

Toward More Readable Big Five Personality Inventories

René Mõttus, Helle Pullmann, and Jüri Allik

Estonian Center of Behavioral and Health Sciences, Department of Psychology, University of Tartu, Estonia

Abstract. The Estonian version of the International Personality Item Pool NEO (IPIP-NEO; Goldberg, 1999) was administered to 297 participants in parallel with the Estonian version of the NEO-PI-R (Kallasmaa, Allik, Realo, & McCrae, 2000). On average, the EPIP-NEO items were 3 words, 7 syllables, and 18 characters shorter than the NEO-PI-R items. By all relevant psychometrical properties the EPIP-NEO was comparable to the NEO-PI-R. The mean convergent correlation between the facet scales was .73. The scales with shorter and grammatically simpler items tended to have higher internal consistency. In an independent cross-validation sample the initial results were generally replicated. The scales also demonstrated an adequate cross-observer agreement. It is concluded that the EPIP-NEO, as a more readable personality inventory compared to the NEO-PI-R, is suitable for a wider range of samples with different levels of reading skills.

Keywords: International Personality Item Pool, Revised NEO Personality Inventory, readable personality inventories, internal consistency, comparative validity, cross-validation, cross-observer agreement

Introduction

The Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) is one of the most comprehensive and widely used instruments for measurement of the Big Five personality dimensions – Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A), and Conscientiousness (C). Although the NEO-PI-R was originally designed for adult populations, developing separate norms later made it applicable to college students as well (Sherry, Henson, & Lewis, 2003). Recent studies have shown that generally even 12-year-old children are able to understand and respond properly to items of the NEO personality inventories (Allik, Laidra, Realo, & Pullmann, 2004; De Fruyt, Mervielde, Hoekstra, & Rolland, 2000; Markey, Markey, Tinsley, & Ericksen, 2002; McCrae et al., 2002; Parker & Stumpf, 1998). However, several studies have also indicated that the application of the NEO questionnaires may have some limitations. For example, McCrae and his colleagues (McCrae et al., 2002) administered the NEO-PI-R to high school students with instructions to leave blank any item they did not understand and found that 30 items of 240 were “difficult” containing either obscure terms (such as *fastidious* and *lackadaisical*) or being otherwise not easy to comprehend. Some other studies have also suggested that several items of the NEO-PI-R should be modified since they are either out of date (“I believe that the ‘new morality’ of permissiveness is no morality at all”) or too difficult for some populations (Costa & McCrae, 1997; McCrae & Cos-

ta, 2004). Allik and colleagues (Allik et al., 2004) found that the adult factor structure of the Estonian NEO Five-Factor Inventory (NEO-FFI) was less clearly replicated in 12-year-olds than in older adolescents and adult samples. Because similar slight deviations were observed in less educated groups (Allik & McCrae, 2004; Toomela, 2003; Austin, Deary, & Gibson, 1997), it was suggested that a certain minimal amount of ability is required for observing one’s own personality dispositions and giving reliable self-reports on the basis of these observations (Allik et al., 2004). Consistently, clinicians have observed that reading and comprehension deficits may be responsible for a significant change in the validity of some scales (Knäuper & Wittchen, 1994; Krakauer, Archer, & Gordon, 1993; Paolo, Ryan, & Smith, 1991). Thus, although the NEO-PI-R has widely proven its validity, a more readable measure of personality would be desirable for a wider range of application, including samples of young adolescents and less educated individuals (cf. McCrae, Costa, & Martin, 2005).

One promising candidate for a more universal personality measure is the IPIP-NEO questionnaire that has been developed on the basis of the International Personality Item Pool (IPIP; Goldberg, 1999; Johnson, 2005). The IPIP-NEO items are short, using mainly familiar words, and with a simple grammar. The IPIP-NEO was designed to match the original NEO-PI-R structure in which five personality dimensions are described by 30 facets. The IPIP-NEO facet scales consist of 10 items each, whereas facet scales of the NEO-PI-R include eight items. According to Goldberg (1999) the average of the coefficient α values of the IPIP-

NEO scales was a little higher than that of the NEO-PI-R (.80 and .75, respectively), which is almost exactly what could be predicted from the Spearman-Brown prophecy formula. The average cross-instrument agreement between corresponding scales of the NEO-PI-R and the IPIP-NEO was .73 (Goldberg, 1999). Thus, the psychometric quality of the IPIP-NEO scales looks promising. Nevertheless, more studies, including translations into different languages, are needed for generalizability of this preliminary observation.

It is possible, however, that shorter, less sophisticated, and linguistically minimalist personality items are not able to convey the original meaning. For example, it has been claimed that the Openness to experience dimension is not well represented in natural language (McCrae, 1990) and it may also be problematic to represent it with unsophisticated statements. It is without doubt that simpler and less specific statements can become more ambiguous and can bring in undesirable shades of meaning. In addition, it is possible that easier items are more vulnerable to socially desirable responding, which can also compromise validity (cf. Stricker, 1963).

The main goal of this study was to evaluate psychometric properties of the Estonian version of the IPIP-NEO, the Estonian Personality Item Pool NEO (EPIP-NEO), which was administered in parallel with the Estonian version of the NEO-PI-R (Kallasmaa et al., 2000). The aim was to establish to what extent the new and linguistically simpler personality inventory retained the intended five-factor structure, and showed acceptable internal consistency and external validity compared to the original NEO-PI-R. However, it was possible that the EPIP-NEO had an advantage over the NEO-PI-R because items of the former had been selected on the basis of the same data that was used for the comparisons. For this reason, a cross-validation sample was used to evaluate the properties of the EPIP-NEO in independent data. In addition, cross-validation data also allowed examination of the cross-observer agreement of the scales.

Method

Participants

The initial sample consisted of 297 participants (94 men and 203 women) with a mean age of 31.34 years ($SD = 13.73$, ranging from 16 to 86; for one participant the age was unknown). About 20% of the participants had completed a primary or basic education, 44% had finished a secondary or vocational school, 35% had a university degree; about 1% declined to declare their educational level. Thus, in respect to the educational level the sample was relatively heterogeneous. The subjects were instructed to fill out the questionnaires on two subsequent days. In order to avoid practice effects roughly half of the subjects filled out the

EPIP-NEO on the first day and other half on the second day.

The cross-validation sample consisted of 182 participants (117 women and 54 men, 11 participants did not report their gender) with a mean age of 42.41 years ($SD = 17.78$; ranging from 16 to 83; for nine participants the age was unknown). These participants also had various educational backgrounds, as approximately 18% of participants had a basic or primary education, 54% had completed a vocational or secondary school, and 23% had a university degree; about 5% did not report their educational level. In the cross-validation sample, 168 participants' (*targets*) personality traits were also rated by two well-acquainted informants (*judges*), who were mostly close relatives, spouses, friends, or colleagues. Most of the judges also reported how long they had been acquainted with their target ($n = 324$) and how well they knew him or her ($n = 314$). On average, the judges knew the target for 17.86 years ($SD = 12.59$), more than 90% knew the target for at least 3 years and less than 2% knew the target for less than a year. About 90% of the judges reported that they knew their target at least "satisfactorily well."

Measures

EPIP-NEO

The first author of this article translated the original IPIP-NEO items retaining their short and simple style. Five experts, including the coauthors, revised the translated items in order to make them as unsophisticated and readable as possible. In case of necessity grammar was simplified. All words considered uncommon by the experts and foreign words were replaced with more common nonforeign terms. Since balanced scales are suggested in order to diminish the confounding effects of acquiescent responding (McCrae, Herbst, & Costa, 2001), a few additional items were generated for facet scales with unequal number of inverted and noninverted items. The final item pool of 397 items was back-translated by an English teacher who was unfamiliar with the inventory. The back-translated items were reviewed by John A. Johnson and according to his suggestions some revisions were made. In order to make the EPIP-NEO similar to the NEO-PI-R, eight items were selected for each facet scale. The selection was based on convergent and discriminant relations between items: Selected items correlated most with the other items of the intended scale and least with other scales. All facet scales except two (O6: Liberalism and A4: Cooperation) were balanced containing an equal number of positively and negatively keyed items. The selection of items was carried out only on the basis of the initial sample. In the cross-validation sample scale scores were comprised of the previously selected items. Judges used the same questionnaire, except that the items were worded in the third person form (i.e., "I am . . ." was replaced with "She/he is . . .").

NEO-PI-R

The main psychometric properties of the Estonian version of the NEO-PI-R are fully described elsewhere (Kallasmaa et al., 2000).

Validity Criteria

In order to validate both self-report personality questionnaires, a subset of participants of the initial sample ($n = 239$, 75 men and 164 women; mean age 33.6 years, $SD = 13.73$) used a 10-point scale (1 – never ... 10 – several times a day) to rate the frequency of several concrete behavioral acts that were considered typical indicators of either neuroticism or extraversion. First, the respondents were asked to indicate how frequently they had taken medication for depression. Based on their ratings, the subjects were assigned to two groups: those who had taken antidepressants during the last year (28%) and those who did not report taking the medication (72%). Second, an indicator of social activity was compiled from three items (“I am at a party or crowded event,” “I meet my friends,” and “I ask my friends whether they have plans for a party”). The scale had an internal consistency of .78.

Results and Discussion

Linguistic and psychometric properties of the EPIP-NEO and the NEO-PI-R scales are presented in Table 1. The EPIP-NEO items were on average 3 words, 7 syllables, and 18 characters shorter than the NEO-PI-R items. In Estonian, the number of commas can also be used as a fairly good indicator of grammatical complexity. The EPIP-NEO had on average 0.10 whereas the NEO-PI-R included 0.48 commas per item.

On average, the Cronbach's α s (see Table 1) of the EPIP-NEO facet scales were slightly higher than those of the NEO-PI-R facet scales (mean values were .79 and .76, respectively). Interestingly, internal consistencies of scales were related to the linguistic properties of items: The scales with shorter items tended to be internally more consistent. The correlations between scales' average number of letters per items and the Cronbach's α s were $r = -.56$ ($p = .001$) and $r = -.33$ ($p = .08$), respectively for the EPIP-NEO and the NEO-PI-R, and $r = -.42$ ($p = .001$) across all 60 facet scales. Although the magnitude of the relationship was affected by scales with more extreme item length and internal consistency, the tendency of scales with longer items to be internally less consistent was still noticeable. For example, ignoring the scales of the Openness dimension, which generally consisted of longer items, did not eradicate the relationship. Other supposedly relevant indicators of linguistic complexity (e.g., scales' average number of syllables, com-

mas and words per item) were similarly related to internal consistency.

In the cross-validation sample the Cronbach's α s were .95, .93, .89, .88, and .92 respectively for the N, E, O, A, and C domains. The α coefficients of the facet scales varied from .56 to .90 with an average value of .78. The mean α coefficient of the 30 NEO-PI-R facet scales reported by Kallasmaa and colleagues in an independent study (Kallasmaa et al., 2000) was .75. Thus, it is likely that in respect to internal consistency the EPIP-NEO is at least as good as the original NEO-PI-R. The previously reported relationship between scales' internal consistency and linguistic properties also appeared in independent data. For example, the correlation between scales' average number of letters per item and the Cronbach's α s was $r = -.54$ ($p < .01$). Comparable patterns of correlations also appeared for the other linguistic properties.

The mean values and standard deviations of the EPIP-NEO and the NEO-PI-R facet scales were rather similar. The mean profiles of the EPIP-NEO and the NEO-PI-R 30 facet scales were highly correlated, $r = .86$ ($p < .001$). The convergent correlations between the corresponding facet scales (last column in Table 1) ranged from $r = .45$ to .84 with an average correlation of $r = .73$.

Principal component analysis (PCA) followed by Varimax rotation applied to the 30 facet scales of the EPIP-NEO revealed a typical five-factor structure (see Table 2). The first seven eigenvalues were 6.53, 5.06, 2.98, 2.49, 1.67, .99, and .93. Parallel analysis (Zwick & Velicer, 1986) and other criteria (e.g., scree test) suggested retaining five factors that accounted for 62.4% of the total variance (in case of the NEO-PI-R the respective figure was 61.9%). The factor congruence coefficients between Varimax-rotated factor structures of the EPIP-NEO and the NEO-PI-R were .98, .97, .95, .95, and .97 for the N, E, O, A, and C factors, respectively. Generally, the typical five-factor structure of the NEO-PI-R was replicated by both inventories. After Procrustes rotation (McCrae, Zonderman, Costa, Bond, & Paunonen, 1996) targeted at the North-American normative structure of the NEO-PI-R (Costa & McCrae, 1992), the factor congruence coefficients between North-American factors and respective factors obtained from the current data ranged from .95 to .96 and from .95 to .98 for the EPIP-NEO and the NEO-PI-R, respectively. However, some unintended primary loadings were observed for both measures. First, the N5 (Impulsiveness) of the EPIP-NEO had its primary loading on the C factor and a similar tendency appeared in the structure of the NEO-PI-R as the N5 had equal loadings on the intended N factor and C factor. Incidentally, the “wrong” loading of the N5 on the C factor was not specific to this sample as the same has been seen previously (Kallasmaa et al., 2000). Analogously for both inventories, A5 (Modesty) had its primary loading on the E factor. Again, a similar deviation has been reported previously (Kallasmaa et al., 2000). More specific to the EPIP-NEO was the unintended primary loading of the O4 (Adventurousness) on the E factor. The facet scale A2

Table 1. Linguistic and psychometric properties of the EPIP-NEO and the NEO-PI-R scales.

	EPIP-NEO					NEO-PI-R					<i>r</i>				
	Words	Sylla- bles	Letters	Com- mas	α	<i>Mean</i>	<i>SD</i>	Words	Sylla- bles	Letters		Com- mas	α	<i>Mean</i>	<i>SD</i>
Neuroticism	4.94	10.21	24.73	0.06	.95	75.82	28.08	7.54	16.63	40.46	0.38	.94	85.84	26.09	.89
Extraversion	4.69	10.27	25.10	0.02	.93	111.05	25.38	7.56	17.67	42.79	0.27	.94	110.82	27.05	.90
Openness	5.98	13.56	32.79	0.13	.89	128.16	19.93	9.77	22.94	56.69	0.67	.89	117.31	21.50	.86
Agreeableness	5.65	12.06	28.96	0.23	.90	126.26	20.59	8.33	18.67	45.69	0.61	.87	118.01	18.28	.83
Conscientiousness	5.54	11.29	27.42	0.08	.93	120.89	23.79	8.21	18.40	44.46	0.48	.91	112.36	22.14	.87
N1 Anxiety	4.75	10.00	24.00	0.00	.86	15.97	6.70	6.38	13.63	32.63	0.25	.85	15.98	6.31	.82
N2 Anger	3.88	8.13	19.63	0.00	.91	11.98	7.26	6.88	15.75	38.13	0.38	.79	12.73	5.43	.78
N3 Depression	4.88	9.63	23.00	0.13	.89	11.32	6.72	7.00	15.25	37.25	0.25	.86	14.19	6.59	.84
N4 Self-Consciousness	5.25	10.63	26.75	0.00	.78	12.40	5.59	10.13	21.38	52.75	0.63	.75	15.30	5.20	.73
N5 Immoderation	5.50	11.38	26.63	0.25	.73	14.60	5.18	6.25	14.13	34.00	0.25	.70	17.45	4.94	.66
N6 Vulnerability	5.38	11.50	28.38	0.00	.84	9.55	5.15	8.63	19.63	48.00	0.50	.85	10.19	5.04	.80
E1 Friendliness	5.50	12.38	30.25	0.00	.83	21.82	5.43	7.38	18.00	42.88	0.25	.76	21.73	5.15	.77
E2 Gregariousness	4.25	10.00	25.38	0.00	.83	17.69	6.33	9.00	19.50	50.88	0.75	.85	16.56	6.53	.84
E3 Assertiveness	4.88	10.63	25.50	0.13	.75	17.24	5.19	8.38	19.38	46.00	0.25	.84	16.64	5.92	.80
E4 Activity Level	4.25	9.50	23.13	0.00	.74	16.99	5.21	6.88	15.25	33.88	0.13	.83	17.92	6.26	.69
E5 Excitement-Seeking	5.13	11.00	26.88	0.00	.78	16.59	6.21	6.88	17.63	43.50	0.13	.73	17.35	5.66	.79
E6 Cheerfulness	4.13	8.13	19.50	0.00	.86	20.72	5.94	6.88	16.25	39.50	0.13	.86	20.63	6.25	.77
O1 Imagination	4.88	11.50	26.13	0.00	.86	23.15	5.77	8.63	21.25	50.25	0.50	.83	20.91	5.89	.78
O2 Artistic Interests	5.38	11.13	29.38	0.13	.79	24.61	5.24	8.75	18.25	47.13	0.50	.80	20.91	5.90	.76
O3 Emotionality	6.00	13.25	31.88	0.13	.83	22.75	5.24	9.25	22.25	52.63	0.38	.80	22.18	4.91	.78
O4 Adventurousness	5.75	12.25	29.38	0.13	.69	18.77	4.40	9.50	21.75	56.13	0.63	.65	15.16	4.78	.63
O5 Intellect	5.88	14.88	35.88	0.00	.79	22.26	5.36	8.75	23.13	56.38	0.38	.86	18.89	6.63	.74
O6 Liberalism	8.00	18.38	44.13	0.38	.58	16.62	4.54	13.75	31.00	76.75	1.63	.61	19.26	4.42	.52
A1 Trust	5.75	12.63	28.25	0.38	.84	20.78	5.34	7.75	18.25	43.13	0.88	.79	20.71	5.07	.78
A2 Morality	6.63	12.38	31.25	0.50	.79	22.79	5.54	9.13	20.38	47.88	0.88	.76	19.26	5.46	.72
A3 Altruism	5.38	11.00	28.50	0.13	.79	22.52	4.39	7.88	18.50	46.25	0.38	.63	21.31	3.76	.45
A4 Cooperation	4.50	9.75	23.75	0.00	.69	19.64	4.86	8.50	17.50	43.50	0.63	.62	15.81	4.38	.63
A5 Modesty	5.38	12.75	28.75	0.00	.79	17.24	5.71	7.38	15.00	37.50	0.50	.8	18.91	5.63	.80
A6 Sympathy	6.25	13.88	33.25	0.38	.77	23.29	4.87	9.38	22.38	55.63	0.38	.55	22.02	3.91	.66
C1 Self-Efficacy	5.38	10.38	24.88	0.25	.74	20.18	4.23	7.13	17.13	42.13	0.38	.66	19.09	4.36	.68
C2 Orderliness	5.38	10.88	26.13	0.00	.84	19.86	6.18	8.75	18.50	43.75	0.50	.70	17.90	5.00	.71
C3 Dutifulness	4.88	9.75	24.25	0.13	.80	22.62	4.93	7.25	17.13	41.00	0.25	.67	22.08	4.47	.69
C4 Achievement Striving	6.38	12.38	30.25	0.00	.75	20.95	5.04	8.50	18.38	46.38	0.50	.75	17.72	5.40	.70
C5 Self-Discipline	5.50	10.75	26.00	0.00	.87	17.68	6.36	11.13	23.50	55.00	1.00	.77	18.22	5.08	.82
C6 Cautiousness	5.75	13.63	33.00	0.13	.77	19.61	5.06	6.50	15.75	38.63	0.25	.73	17.36	4.91	.73

NOTE: $n = 297$; Words = average number of words per item; Syllables = average number of syllables per item; Letters = average number of letters per item (without spaces and punctuation marks); Commas = average number of commas per item; α = Cronbach's α ; *Mean* = mean score; *SD* = standard deviation of scores; *r* = correlation between corresponding scales of the EPIP-NEO and the NEO-PI-R (all correlations are significant at $p < .01$). Only facet names of the EPIP-NEO are provided.

Table 2. Varimax rotated factor structures and congruence coefficients of the EPIP-NEO and the NEO-PI-R

	EPIP-NEO						NEO-PI-R					
	<i>N</i>	<i>E</i>	<i>O</i>	<i>A</i>	<i>C</i>	<i>Cong</i> ¹	<i>N</i>	<i>E</i>	<i>O</i>	<i>A</i>	<i>C</i>	<i>Cong</i> ²
N1 Anxiety	-.85	-.08	.01	.09	.01	.98	-.85	-.03	.05	-.05	-.01	.98
N2 Anger	-.75	.08	-.04	-.31	.02	.96	-.81	.09	-.01	-.27	-.09	.98
N3 Depression	-.80	-.16	.02	-.14	-.14	.96	-.84	-.24	.03	-.03	-.17	.99
N4 Self-Consciousness	-.61	-.43	-.21	.10	-.29	.95	-.68	-.36	-.17	.03	-.16	.98
N5 Immoderation	-.52	.20	.13	-.02	-.60	.95	-.47	.28	.10	.05	-.47	.97
N6 Vulnerability	-.70	-.32	-.15	-.01	-.34	.99	-.67	-.32	-.16	.10	-.32	.98
E1 Friendliness	.29	.72	.09	.29	.12	.90	.18	.71	.21	.38	.09	.93
E2 Gregariousness	.04	.78	-.20	.21	-.07	.97	.10	.75	-.03	.11	-.01	.95
E3 Assertiveness	.05	.76	.17	-.27	.12	.94	.13	.65	.25	-.28	.21	.96
E4 Activity Level	.01	.62	.12	-.07	.27	.98	.07	.68	.09	.13	.43	.95
E5 Excitement-Seeking	.04	.70	.16	-.10	-.23	.96	.04	.72	.21	-.12	-.05	.98
E6 Cheerfulness	.18	.69	.08	.33	-.01	.97	.26	.70	.17	.29	.02	.95
O1 Imagination	-.09	.34	.63	.04	-.19	.94	-.22	.26	.64	-.02	-.19	.97
O2 Artistic Interests	.01	.08	.61	.41	.07	.96	-.06	.01	.59	.43	.11	.94
O3 Emotionality	-.30	.33	.61	.28	.14	.96	-.24	.36	.65	.26	.16	.97
O4 Adventurousness	.07	.51	.30	.08	-.14	.84	.18	.23	.46	.02	-.22	.94
O5 Intellect	.18	.10	.77	-.01	.04	.99	.10	.05	.80	-.09	.03	.98
O6 Liberalism	.20	-.05	.51	-.09	-.30	.92	.35	.03	.58	-.04	-.21	.96
A1 Trust	.33	.12	.14	.52	-.08	.97	.31	.07	.16	.67	-.02	.98
A2 Morality	.06	-.22	.00	.60	.51	.95	.11	-.50	-.04	.49	.22	.97
A3 Altruism	-.01	.09	.06	.75	.28	.91	.06	.16	.06	.66	.24	.97
A4 Cooperation	.38	-.27	.05	.67	.09	.99	.23	-.44	-.07	.58	-.02	.98
A5 Modesty	-.03	-.57	-.31	.42	.16	.90	-.12	-.57	-.13	.46	.03	.94
A6 Sympathy	-.17	.25	.10	.76	.15	.93	-.24	.11	.02	.70	.07	.91
C1 Self-Efficacy	.42	.33	.18	-.06	.60	.99	.45	.22	.04	-.04	.67	.99
C2 Orderliness	-.10	-.06	-.17	.11	.72	.95	-.03	.09	-.09	.01	.77	.99
C3 Dutifulness	.04	-.08	-.11	.33	.74	.95	.06	-.15	-.11	.33	.70	.96
C4 Achievement Striving	.01	.21	.17	.06	.73	.97	-.01	.25	-.03	.02	.80	.96
C5 Self-Discipline	.19	.13	-.14	.11	.78	.98	.22	.12	.05	.12	.78	.98
C6 Cautiousness	.17	-.24	.00	.07	.64	.99	.19	-.34	-.15	.11	.61	.98
Factor congruence coefficients						Factor congruence coefficients						
.96 .95 .95 .95 .96						.96 .98 .95 .95 .98						

Note: $n = 297$; $Cong^1$ = congruence coefficient of the EPIP-NEO with American normative structure of the NEO-PI-R (Costa & McCrae, 1992) after Procrustes rotation; $Cong^2$ = congruence coefficient of the NEO-PI-R with American normative structure of the NEO-PI-R (Costa & McCrae, 1992) after Procrustes rotation. Factor loadings above .40 are in bold. Only facet names of the EPIP-NEO are provided.

(Straightforwardness) of the NEO-PI-R had a “wrong” primary loading on the E factor. Interestingly, the corresponding facet A2 (Morality) of the EPIP-NEO had a strong secondary loading on the C factor. The Procrustes-rotated structure of the EPIP-NEO, targeted at the normative North-American Varimax-rotated structure of the NEO-PI-R (Costa & McCrae, 1992), slightly more resembled the simple structure since the unintended primary loadings of the N5 and O4 diminished remarkably and the primary loading of A5 resettled on the intended A factor.

In the cross-validation sample, the Varimax-rotated factor structure (PCA) of the 30 EPIP-NEO facet scales was less clear since eight facet scales had primary loadings on other than intended factors. However, it has been argued that Varimax might not be the most suitable method of rotation when replicating the factor structure of a five-factor measure, especially in smaller samples (Allik & McCrae, 2004), and that targeted rotation is preferable for this purpose (McCrae et al., 1996). After Procrustes rotation at the North-American normative structure of the NEO-PI-R (Costa & McCrae, 1992) the factor structure was slightly clearer, as four facets had their primary loading on an unintended factor. The factor congruence coefficients between the present factors and respective North-American factors were .92, .92, .92, .93, and .95 for the N, E, O, A, and C, respectively. Thus, although the factor structure that appeared in the cross-validation sample was far from the intended structure, it demonstrated satisfactory congruence with the normative structure of the NEO-PI-R. In the averaged ratings of two judges the Varimax-rotated structure was more similar to the expected factor structure because only three facet scales had primary loadings on other than the intended factor. After Procrustes rotation at the normative structure as reported by Costa and McCrae (1992), the factor congruence coefficients were .90, .92, .94, .89, and .94, for the N, E, O, A, and C, respectively.

Analysis of Validity

Age Differences

The changes in the mean levels of the five personality factors across the life span are well known and replicated in different cultures (McCrae et al., 1999). Similarly to previous results, the N, E, and O domain scales correlated negatively ($r = -.15, -.38, \text{ and } -.19$, respectively, $p < .01$) and the A and C scales positively ($r = .28$ and $.16$, respectively, $p < .01$) with age. A similar pattern of correlations appeared between the NEO-PI-R domains and age.

In cross-validation data ($n = 173$) the results were partially replicated. The E domain correlated negatively and the A and C domains positively with age (respectively $r = -.17, p < .05$; $r = .36, p < .01$; and $r = .18, p < .05$). The negative correlation between age and the O domain ($r = -.12$) did not reach the level of statistical significance. In-

terestingly, in cross-validation data the N domain was positively correlated with age ($r = .19, p < .05$). Within the N domain, however, only the N1 (Anxiety) was significantly positively related to age ($r = .18, p < .05$).

Gender Differences

Consistent with previous results (Costa, Terracciano, & McCrae, 2001), women scored significantly higher on the N domain, $t(295) = 3.18, p < .01$, and the A domain of the EPIP-NEO, $t(295) = 4.50, p < .001$. With the NEO-PI-R the corresponding gender differences were rather similar. Neither of the measures reported significant gender differences on E, though in both cases women scored significantly higher in E1 (Warmth). Women had higher mean scores on O for both the EPIP-NEO, $t(295) = 3.23, p < .01$; and NEO-PI-R, $t(295) = 2.44, p < .05$. Curiously, women had significantly higher mean values on the EPIP-NEO C domain, $t(295) = 3.01, p < .01$. Though the gender difference on C did not reach the level of significance in case of the NEO-PI-R, there also appeared to be similar significant differences in two facet scales of the C domain. Thus, except for the C domain, the gender differences in the EPIP-NEO scores generally corresponded to previous findings (Costa et al., 2001) and present results with the NEO-PI-R.

In cross-validation data, the gender differences were generally replicated: Women had significantly higher mean scores on the N, O, and A dimensions, $t(169) = 2.31, p < 0.05$; $t(169) = 2.63, p < 0.01$; and $t(169) = 2.91, p < 0.01$, respectively. For the other domains gender differences were not significant.

Behavioral Criteria

The comparison of the mean values of the EPIP-NEO domains and behavioral criterion of depression indicated that subjects who reported taking medication for depression had significantly higher scores on the N scale, $t(237) = 4.40, p < .001$, and lower scores on the E scale, $t(237) = -2.70, p < .01$, compared to those who did not take medication. The other three domains were not significantly related to taking medication. This result is perfectly in line with previous findings that clinical depression is primarily related to Negative Affectivity or Neuroticism and Positive Affectivity or Extraversion (see Clark, Watson, & Mineka, 1994; Terracciano, McCrae, & Costa, 2003). Nearly identical relationships were found for the N and E domain scales of the NEO-PI-R. Further, extraverted people are usually considered to be socially active, seeing their friends frequently, and going to parties or other social events. According to this description it is obvious that individuals with higher values on the Extraversion scale should report the relevant behavior taking place more frequently. As expected, the indicator of social ac-

tivity was significantly related to the E scale of the EPIP-NEO ($r = .51, p < .001$). There was also a moderate relationship with the O scale ($r = .31, p < .001$), which is predictable considering the correlation between the E and O personality domains (in the current case $r = .50, p < .001$). Social activity was not significantly related to the other three EPIP-NEO domains. The correlations between the social activity indicator and the original NEO-PI-R domains were similar.

Cross-Observer Validity

A powerful indicator of accuracy of the assessment of personality traits is cross-observer agreement (Funder, 1995). The correlations between self-ratings and averaged ratings of two judges ($n = 168$) were .57, .67, .40, .42, and .41 ($p < .001$) for the N, E, O, A, and C domains, respectively. At the facet level, the cross-observer correlations ranged from .24 (A4: Cooperation) to .59 (O6: Cheerfulness) with an average correlation of .45 (all correlations were significant at $p < .001$, except A4, which had a correlation significant at $p < .01$). When the cross-observer agreement was calculated only for those targets known at least “satisfactory well” by both informants ($n = 120$), the generally improved correlations were .63, .71, .43, .41, and .45 (respectively for the N, E, O, A, and C, all significant at $p < .001$) with an average correlation of .49 at the facet level. Thus, just as expected, knowing the target better raised the cross-observer correlations. To conclude, the level of cross-observer agreement was fully comparable to the previous findings reported by McCrae et al. (2004), thus, demonstrating completely acceptable cross-observer validity of the EPIP-NEO scales.

Conclusions

Although it might be argued that linguistically minimalist personality questionnaire items could have a wider bandwidth of meaning compared to more specifically formulated statements, the current study provides a convincing demonstration that scales with short and simple items can be reliable and valid.

The results of this study confirmed that the EPIP-NEO as a linguistically simple Big Five personality inventory had psychometrical properties comparable to those of the well established but linguistically more sophisticated NEO-PI-R. In this respect, the EPIP-NEO can be compared with a short and more readable version of the MMPI, which also demonstrated psychometrical properties similar to its original version (Ward & Selby, 1980). More specifically, the EPIP-NEO scales had generally even higher internal consistencies than the original NEO-PI-R scales. The factor structure of the EPIP-NEO replicated the original five-factor structure of the NEO-PI-R

(Costa & McCrae, 1992) at least satisfactorily. The cross-instrument agreement between the EPIP-NEO and the NEO-PI-R was excellent at the domain level and with few exceptions also at the level of facet scales. Further, the analysis of external validity lent support to the convergent and discriminant validity of the linguistically simple EPIP-NEO as the instrument generally replicated the established theoretical and empirical relationships with different external variables. Another important evidence of the validity of the EPIP-NEO scales came from good cross-observer agreement.

An important result was that scales with shorter items (i.e., containing a smaller number of letters, syllables, words, and commas) had a tendency to have higher internal consistency. Assuming that shorter items tend to be more readable and, in turn, more easily understandable (but see also Schinka & Borum, 1993), this observation is consistent with previous reports that more comprehensible items lead to more stable and valid scales (Angleitner, John, & Löhr, 1986; Holden, Fekken, & Jackson, 1985). However, it is also possible that some personality traits are more “complicated” in nature. The “complicatedness” could have two effects: More “complicated” traits might need more sophisticated statements to describe them and, independently from the linguistic properties of items, they could have less consistent descriptors – behavioral, affective, or thought patterns that are used to describe these traits – that, in turn, could lead to lower internal consistency in the assessment scales. In this case, there would be no causal relationship between items’ surface characteristics and internal consistencies since both would covary as functions of an underlying variable, i.e., “complicatedness” of the trait.

To sum up, the EPIP-NEO, being a linguistically simpler and, therefore, probably more readable instrument, has the potential to be a useful personality questionnaire for wider range of samples. Recently, McCrae, Costa et al. (2005) also developed a more readable version of the NEO Personality Inventory, NEO-PI-3, and consistent with the present results, concluded that the more readable version apparently had better reliability and validity in samples with different reading skills (McCrae, Martin, & Costa, 2005).

Acknowledgments

The authors are indebted to Liina Haring and Airis Erme for their efforts with data collection, and to Kenn Konstabel for his extremely helpful comments throughout the study. We also thank Anu Realo, Pirjo Saadla, and John A. Johnson for their help with the item translation. This project was supported by the Estonian Science Foundation Grant 5677 and an Estonian Ministry of Education and Science Grant 0180543 provided to the third author.

References

- Allik, J., Laidra, K., Realo, A., & Pullmann, H. (2004). Personality development from 12 to 18 years of age: Changes in mean levels and structure of traits. *European Journal of Personality, 18*, 445–462.
- Allik, J., & McCrae, R.R. (2004). Escapable conclusions: Toomela (2003) and the universality of trait structure. *Journal of Personality and Social Psychology, 87*, 261–265.
- Angleitner, A., John, O.P., & Löhr, F.-J. (1986). It's *what* you ask and *how* you ask it: An item-metric analysis of personality questionnaires. In A. Angleitner & J.S. Wiggins (Eds.), *Personality assessment via questionnaires*. (pp. 61–108). Berlin: Springer-Verlag.
- Austin, E.J., Deary, I.J., & Gibson, G.J. (1997). Relationships between ability and personality: Three hypotheses tested. *Intelligence, 25*, 49–70.
- Clark, L.A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103*, 103–116.
- Costa, P.T.Jr., & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEOFFI) – professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Costa, P.T., & McCrae, R.R. (1997). Stability and change in personality assessment: The Revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment, 68*, 86–94.
- Costa, P.T., Terracciano, A., & McCrae, R.R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322–331.
- De Fruyt, F., Mervielde, I., Hoekstra, H.A., & Rolland, J.P. (2000). Assessing adolescents' personality with the NEO-PI-R. *Assessment, 7*, 329–345.
- Funder, D.C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652–70.
- 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*. (pp. 7–28). Tilburg: Tilburg University Press.
- Holden, R., Fekken, G.C., & Jackson, D.N. (1985). Structured personality test item characteristics and validity. *Journal of Research in Personality, 19*, 386–394.
- Johnson, J.A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39*, 103–129.
- Kallasmaa, T., Allik, J., Realo, A., & McCrae, R.R. (2000). The Estonian version of the NEO-PI-R: An examination of universal and culture-specific aspects of the Five-Factor Model. *European Journal of Personality, 14*, 265–278.
- Knäuper, B. & Wittchen, H.-U. (1994). Diagnosing major depression in the elderly: Evidence for response bias in standardized diagnostic interviews? *Journal of Psychiatric Research, 28*, 147–164.
- Kraskauer, S.Y., Archer, R.P., & Gordon, R.A. (1993). The development of the Items-Easy (Ie) and Items-Difficult (Id) subscales for the MMPI-a. *Journal of Personality Assessment, 60*, 561–571.
- Markey, P.M., Markey, C.N., Tinsley, B.J., & Ericksen, A.J. (2002). A preliminary validation of preadolescents' self-reports using the Five-Factor Model of personality. *Journal of Research in Personality, 36*, 173–181.
- McCrae, R.R. (1990). Traits and trait names – how well is openness represented in natural languages. *European Journal of Personality, 4*, 119–129.
- McCrae, R.R., & Costa, P.T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences, 36*, 587–596.
- McCrae, R.R., Costa, P.T., De Lima, M.P., Simoes, A., Ostendorf, F., Angleitner, A., Marusic, I., Bratko, D., Caprara, G.V., Barbaranelli, C., Chae, J.H., & Piedmont, R.L. (1999). Age differences in personality across the adult life span: Parallels in five cultures. *Developmental Psychology, 35*, 466–477.
- McCrae, R.R., Costa, P.T., & Martin, T.A. (2005). A more readable revised NEO Personality Inventory. *Journal of Personality Assessment, 84*, 261–270.
- McCrae, R.R., Martin, T.A., Costa, P.T. (2005). Age trends and age norms for the NEO Personality Inventory – 3 in adolescents and adults. *Assessment, 12*, 363–373.
- McCrae, R.R., Costa, P.T., Martin, T.A., Oryol, V.E., Rukavishnikov, A.A., Senin, I.G., Hrebickova, M., & Urbanek, T. (2004). Consensual validation of personality traits across cultures. *Journal of Research in Personality, 38*, 179–201.
- McCrae, R.R., Costa, P.T., Terracciano, A., Parker, W.D., Mills, C.J., De Fruyt, F., & Mervielde, I. (2002). Personality trait development from age 12 to age 18: Longitudinal, cross-sectional, and cross-cultural analyses. *Journal of Personality and Social Psychology, 83*, 1456–1468.
- McCrae, R.R., Herbst, J.H. & Costa, P.T. Jr. (2001). Effects of acquiescence on personality factor structures. In R. Riemann, F.M. Spinath, & F. Ostendorf (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp. 217–231). Berlin: Pabst Science Publishers.
- 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.
- Paolo, A.M., Ryan, J.J., & Smith, A.J. (1991). Reading difficulty of MMPI-2 subscales. *Journal of Clinical Psychology, 47*, 529–532.
- Parker, W.D., & 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.
- Schinka, J.A., & Borum, R. (1993). Readability of adult psychopathology inventories. *Psychological Assessment, 5*, 384–386.
- Sherry, A., Henson, R.K., & Lewis, J.G. (2003). Evaluating the appropriateness of college-age norms for use with adolescents on the NEO Personality Inventory-Revised. *Assessment, 10*, 71–78.
- Stricker, L.J. (1963). Acquiescence and social desirability response styles, item characteristics, and conformity. *Psychological Reports, 12*, 319–341.
- Terracciano, A., McCrae, R.R., & Costa, P.T. (2003). Factorial and construct validity of the Italian Positive and Negative Affect Schedule (PANAS). *European Journal of Psychological Assessment, 19*, 131–141.
- Toomela, A. (2003). Relationships between personality structure,

structure of word meaning, and cognitive ability: A study of cultural mechanisms of personality. *Journal of Personality and Social Psychology*, 85, 723–735.

Ward, L.C., & Selby, R.B. (1980). Abbreviation of the MMPI with increased comprehensibility and readability. *Journal of Clinical Psychology*, 36, 180–186.

Zwick, W.R., & Velicer, W.F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432–442.

René Mõttus

Department of Psychology
University of Tartu
Tiigi 78
Tartu 50410
Estonia
Tel. +372 7375 902
Fax +372 7376 152
E-mail Rene.Mottus@ut.ee