFI scale, r_{Factor} = correlation with NEO-PI-R factor. HS, High School sample, $N = 1959$. BLSA, Baltimore Longitudinal Study of Aging sample, $N = 1492$.

^a Number of reverse-keyed items.

Table 3
Means and standard deviations in two samples by gender

Scale	Version	Men				Women				Combined			
		HS		BLSA		HS		BLSA		HS		BLSA	
		M	S.D.	M	S.D.	M	S.D.	M	S.D.	M	S.D.	M	S.D.
N	FFI	22.46	8.03	14.62	6.76	25.73	7.89	16.77	7.91	24.65	8.07	15.77	7.47
	FFI-R	23.15	8.01	15.79	6.68	26.02	7.76	16.50	7.00	25.08	7.95	16.83	7.36
E	FFI	29.40	6.49	27.85	6.17	31.18	6.70	29.06	6.29	30.58	6.67	28.50	6.26
	FFI-R	30.62	6.81	28.74	6.48	32.40	6.95	29.78	6.41	31.80	6.94	29.29	6.46
O	FFI	28.06	6.73	28.33	5.98	28.59	6.50	30.18	6.09	28.40	6.57	29.32	6.11
	FFI-R	30.59	7.05	30.19	5.92	31.49	6.91	32.25	6.13	31.18	6.96	31.29	6.12
A	FFI	27.23	6.46	32.31	4.76	28.86	6.23	34.34	4.98	28.31	6.34	33.39	4.98
	FFI-R	26.54	6.93	31.11	5.13	28.86	6.80	33.54	5.43	28.09	6.93	32.41	5.42
C	FFI	26.47	7.33	33.18	6.15	27.93	7.23	33.73	6.52	27.45	7.30	33.48	6.36
	FFI-R	26.05	7.41	32.88	6.06	27.48	7.34	33.59	6.50	27.00	7.40	33.26	6.30

HS, High School sample; BLSA, Baltimore Longitudinal Study of Aging sample. For the High School sample, 1312 girls and 635 boys reported gender.

Table 4
Convergent correlations for NEO-FFI and NEO-FFI-R Scales

Criterion	NEO-FFI	NEO-FFI-R
<i>Mean peer-rated adjective factors (N = 133)</i>		
N	0.33	0.32
E	0.38	0.37
O	0.42	0.43
A	0.36	0.33
C	0.29	0.29
<i>Self-reported adjective factors (N = 214)</i>		
N	0.44	0.45
E	0.59	0.61
O	0.45	0.44
A	0.57	0.51
C	0.49	0.50
<i>Self-reported NEO-PI-R domain scales (N = 240)</i>		
N	0.75	0.76
E	0.80	0.80
O	0.74	0.76
A	0.71	0.75
C	0.77	0.78

All correlations significant at $P < 0.001$.

Table 5
Intercorrelations among NEO-FFI and NEO-FFI-R scales

Scale	NEO-FFI Scale					NEO-FFI-R Scale				
	N	E	O	A	C	N	E	O	A	C
N										
E	–0.38					–0.35				
O	–0.04	0.24				0.03	0.28			
A	–0.29	0.22	0.10			–0.23	0.11	0.03		
C	–0.42	0.32	0.01	0.17		–0.43	0.33	0.04	0.15	

Correlations above the diagonals are within the high school sample; correlations below the diagonal are within the BLSA sample.

But good instruments can be improved, and test developers have an obligation to evaluate the research literature and undertake revisions when appropriate. In the present analyses we have suggested changes to 14 of the 60 NEO-FFI items. Because all the new items are taken from the NEO-PI-R, researchers who have collected NEO-PI-R data can score the NEO-FFI-R. Researchers who wish to use the NEO-FFI-R itself in future research can obtain a license from the publisher to do so. However, the improvements over the original NEO-FFI are modest, and for most purposes the published version of the instrument, with its established norms, will continue to

suffice. The NEO-FFI-R may prove most useful in special populations where literacy is likely to be a problem, although it also works well in normal volunteer populations.

The review of item factor analyses included data from seven countries and five different languages, and yet the same items were repeatedly identified as better or worse. This suggests that the universality of personality traits (McCrae & Costa, 1997) may extend even to the level of items. It also suggests the hypothesis that the same revisions adopted here might also make modest improvements to the quality of the NEO-FFI in translation.

References

- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*, 187–215.
- Borkenau, P., & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae: handanweisung*. Göttingen: Hogrefe.
- Brieger, P., Sommer, S., Blöink, R., & Marneros, A. (2000). The relationship between Five-Factor personality measurements and ICD-10 personality disorder dimensions: results from a sample of 229 subjects. *Journal of Personality disorders*, *14*, 282–290.
- Costa, P. T. Jr., & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T. Jr., & McCrae, R. R. (1989). *The NEO-PI/NEO-FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T. Jr., & 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.
- De Fruyt, F., Mervielde, I., Hoekstra, H. A., & Rolland, J.-P. (2000). Assessing adolescents' personality with the NEO-PI-R. *Assessment*, *7*, 329–345.
- Egan, V., Deary, I., & Austin, E. (2000). The NEO-FFI: Emerging British norms and an item-level analysis suggest N, A, and C are more reliable than O and E. *Personality and Individual Differences*, *29*, 907–920.
- Holden, R. R., & Fekken, G. C. (1994). The NEO Five-Factor Inventory in a Canadian context: psychometric properties for a sample of university women. *Personality and Individual Differences*, *17*, 441–444.
- Hřebíčková, M., Urbánek, T., Čermák, I., Szarota, P., Ficková, E., & Orlická, L. (2002). The NEO Five-Factor Inventory in Czech, Polish, and Slovak contexts. In R. R. McCrae, & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 53–78). New York: Kluwer Academic/Plenum Publishers.
- McCrae, R. R., & Costa, P. T. Jr. (1987). Validation of the Five-Factor Model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*, 81–90.
- McCrae, R. R., & Costa, P. T. Jr. (1989). Rotation to maximize the construct validity of factors in the NEO Personality Inventory. *Multivariate Behavioral Research*, *24*, 107–124.
- McCrae, R. R., & Costa, P. T. Jr. (1997). Personality trait structure as a human universal. *American Psychologist*, *52*, 509–516.
- McCrae, R. R., Costa, P. T. Jr., Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., & Mervielde, I. (2002). Personality trait development from 12 to 18: longitudinal, cross-sectional, and cross-cultural analyses. *Journal of Personality and Social Psychology*, *83*, 1456–1468.
- Parker, W., & Stumpf, H. (1998). A validation of the Five-Factor Model of personality in academically talented youth across observers and instruments. *Personality and Individual Differences*, *25*, 1005–1025.
- Pytlik Zillig, L. M., Hemenover, S. H., & Dienstbier, R. A. (2002). What do we assess when we assess a Big 5 trait? A content analysis of the affective, behavioral, and cognitive processes represented in Big 5 personality inventories. *Personality and Social Psychology Bulletin*, *28*, 847–858.
- Reise, S. P., & Henson, J. M. (2000). Computerization and adaptive administration of the NEO-PI-R. *Assessment*, *7*, 347–364.
- Riemann, R., Angleitner, A., & Strelau, J. (1997). Genetic and environmental influences on personality: a study of twins reared together using the self- and peer report NEO-FFI scales. *Journal of Personality*, *65*, 449–475.

- Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality*, *69*, 617–640.
- Shock, N. W., Greulich, R. C., Andres, R., Arenberg, D., Costa, P. T. Jr., Lakatta, E. G., & Tobin, J. D. (1984). *Normal human aging: the Baltimore Longitudinal Study of Aging (NIH Publication No. 84-2450)*. Bethesda, MD: National Institutes of Health.
- Yoshimura, K., Ono, Y., Nakamura, K., Nathan, J. H., & Suzuki, K. (2001). Validation of the Japanese version of the NEO Five-Factor Inventory in a large community sample. *Psychological Reports*, *88*, 443–449.