Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Research in Personality 38 (2004) 179–201
JOURNAL OF RESEARCH IN PERSONALITY
www.elsevier.com/locate/jrp

Consensual validation of personality traits across cultures☆

Robert R. McCrae,[a,*] Paul T. Costa, Jr.,[a] Thomas A. Martin,[b] Valery E. Oryol,[c] Alexey A. Rukavishnikov,[c] Ivan G. Senin,[c] Martina Hřebíčková,[d] and Tomáš Urbánek[d]

[a] *Laboratory of Personality and Cognition, National Institute on Aging, National Institutes of Health, Box #03, Gerontology Research Center, 5600 Nathan Shock Drive, Baltimore, MD 21224-6825, USA*
[b] *Department of Psychology, Susquehanna University, Selinsgrove, PA, USA*
[c] *Department of Psychology, Yaroslavl State University, Yaroslavl, Russia*
[d] *Institute of Psychology, Academy of Sciences of the Czech Republic, Brno, Czech Republic*

Abstract

Cross-observer agreement on personality trait ratings has been interpreted as particularly powerful evidence of the veridicality of personality traits, but cross-cultural studies of consensual validity are relatively rare. In this article we review the available literature on cross-observer agreement on traits of the Five-Factor Model, and provide new data from Russia and the Czech Republic. Russian and Czech versions of the Revised NEO Personality Inventory showed adequate internal consistency and replicated the American factor structure and gender differences. Cross-observer correlations showed moderate to high agreement, especially for Extraversion. Despite cultural differences in individualism/collectivism that affect many psychological processes, these data suggest that personality traits exist and function in much the same way in these cultures.

Keywords: Cross-cultural; Personality traits; Self/other agreement

☆ Portions of this article were presented at the Third Annual Association for Research in Personality Conference, February, 2003, Los Angeles, and at the Annual Convention of the American Psychological Association, August, 2003, Toronto. This research was supported in part by Grant 406/01/1507 from the Grant Agency of the Czech Republic and is related to research plan AV 0Z7025918 of the Institute of Psychology, Academy of Sciences of the Czech Republic.
* Corresponding author. Fax: 1-410-558-8108.
E-mail address: jeffm@lpc.grc.nia.nih.gov (R.R. McCrae).

0092-6566/$ - see front matter
doi:10.1016/S0092-6566(03)00056-4

1. Introduction

Personality traits were widely disparaged in the 1970s (e.g., Shweder, 1975). In response, defenders of traits pointed to their longitudinal stability (Block, 1977) and their consistency when aggregated across occasions (Epstein, 1977), but some of the most persuasive evidence came from studies showing consensual validation of personality traits (Funder, 1980; McCrae, 1982; Woodruffe, 1985). In these heteromethod studies, agreement across observers could not plausibly be explained by response sets or shared cognitive schemas; instead, it pointed to the existence of objective psychological attributes that both respondents perceived more or less accurately (Funder, 1995). Cross-observer agreement also played a major role in establishing the Five-Factor Model (FFM) of personality as a widely accepted taxonomy of traits (McCrae & Costa, 1987). Today, there is continued interest in cross-observer agreement, much of it concerned with accounting for variations in the strength of agreement across different traits and raters (Funder, Kolar, & Blackman, 1995; McCrae, Stone, Fagan, & Costa, 1998).

Cross-cultural research on personality, however, has not systematically addressed the question of whether there is cultural variation in cross-observer agreement. There is some reason to expect that there would be. Cultural psychologists (Cross & Markus, 1999) have argued that personality traits, as distinctive and enduring aspects of individuals, are essentially a Western phenomenon; in non-Western, collectivistic societies, personality characteristics are fluid, determined more by transient interpersonal situations than by enduring traits (Markus & Kitayama, 1991). Indeed, some have argued that the trait construct is meaningless in non-Western societies (see Church, 2000, for a fuller discussion). In a careful review of the evidence from studies in the Philippines, Church and Katigbak (2000) concluded that traits do exist and function in that collectivistic society, but they noted that evidence on cross-observer agreement was weak: "Future studies should compare the extent of interjudge agreement cross-culturally using reliable and equivalent measures and well-acquainted judges in each culture" (p. 87).

There is now considerable evidence that the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) provides reliable and valid measures of personality traits in a wide variety of cultures, from Zimbabwe to the Russian Arctic (Draguns, Krylova, Oryol, Rukavishnikov, & Martin, 2000; Piedmont, Bain, McCrae, & Costa, 2002). The NEO-PI-R assesses 30 traits, or facets, six for each of the factors in the FFM. Using translations prepared by psychologists from around the world, the American factor structure has been replicated in a wide range of cultures (Rolland, 2002). Gender differences (Costa, Terracciano, & McCrae, 2001) and maturational trends (McCrae et al., 1999) on NEO-PI-R scales have also been widely replicated. In general, these studies suggest that the NEO-PI-R functions much the same in all cultures. However, there are also some cross-cultural differences: Standard deviations of NEO-PI-R scales are consistently smaller in Asian countries than in the West (McCrae, 2002), and gender differences are less marked among Asians and Black Africans than among Americans and Europeans (Costa et al., 2001).

The present article reviews the cross-cultural literature on cross-observer agreement and provides new data from two Slavic cultures: Russia and the Czech Republic. Although both of these are Western nations, they are considerably less individualistic than the United States: On Hofstede's (2001) Individualism-Collectivism dimension they score 39 and 58, respectively, compared to 91 for the US. Indeed, by this measure, Russia is more collectivistic than Japan, which scores 46. For most of the past 50 years, both Russia and Czechoslovakia were controlled by Communist governments that actively suppressed the rights of individuals. If stable and observable traits are found only in highly individualistic cultures, then interjudge agreement should be considerably lower in these two nations than in the United States.

Cross-observer agreement can be assessed whenever a set of targets are each rated by at least two observers (one of whom may be the self). The magnitude of the agreement depends on who the raters are and how the data are treated. Consistent with the principle of aggregation, scores based on mean ratings from a group of raters typically have higher validity than scores based on single ratings (McCrae & Costa, 1989). Agreement also varies with degree of acquaintance (Funder et al., 1995), with married couples typically showing the highest levels of concordance. In the present study we will examine self/other ratings using spouses and siblings as raters in Russia and spouses, siblings, other relatives, and friends as raters in the Czech Republic. We will also examine informant/informant correlations in the Czech data.

1.1. Cross-observer agreement in North America and elsewhere

Although cross-observer studies on the FFM can be traced back as far as the 1960s (Norman & Goldberg, 1966), the number of studies is not large. In German samples, Amelang and Borkenau (1982) and Ostendorf (1990) reported peer/self agreement on adjective factors that could be interpreted as representing the FFM. Lanning (1994) examined interjudge agreement for FFM factors derived from the California Q-Set (Block, 1961), and found that all and only these factors, plus Attractiveness, showed high interrater agreement. In a confirmatory factor analysis of the Big Five Questionnaire, Barbaranelli and Caprara (2002) showed that all five factors were defined by both self-reports and peer ratings.

Several studies of cross-observer agreement have utilized the NEO-PI-R, which is one of the few published personality inventories that offer an observer rating form as well as a self-report form. Table 1 summarizes studies conducted on three versions of the instrument: the 240-item NEO-PI-R; the original NEO Personality Inventory (NEO-PI; Costa & McCrae, 1985), a 180-item instrument with short scales to measure Agreeableness and Conscientiousness; and the brief, 60-item NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992). The table also reports findings from five other instruments that were designed to measure the FFM. To maintain comparability, only studies in which correlations were based on single raters (rather than mean rating from multiple observers) are used. Table 1 includes the major North American studies, together with all known cross-observer studies conducted outside the US.

Table 1
Cross-observer agreement in studies using measures of the Five-Factor Model with single raters

Source	Country	Language	Instrument	*N*	Type	Factor				
						N	**E**	**O**	**A**	**C**
North American studies										
McCrae and Costa (1989)	US	English	BAF[a]	718	Self/peer	.40	.40	.41	.41	.34
				712	Peer/peer	.34	.42	.43	.39	.38
Costa and McCrae (1992)	US	English	NEO-PI-R	193	Peer/peer	.43	.42	.45	.49	.22
				250	Self/peer	.36	.44	.53	.41	.40
				68	Self/spouse	.60	.73	.65	.62	.34
Piedmont (1994)	US	English	NEO-PI-R	97	Peer/peer	.52	.51	.31	.41	.54
Funder et al. (1995)	US	English	NEO-PI	121	Self/parent	.50	.42	.50	.34	.35
				274	Self/peer[b]	.28	.47	.43	.40	.39
				217	Parent/peer[b]	.25	.40	.34	.20	.34
				278	Peer/peer	.42	.57	.36	.36	.36
				88	Parent/parent	.69	.62	.45	.64	.49
McCrae et al. (1998)	US	English	NEO-PI-R	94	Self/spouse	.46	.74	.53	.58	.49
Parker and Stumpf (1998)	US	English	NEO-FFI	147	Self/parent	.37	.56	.39	.45	.49
Paunonen and Ashton (2002)	Canada	English	FF-NPQ[c]	96	Self/peer	.39	.45	.38	.40	.41
Kurtz and Sherker (2003)	US	English	NEO-FFI	103	Self/peer[d]	.41	.37	.41	.37	.56
Cross-cultural studies										
Spirrison and Choi (1998)	South Korea[e]	Korean	NEO-PI-R	38	Self/spouse	.36	.63	.56	.61	.61
Xiu, Wu, Wu, and Shui (1996)	P.R. China	Chinese	NEO-FFI	119	Self/spouse	.23	.44	.49	.41	.32
Yang et al. (1999)	P.R. China	Chinese	NEO-PI-R	160	Self/spouse	.45	.46	.39	.32	.51
Zawadski et al. (1998)	Poland	Polish	NEO-FFI	1092	Self/peer[b]	.53	.60	.52	.42	.45
				1092	Peer/peer[b]	.37	.47	.41	.32	.44
Szarota, Zawadski, and Strelau (2002)	Poland	Polish	PAL[f]	1278	Self/peer[b]	.52	.40	.27	.24	.31
				639	Peer/peer[b]	.46	.40	.41	.27	.41

Riemann, Angleitner, and Strelau (1997)	Germany	German	NEO-FFI	1928	Peer/peer[g]	.46	.48	.42	.42	.44
Ostendorf and Angleitner (in press)	Germany	German	NEO-PI-R	750	Peer/peer	.44	.51	.47	.47	.40
Barbaranelli, Caprara, Rabasca, and Pastorelli (2003)	Italy	Italian	BFQ-C[h]	639	Adult/adult[b]	.16	.28	.59	.28	.45
				2074	Adult/child[b]	.19	.25	.48	.24	.19
Perugini and Ercolani (1998)	Italy	Italian	FFPI[i]	112	Self/peer	.54	.52	.33	.39	.51
Saroglou and Fiasse (2003)	Belgium	French	NEO-PI-R	122	Self/mother	.48	.56	.66	.55	.58
Martin, Costa, Oryol, Rukavishnikov, and Senin (2002)	Russia	Russian	NEO-PI-R	60	Self/spouse	.43	.50	.45	.31	.35
Mdn for North American studies						*.41*	*.45*	*.43*	*.41*	*.39*
Mdn for cross-cultural studies						*.45*	*.48*	*.46*	*.36*	*.44*
Mdn for self/spouse						*.44*	*.57*	*.51*	*.50*	*.42*
Mdn for other types						*.42*	*.45*	*.42*	*.40*	*.41*
Mdn						*.43*	*.47*	*.43*	*.40*	*.41*

[a] Bipolar adjective factors (Varimax rotation).
[b] Median of two or more groups.
[c] Five-Factor Nonverbal Personality Questionnaire.
[d] At 15 weeks' acquaintance.
[e] Korean citizens studying in the US.
[f] Polish Adjective List.
[g] Spearman–Browne correction removed.
[h] Big Five Questionnaire—Children Version.
[i] Five-Factor Personality Inventory (measures Autonomy instead of Openness).

None of the samples reported in Table 1 is strictly representative of its country or culture, nor are they directly comparable to each other: Zawadzki, Strelau, Szczepaniak, and Śliwińska (1998) studied Polish twins, Yang et al. (1999) assessed Chinese psychiatric patients, and Parker and Stumpf (1998) reported data from intellectually gifted adolescents. It is perhaps all the more remarkable that despite these sampling differences, results are fairly consistent across samples, raters, and factors. The largest correlations are found when Extraversion is rated and when self-reports are correlated with spouse ratings, but all domains and all types of rater show significant and substantial agreement. If we classify Spirrison and Choi's (1998) study of Korean students in America as a cross-cultural study, these data suggest that cross-observer agreement is as high in less individualistic cultures as it is in North America.

Not reported in the table are cross-observer correlations for the 30 individual facet scales of the NEO-PI-R. Perhaps because they are shorter and thus less reliable, these values are somewhat lower than those reported in Table 1. For example, in one American study, the median peer/peer correlation for the 30 facets was .36; the median self/peer correlation was .34 (Costa & McCrae, 1992).

The last study in Table 1 reports pilot data from Russia, suggesting that personality traits can be consensually validated in a collectivistic country with a recent history of enforced socialism. The present article reports data from a much larger, independent Russian sample, together with data from a Central European culture, the Czech Republic.

1.2. The cross-cultural generalizability of differential consensual validity

There is continuing interest in whether some traits are more easily and accurately rated than others (Funder, 1995). Borkenau and Liebler (1995) showed that Extraversion is the factor most easily inferred from observable cues; strangers were able to make accurate ratings of Extraversion after watching a brief videotape. John and Robins (1993; see also Asendorpf & Ostendorf, 1998), in a study of adjective ratings, reported that highly evaluative traits showed poor interrater agreement.

In principle, it should be possible to predict the magnitude of agreement based on an understanding of the psychological processes involved in making self-reports and ratings. For example, ratings on highly evaluative traits may be biased by the rater's personal feelings for the target. These underlying processes may or may not be the same in different cultures. The same trait might be highly desirable in some cultures, but not in others; if so, John and Robins's work suggests that interjudge agreement on that trait should be higher in cultures where it is not highly evaluative.

There do not appear to have been any systematic cross-cultural investigations of the relative consensual validity of different traits, although at least one comparison can be made: Borkenau and Liebler (1995), with a German sample, and John and Robins (1993), with American samples, found that traits related to Extraversion elicited highest inter-judge agreement. In the present study we will explore this issue

using the 30 facets of the NEO-PI-R in previously reported American, German, and Chinese samples, and new Russian and Czech samples.[1]

2. Method

2.1. Participants

The Russian sample consisted of 800 respondents (387 men and 413 women) ranging from 15 to 80 years of age ($M = 31.2$); 26% were 15–21 years of age, 49% were 22–40 years of age, and 25% were 41–80 years of age. All respondents were paired with a sibling ($n = 166$) or a spouse ($n = 634$). Among the youngest group, 74% of respondents were paired with a sibling and 26% with a spouse, whereas 97% of the 22–40 group and 100% of the 41–80 group were paired with their spouse. Of the respondents paired with siblings, 68% were paired with a sibling of the other sex, 24% were pairs of sisters, and 8% were pairs of brothers. The Russian respondents did not provide information on educational level.

The Czech sample included 1365 respondents (578 men, 764 women; 23 did not report gender), who were recruited in a series of studies. They ranged in age from 14 to 83 years, with a mean age of 30.92 ($SD = 13.62$ years); 11% were aged 14–21 years, 53% were from 22 to 40, and 34% from 41 to 83. Information on education was provided by 1076 respondents: 16% of the respondents had primary education (10 years of mandatory education in the Czech Republic), 60% had secondary and 24% university education.

Peer-ratings were provided by 910 raters (366 men, 522 women; 20 did not report gender) aged 14–83 years ($M = 35.66$; $SD = 14.15$ years) who participated in one of two research designs. In the self/other agreement studies ($N = 616$), each target provided a self-report and was rated by one informant. In the consensus study, 49 targets (26 men and 23 women aged 16–67 years) provided a self-report and were each rated by three informants. The target then rated each of the three informants. This design yielded $49 \times 6 = 294$ observer ratings.

For the analysis of self/other agreement, the 616 targets from the Czech self/other agreement studies were supplemented by the 49 targets and one of their raters (who had been rated by the target) from the consensus study. The rater chosen was a spouse if available; if not, a sibling; then another relative; then a friend. The resulting sample consisted of 298 men and 416 women aged 15–81 ($M = 36.10$, $SD = 14.08$ years).

2.2. Instrument and procedure

The NEO-PI-R is a 240-item measure of the five basic personality factors or domains: Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness

[1] These are the only studies from Table 1 that report facet data from the full NEO-PI-R for self/other agreement.

(A), and Conscientiousness (C). Each factor is represented by six 8-item facet scales. Items are answered on a five-point Likert scale from *strongly disagree* to *strongly agree*; scales are balanced to control effects of acquiescence. Form S is the self-report version; a parallel, third-person Form R is used for observer ratings. Evidence on the reliability and validity of the instrument is given elsewhere (Costa & McCrae, 1992; Costa, McCrae, & Jónsson, 2002).

A bilingual Russian student worked with an American psychologist to develop an initial Russian draft of the NEO-PI-R. This draft was then reviewed and modified by a team of Russian and American psychologists. A translator unfamiliar with the original English version conducted a back-translation examined by the authors of the original instrument. This draft was administered to 178 undergraduates and adults in Yaroslavl, Russia (Martin et al., 1997). Analysis of these data identified 64 items with item-facet and item-domain correlations below .20 as candidates for revision. The team met in person to review each of these items and to develop alternatives translations for them. An item analysis of a 278-item version of the NEO-PI-R administered to 350 university students and adults, again in the Yaroslavl Region of Russia, served as the foundation for selecting the best translation for inclusion in the final 240-item set that was used in the present study.

The internal consistency coefficients of the Russian version are quite similar to those of the original English version, as is the factor structure (Martin et al., 2002). Studies employing the final Russian NEO-PI-R have explored the equivalence of the Estonian and Russian translations (Konstabel, 1999), bilingual test–retest coefficients for the English and Russian versions (Simahodskaya, 2000), and one-year stability coefficients for the Russian translation (Martin et al., 2002). Bilingual test–retest coefficients for the five domains ranged from .85 to .93 for Estonian and Russian, and from .88 to .98 for English and Russian. Domain stability coefficients ranged from .65 to .89.

However, two items showed little or no correlation with their respective facets and domains in this study, and were excluded from the following analysis. They are items 88 and 208, both of which belong to the Values facet of the Openness domain. They back-translate from Russian as "While solving moral questions one should take religious principles into consideration" and "I think that by the age of 25 any person should have his or her own convictions," respectively. For reasons discussed elsewhere (Martin et al., 2002; Martin et al., 1997), these items, and indeed the facet more generally, have not performed well in the Russian context.

Data for this study were collected by six examiners in four cities: Yaroslavl, Ivanovo, Vladimir, and Vologda. These ancient cities of Central Russia lie within a 400 km radius north and east of Moscow. The examiners were psychologists who reside and teach in the city in which they recruited respondents; both examiners and respondents were paid modest stipends for their participation in this study. Examiners recruited individuals they knew and in turn solicited recommendations for additional contacts from respondents. Typically respondents filled out the questionnaires in their own homes under the supervision of an examiner. All respondents completed

self-reports and were rated by their sibling or spouse, so each respondent simultaneously served as a target and an informant.[2]

The Czech version of the NEO Five-Factor Personality Inventory (NEO-FFI) has been used in several research studies (Hřebíčková, 1997; Hřebíčková et al., 2002) and the Czech NEO-FFI has now been published (Hřebíčková & Urbánek, 2001). To extend this work to the full NEO-PI-R, the usual procedures for translation of the NEO-PI-R were used. Three independent translators translated the NEO-PI-R items from English into Czech; two were psychologists and one was a professional translator. The preliminary Czech version of the items was based on a discussion of these three translations. A back-translation into English was done by a translator unfamiliar with the NEO-PI-R and reviewed by the authors of the original version.

Data were collected with the help of undergraduate students of psychology who were paid for their work. The students were requested to ask their relatives, partners, friends, and acquaintances to take part in a psychological study designed to investigate the relationships between various personality questionnaires, or the convergence of personality ratings across self- and partner ratings. Subjects were not paid for their participation and all testing materials were answered at home. In cases where the participants expressed an interest in feedback, they were given *Your NEO Summary* (Costa & McCrae, 1992) in Czech translation.

3. Results and discussion

3.1. Russian findings

Table 2 reports basic psychometric characteristics of the Russian version of the NEO-PI-R. Coefficient αs for the five domains are uniformly high, ranging from .86 to .92 in self-reports and from .85 to .93 in observer ratings. Internal consistencies are lower for the 8-item facet scales, ranging from .45 to .82. The median α for Form S is .70, comparable to the .71 found in American data (Costa & McCrae, 1992). Median α for Form R is .72.

[2] Inadvertently, the Russian observer ratings were not made with a separate third-person version of the NEO-PI-R. Instead, respondents were instructed to use the same set of items they had completed as a self-report in describing their spouse or sibling, by substituting "He" or "She" for "I," etc., as they read the items. Although the instructions made it quite clear that the second administration was supposed to be an informant rating, it is possible that some people were confused by the format and instead provided a second self-report. To evaluate this possibility, we correlated domain scores from the first questionnaire completed with scores from the second. If both were mistakenly construed as self-reports, these would be retest correlations and should all exceed .80 (cf. Martin et al., 2002). Instead, the correlations ranged from .29 to .39, *Mdn* = .33. By comparison, the same analysis in the Czech data, where separate Form S and Form R versions were used, yielded correlations from .05 to .36, *Mdn* = .16. The slightly higher values in the Russian data suggest that a few respondents may have been confused by the wording of the items. Note that the effect of introducing this error into the data should be to lower cross-observer correlations. The Russian correlations can thus be construed as conservative estimates of self/other agreement.

Table 2
Psychometric characteristics of the Russian NEO-PI-R

NEO-PI-R scale	Self report (Form S)						Observer rating (Form R)						r_{SR}
		Principal component						Principal component					
	α	N	E	O	A	C	α	N	E	O	A	C	
N: Neuroticism	.90						.88						.51
E: Extraversion	.89						.90						.57
O: Openness	.86						.85						.47
A: Agreeableness	.86						.89						.48
C: Conscientiousness	.92						.93						.43
N1: Anxiety	.80	**.84**	.01	.05	.09	−.02	.77	**.83**	−.00	.09	.14	.05	.47
N2: Angry Hostility	.73	**.71**	.14	−.05	−.31	−.14	.73	**.65**	.11	−.14	**−.40**	−.21	.44
N3: Depression	.69	**.72**	−.09	.01	−.03	−.25	.63	**.71**	−.16	.05	.05	−.14	.38
N4: Self-Consciousness	.59	**.73**	−.10	−.06	.20	−.02	.55	**.69**	−.09	.01	.27	.09	.39
N5: Impulsiveness	.65	.39	.22	.26	−.29	**−.40**	.63	.30	.31	.14	−.30	**−.48**	.41
N6: Vulnerability	.70	**.78**	.04	−.02	.06	−.26	.69	**.75**	.06	−.03	−.02	−.34	.42
E1: Warmth	.72	.04	**.71**	.04	**.42**	.06	.74	.02	**.70**	.05	**.43**	.05	.51
E2: Gregariousness	.82	−.07	**.78**	.02	−.00	−.08	.82	−.06	**.79**	−.06	−.03	−.16	.50
E3: Assertiveness	.72	−.06	**.65**	.03	−.25	.21	.74	−.15	**.67**	.06	−.29	.18	.46
E4: Activity	.69	.08	**.68**	.06	−.06	.22	.72	.03	**.66**	.05	−.06	.23	.38
E5: Excitement Seeking	.64	−.07	**.50**	.26	−.34	−.17	.61	−.05	**.52**	.19	−.31	−.27	.46
E6: Positive Emotions	.73	.06	**.72**	.23	.12	−.04	.72	.07	**.72**	.20	.11	−.09	.57

O1: Fantasy	.75	.10	.03	**.66**	−.16	−.27	.74	.18	.13	**.60**	−.07	−.31	.37
O2: Aesthetics	.77	.13	.07	**.74**	.15	.15	.78	.19	.07	**.72**	.17	**.14**	.48
O3: Feelings	.65	.29	.26	**.68**	.01	.00	.65	.36	.35	**.56**	.00	.01	.42
O4: Actions	.58	−.16	.32	**.50**	.00	−.25	.52	−.17	.39	**.43**	−.03	−.20	.39
O5: Ideas	.70	−.11	.02	**.67**	−.13	.16	.75	−.05	−.08	**.69**	−.08	.19	.40
O6: Values	.48	−.26	.00	**.60**	.12	−.13	.47	−.22	.06	**.54**	.10	.04	.40
A1: Trust	.73	−.21	.12	.14	**.69**	.01	.77	−.06	.18	.10	**.68**	.02	.39
A2: Straightforwardness	.72	.09	−.13	−.15	**.70**	.14	.72	.07	−.23	−.08	**.67**	.16	.43
A3: Altruism	.65	.03	.20	.18	**.64**	.30	.69	.11	.19	.19	**.71**	.25	.41
A4: Compliance	.60	−.14	−.27	−.00	**.71**	.04	.64	−.15	−.19	.02	**.75**	.04	.36
A5: Modesty	.67	.12	−.14	−.30	**.56**	.01	.70	.09	−.14	−.18	**.63**	.05	.43
A6: Tender-Mindedness	.45	.13	.14	.06	**.64**	.04	.54	.24	.09	.14	**.65**	.08	.39
C1: Competence	.63	−.29	.12	.08	−.01	**.74**	.60	−.19	.10	.08	−.01	**.79**	.33
C2: Order	.75	−.01	−.01	−.04	.08	**.77**	.80	.05	−.07	−.04	.12	**.77**	.42
C3: Dutifulness	.68	.02	−.04	.00	.31	**.75**	.74	.13	−.02	.03	.33	**.77**	.43
C4: Achievement Striving	.72	−.15	.29	.07	−.15	**.73**	.74	−.09	.24	.22	−.12	**.74**	.42
C5: Self-Discipline	.79	−.17	.10	−.06	.12	**.82**	.82	−.07	.02	−.01	.14	**.85**	.39
C6: Deliberation	.76	−.19	−.16	−.19	.15	**.69**	.77	−.20	−.21	−.04	.14	**.72**	.36
Factor congruence[a]		.96	.94	.94	.97	.97		.93	.94	.95	.96	.95	

Note. $N = 800$. The table reports Varimax-rotated factor loadings; loadings $\geqslant 0.40$ in absolute magnitude are given in boldface. r_{SR} = correlation between self-reports and observer ratings; all are significant at $p < .001$.

[a] Congruence with American normative (self-report) factor structure (Costa & McCrae, 1992).

Table 2 also includes results of Varimax-rotated principal components analyses of the 30 facet scales. Despite the low internal consistencies of O6: Values and A6: Tender-Mindedness, these scales define the intended factor in both self-reports and observer ratings. Indeed, with the exception of N5: Impulsiveness, which loads chiefly on the C factor,[3] all facets have their primary loading on the intended factor, and the pattern of secondary loadings is similar to that seen in American samples. Factor congruence coefficients comparing these factors to the American normative factors ranged from .93 to .97, well above the .90 usually interpreted as evidence of replication.

When data from men and women were factored separately, the expected structure was found in both self-reports and observer ratings, with congruence coefficients (compared to the American structure) ranging from .91 to .97. Gender differences in self-reports closely paralleled those found pan-culturally (Costa et al., 2001). Gender differences as perceived by observers were less marked, but similar in nature: Women were seen to be significantly higher in Anxiety, Vulnerability, Warmth, Gregariousness, Activity, Positive Emotions, Aesthetics, Feelings, Actions, Altruism, Tender-Mindedness, Dutifulness, and Self-Discipline; men were seen as significantly higher in E5: Excitement Seeking. A conspicuous difference across forms was seen for Impulsiveness, which was higher for women in self-reports, but for men in observer ratings (both $p < .05$).

Together, these analyses suggest that the Russian translation of the NEO-PI-R closely parallels the American version in psychometric qualities, and that an examination of cross-observer agreement should be a reasonable test of the generalizability of American findings on consensual validation. The last column of Table 2 presents correlations between Form S and Form R scales. Values for the five domains are very similar to those seen in Table 1 in both the US and other cultures, and somewhat higher than those reported in the small Russian pilot study. At one level, these data provide cross-cultural evidence for the consensual validity of personality traits; at another, they attest to the convergent validity of NEO-PI-R scales in Russian translation. The full 5×5 correlation matrix also shows evidence of discriminant validity: The median absolute discriminant validity coefficient is .06; the largest, between self-reported A and observer rated C, is .19.

Convergent correlations for the individual facet scales ranged from .33 to .57 with a median of .42. These values are somewhat higher than American peer/peer agreements, but less than American self/spouse correlations, where the median was .54 (Costa & McCrae, 1992). To assess the relative importance of trait and method variance in these data, a joint factor analysis of the 60 facets from Forms S and R was conducted (cf. McCrae & Costa, 1983). It showed five clear factors; congruence coefficients with the American normative Form S structure (doubled to form a 5×60 target matrix) ranged from .87 for O to .96 for A.

[3] In cross-cultural studies, N5: Impulsiveness typically has loadings on N, E, and (low) C factors, with the relative magnitude of these three loadings varying in seemingly arbitrary ways (see Konstabel, Realo, & Kallasmaa, 2002).

Table 3
Cross-observer correlations for NEO-PI-R domains within subgroups

Subgroup	*N*	NEO-PI-R domain				
		N	**E**	**O**	**A**	**C**
Russians						
Target						
Men	387	.53	.62	.52	.53	.45
Women	413	.50	.52[a]	.42	.43	.41
Adolescents	249	.40	.44	.33	.38	.28
Adults	551	.58[b]	.63[b]	.53[b]	.51[b]	.52[b]
Rater						
Spouse	634	.56	.62	.51	.46	.47
Sibling	166	.34[c]	.40[c]	.34[c]	.46	.29[c]
Czechs						
Target						
Men	298	.50	.62	.61	.54	.54
Women	416	.48	.63	.68	.52	.50
Adolescents	43	.24n.s.	.61	.62	.55	.64
Adults	662	.51	.62	.65	.52	.50
Rater						
Spouse	281	.48	.61	.59	.56	.44
Sibling	55	.21n.s.[c]	.61	.57	.63	.48
Other relative	67	.66[d]	.71	.65	.42	.39
Friend	301	.50[e]	.60	.61	.52	.60[f,g]

Note. Adolescents were age 15–22; adult were age 23+. Significant difference between correlations for: [a]men and women; [b]adolescents and adults; [c]spouse- and sibling raters; [d]sibling- and other relative raters; [e]sibling- and friend raters; [f]spouse- and friend raters; [g]other relative- and friend raters.

The top panel of Table 3 lists cross-observer correlations for the five domains within subsamples. Men and women, adolescents and adults, and those rated by spouses and siblings all show significant cross-observer agreement. Correlations are somewhat higher for men—significantly so in the case of E—perhaps because men are more accurate in self-reports, or perhaps because their spouses are more perceptive raters. Correlations for adults and for those rated by spouses are almost uniformly and substantially higher than correlations for adolescents and for those rated by siblings. These two variables, however, are confounded: Only a third of adolescents had spouse raters, and less than 1% of adults were rated by siblings. Thus, it is unclear whether age or type of rater accounts for this difference.

3.2. Czech findings

Table 4 reports psychometric characteristics of the Czech NEO-PI-R. Domain coefficient αs ranged from .88 to .93 for the two forms. With the exception of Openness to Values, which showed very low internal consistency in both self-reports and observer ratings, facet αs were adequate, ranging from .52 to .84. The median facet

Table 4
Psychometric characteristics of the Czech NEO-PI-R

NEO-PI-R scale	Self-report (Form S)						Observer rating (Form R)						r_{SR}
		Principal component						Principal component					
	α	N	E	O	A	C	α	N	E	O	A	C	
N: Neuroticism	.91						.90						.50
E: Extraversion	.90						.89						.62
O: Openness	.89						.91						.65
A: Agreeableness	.88						.91						.53
C: Conscientiousness	.91						.93						.51
N1: Anxiety	.79	**.86**	−.09	.04	.00	.00	.80	**.86**	−.04	.03	.02	.14	.51
N2: Angry Hostility	.75	**.73**	.05	−.03	**−.43**	−.06	.78	**.63**	.08	−.13	**−.57**	−.03	.47
N3: Depression	.80	**.81**	−.18	.06	.01	−.22	.77	**.81**	−.18	.10	.05	−.17	.46
N4: Self-Consciousness	.68	**.73**	−.25	−.10	.07	−.09	.68	**.74**	−.27	−.05	.11	.00	.43
N5: Impulsiveness	.71	.38	**.45**	.27	−.23	**−.41**	.72	.24	**.45**	.20	−.38	−.37	.45
N6: Vulnerability	.66	**.82**	−.04	.02	.08	−.24	.65	**.80**	.02	−.04	−.07	−.27	.45
E1: Warmth	.74	−.08	**.71**	.15	**.44**	.13	.77	−.10	**.66**	.23	**.49**	.10	.48
E2: Gregariousness	.80	−.10	**.76**	−.03	.03	−.10	.80	−.09	**.74**	.04	.01	−.10	.59
E3: Assertiveness	.80	−.37	**.55**	.14	−.36	.24	.79	−.37	**.52**	.13	−.38	.33	.56
E4: Activity	.59	−.10	**.51**	.15	−.25	.26	.60	−.02	**.50**	.10	−.12	**.41**	.52
E5: Excitement Seeking	.73	−.13	**.50**	.21	−.33	−.33	.73	−.17	**.52**	.28	−.32	−.30	.58
E6: Positive Emotions	.78	−.23	**.68**	.29	.08	−.02	.78	−.20	**.68**	.34	.18	−.02	.47

O1: Fantasy	.81	.16	.15	**.68**	−.07	−.28	.82	.13	.16	**.73**	−.01	−.37	.54
O2: Aesthetics	.79	.15	.05	**.76**	.19	.06	.84	.22	.11	**.76**	.12	.05	.62
O3: Feelings	.77	.30	**.42**	**.65**	.08	.00	.78	.33	**.40**	**.64**	.14	.08	.48
O4: Actions	.68	−.17	.25	**.55**	.00	−.17	.70	−.11	.21	**.65**	.03	−.09	.53
O5: Ideas	.76	−.13	−.07	**.82**	−.08	.09	.81	−.16	−.11	**.80**	−.06	.18	.56
O6: Values	.29	−.28	.13	**.44**	.18	−.07	.33	−.20	.14	**.53**	.19	.08	.33
A1: Trust	.75	−.23	.20	.15	**.59**	−.01	.80	−.12	.34	.12	**.62**	−.07	.45
A2: Straightforwardness	.77	.06	−.14	−.06	**.73**	.19	.80	.02	−.16	.04	**.73**	.19	.43
A3: Altruism	.72	−.04	.31	.08	**.68**	.31	.79	.01	.25	.13	**.78**	.25	.42
A4: Compliance	.69	−.23	−.19	.00	**.73**	.06	.77	−.23	−.15	.05	**.79**	.04	.46
A5: Modesty	.75	.27	−.26	−.32	**.54**	−.02	.82	.22	−.25	−.15	**.67**	.07	.47
A6: Tender-Mindedness	.52	.14	.01	.23	**.60**	−.08	.60	.22	.14	.17	**.66**	.09	.36
C1: Competence	.67	**−.46**	.18	.11	−.09	**.65**	.71	−.37	.06	.14	.02	**.74**	.39
C2: Order	.70	.08	−.05	−.15	.07	**.69**	.79	.19	−.05	−.08	.12	**.78**	.59
C3: Dutifulness	.77	−.04	−.07	−.13	.32	**.79**	.79	.09	−.06	−.10	.36	**.80**	.51
C4: Achievement Striving	.72	−.12	.26	.17	−.20	**.74**	.74	−.09	.18	.17	−.13	**.78**	.42
C5: Self-Discipline	.79	−.27	.04	−.11	.12	**.79**	.83	−.08	−.02	−.08	.18	**.85**	.48
C6: Deliberation	.78	−.31	−.34	−.08	.15	**.58**	.81	−.28	−.35	−.02	.23	**.66**	.40
Factor congruence[a]		.97	.96	.96	.97	.96		.94	.95	.96	.97	.96	

Note. Ns = 1365 self-reports, 908 peer-ratings. The table reports Varimax-rotated factor loadings; loadings $\geqslant 0.40$ in absolute magnitude are given in boldface. r_{SR} = correlation between self-reports and single observer ratings; all are significant at $p < .001$, $N = 714$.

[a] Congruence with American normative (self-report) factor structure.

α for Form S is .75; the median for Form R is .78. These values are somewhat higher than those seen in American and Russian data.

Exploratory factor analyses of both versions show clear replications of the American structure. Only N5: Impulsiveness fails to load chiefly on its intended factor. In this translation, Impulsiveness is related chiefly to Extraversion. Factor congruences (compared to the American structure) ranged from .94 to .97. Congruences were also high between the Russian and Czech versions, ranging from .95 to .99 for Form S and from .96 to .98 for Form R. In the Czech data, factor structures in men and women were virtually identical, with congruence coefficients ranging from .97 to .99. All these findings are consistent with lexical studies that recovered the five factors in Czech adjectives (Hřebíčková, 1997).

Gender differences in this Czech sample closely resembled the pan-cultural pattern. In both self-reports and observer ratings, women scored significantly higher than men in Neuroticism, Openness, and Agreeableness domains, and in Anxiety, Angry Hostility, Depression, Self-Consciousness, Vulnerability, Warmth, Gregariousness, Positive Emotions, Openness to Aesthetics and Feelings, Trust, Straightforwardness, Altruism, Modesty, Tender-Mindedness, Order, and Dutifulness. Men scored consistently higher in Assertiveness, Excitement Seeking, and Openness to Ideas. In the Czech sample, Impulsiveness showed no difference on either Form.

Data on consensual validation are given in the last column of Table 4. Correlations for the five domains range from .50 to .65, values somewhat higher than typically seen in American or other studies. Examination of the full 5×5 correlation matrix shows evidence of discriminant validity: The largest discriminant correlation, between self-reported Openness and observer rated Extraversion, is .27; the median absolute discriminant correlation is .10. Agreement at the facet level is also higher than usual, with a median cross-observer correlation of .47. At $r = .33$, cross-observer agreement equals or exceeds internal consistency for Openness to Values.

After Varimax rotation, five factors from a joint analysis of the 60 Form S and Form R facet scales partially reproduced the intended structure. Clear joint N, A, and C factors were found, with congruence coefficients (compared to the American self-report structure) of .92, .97, and .92, respectively. An O factor was defined by 11 of the 12 O facet scales, but Form R E1:Warmth, E5: Excitement Seeking, and E6: Positive Emotions also loaded on the factor, lowering the congruence coefficient to .82. The E factor (congruence = .68) was defined solely by scales from Form S.[4]

The bottom panel of Table 3 lists cross-observer correlations for the five domains within subsamples. Correlations are uniformly high, except in the case of N, where adolescents and those rated by siblings show lower agreement. In contrast to the Russian findings, sibling raters do not show consistently lower agreement than spouse raters. This might be because in the Czech sample, age and type of rater are not confounded; the mean age of those rated by siblings is 30.2 years. However,

[4] Complete factor loading matrices for the Russian and Czech joint factor analyses are available from the first author.

age is not the explanation for the different findings, because Russians show strong differences between adolescent and adult samples, whereas Czechs do not.

All cross-observer correlations examined so far have been between self-reports and single observers. The Czech consensus study permits an examination of agreement between informant ratings of the same target and an analysis of the effects of aggregation. In the consensus study, three informants rated each of 49 targets. Informant/informant agreement, expressed as intraclass correlations between pairs of raters, were .49, .55, .71, .44, and .47 for N, E, O, A, and C, respectively. These values are comparable in magnitude to those seen for self-other agreement in the Czech sample, and are somewhat higher than values reported by Costa and McCrae (1992) for peer/peer agreement on American targets.

In this subsample, the correlations of single informants with self-reports were .43, .50, .43, .35, and .48, respectively, for N, E, O, A, and C. The principle of aggregation suggests that correlations of mean informant rating with self-reports should be higher, and they are, with corresponding values of .53, .60, .48, .45, and .57.

3.3. *Differential consensual validity*

To compare the rank-ordering of the magnitude of self/other agreement for the 30 NEO-PI-R facets, we calculated Spearman correlations among cross-observer correlations from Tables 2 and 4, from the American NEO-PI-R manual (Costa & McCrae, 1992), and from German (Ostendorf & Angleitner, in press) and Chinese (Yang et al., 1999) studies. These are reported in Table 5. The data provide some support for the hypothesis that differential consensual validity is generalizable across cultures: 13 of the 15 correlations are positive, and five are statistically significant. A closer look at the data, however, shows that the pattern among Chinese respondents does not match the others: None of the correlations approaches significance, and two are negative. This may be due to the fact that the targets in this study were psychiatric patients, or it may reflect cultural differences between East and West.

Table 5
Rank-order correlations among self/other correlations across 30 NEO-PI-R facets from five cultures

	1.	2.	3.	4.	5.
1. American[a]					
2. American[b]	.31				
3. German	.30	.53**			
4. Chinese	.10	−.13	.28		
5. Russian	.43*	.05	.33	.14	
6. Czech	.20	.55**	.63**	−.02	.43*

Note. Data taken from Tables 2 and 4 and from Costa and McCrae (1992); Ostendorf and Angleitner (in press); and Yang et al. (1999).

[a] Self/spouse agreement.

[b] Self/peer agreement.

* $p < .05$.

** $p < .01$.

If data from the four Western cultures are pooled, it is possible to identify the NEO facets that show most and least self/other agreement. Six facets have mean correlations greater than .50: N1: Anxiety, E2: Gregariousness, E3: Assertiveness, E4: Activity, E5: Excitement Seeking, and O2: Aesthetics. Consistent with earlier work, it appears that traits related to Extraversion have the strongest self/other agreement. Eight traits have mean correlations less than .40: N4: Self-Consciousness, N6: Vulnerability, A1: Trust, A5: Modesty, C1: Competence, C3: Dutifulness, C5: Self-Discipline, and C6: Deliberation. It is possible that relatively poor agreement on Conscientiousness facets is due to the high evaluativeness of those traits.

It is possible that cross-cultural generalizability of consensual validity is due to generalizability of reliability: Some facets may consistently show higher cross-observer agreement simply because they are consistently more reliable. To test this, internal consistency coefficients for Form S and Form R were examined for American, German, Chinese (Form S only), Russian, and Czech NEO-PI-R facets. The nine columns of 30 reliability coefficients were intercorrelated, yielding 36 rank-order correlations (e.g., American Form R reliabilities with Chinese Form S reliabilities), of which 34 were significant. We therefore averaged the coefficients across the nine columns. The alpha reliability of this composite was .92. Thus, there is substantial agreement on the differential internal consistency of NEO-PI-R facets. Openness to Actions and Values and Tender-Mindedness had the lowest reliability, as they have had in studies in other cultures including Malaysia (Mastor, Jin, & Cooper, 2000) and the Philippines (Church & Katigbak, 2002); Anxiety and Self-Discipline had the highest mean αs. However, mean internal consistency was not related to mean cross-observer correlation, Spearman $r = .20$, n.s. Thus, the sheer reliability of the facet scales does not appear to account for their differential validity.

It is essential to recall in considering these analyses that the traits in question are measured only by a single instrument. For example, we do not know if the low internal consistency of Openness to Values is an intrinsic feature of the trait, or merely a quirk of the items used in the NEO-PI-R to measure it. Similarly, the consistently higher cross-observer validity of Openness to Aesthetics may suggest that that trait is measured particularly well by the NEO-PI-R. However, the findings for facets of Extraversion probably have a substantive interpretation, because similar results have been reported with different instruments.

4. Comments and conclusions

This article added two cultures to the small list in which the consensual validity of traits in the FFM has been studied. In Russia and the Czech Republic, as in the others, all five factors showed clear evidence of interjudge agreement. Judging by their meaningful responses to personality questionnaire items, it appears that individuals in all these cultures include trait information in their self-concepts and in their social perceptions, which form the basis of self-reports and observer ratings, respectively. The convergence of these two when single targets are assessed by two or more judges

can most parsimoniously be interpreted as evidence of the existence and operation of personality traits.

That conclusion is no longer in dispute when applied to the US and other individualistic Western cultures. But the present review and data suggest that it is equally true for people raised in collectivistic cultures, such as South Korea and China, and in once-collectivized nations, such as Russia and the Czech Republic. Individuals in all these cultures not only have traits—a consequence of the shared human genome and the heritability of personality (Bouchard & Loehlin, 2001; Jang, McCrae, Angleitner, Riemann, & Livesley, 1998)—but also attend to them and form consistent and more-or-less accurate assessments of them in self and others. The extreme social constructivist view that in some cultures, the person is conceived not in terms of abstract traits but only in terms of relations to others (see KağıtçıbaşBerry & Berry, 1989), does not appear to apply to these cultures, and may in fact be relevant, if at all, only to preliterate cultures.[5]

That conclusion must be qualified by two considerations. First, the data were not from probability samples, and may not fully represent their cultures. For example, South Koreans studying in the US (Spirrison & Choi, 1998) may be more Westernized and less collectivistic that the typical South Korean. Second, the new cultures examined here—Russia and the Czech Republic—score low on individualism using Hofstede's (2001) ratings, but might score higher on other indicators of that construct (see Oyserman, Coon, & Kemmelmeier, 2002). Certainly the same results would be more persuasive to many readers if the samples had been from rural Pakistan or Confucian Taiwan.

The present study also provided a cross-cultural perspective on a more differentiated understanding of interjudge agreement. In a review of previous studies, it appeared that the factor on which agreement was highest was Extraversion. That finding was supported in the Russian data, where the highest cross-observer correlation was again for Extraversion, and partially supported in the Czech data, where the correlation for Extraversion was second only to that for Openness. In particular, it appears that the facets of Gregariousness, Assertiveness, Activity, and Excitement Seeking are easily rated by observers, presumably because they involve easily observed behaviors. This pattern was not seen in Chinese data, however, so at present it is not clear whether this phenomenon is generalizable beyond Western cultures.

Finally, there was some evidence that spouse raters are generally superior to others as sources of personality data, at least as judged by agreement with self-reports. This is seen in American data (Costa & McCrae, 1992) and replicated in Russian data, with significantly larger correlations for spouse raters than for sibling raters for four of the five domains. That trend was not seen in Czech data, however. Further, in the Russian data, the nature of the rater was confounded with the age of the

[5] One ethnologist (Bonvillain, 2001) stressed the importance of both social relations and traits in preliterate aboriginal Americans: "People were expected to participate in communal activities, to give economic and ceremonial support to relatives, and to respect each other's autonomy. In general, generosity, even temper, and cooperativeness were highly valued personality traits whereas anger, stinginess, pride, and acquisitiveness were considered shameful attributes" (p. 3).

target. In future studies of rater effects, multiple raters (say, spouse, sibling, friend) should be obtained for the same target to control for target characteristics.

The present study adopted an etic approach to personality assessment, in which an instrument developed in the US was imported to other cultures. Data from many such studies (McCrae & Allik, 2002) make it clear that this is a viable approach to cross-cultural personality research: Traits of the FFM show very similar patterns of structure, development, reliability, and validity across a wide range of cultures. This does not mean, however, that traits serve identical functions in all societies. Lexical studies (Peabody & De Raad, 2002; Saucier, Hampson, & Goldberg, 2000) suggest that the language of personality traits varies somewhat across cultures, suggesting that some trait constructs are more salient in some cultures than in others. The details of how universal traits, unique trait languages, and self- and other perceptions interact across cultures remain to be worked out.

References

Amelang, M., & Borkenau, P. (1982). Über die faktorielle Struktur und externe Validität einiger Fragebogen-Skalen zur Erfassung von Dimensionen der Extraversion und emotionalen Labilität [On the factor structure and external validity of some questionnaire scales measuring dimensions of extraversion and neuroticism]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 3*, 119–146.

Asendorpf, J. B., & Ostendorf, F. (1998). Is self-enhancement healthy? Conceptual, psychometric, and empirical analysis. *Journal of Personality and Social Psychology, 74*, 955–966.

Barbaranelli, C., & Caprara, G. V. (2002). Studies of the Big Five questionnaire. In B. De Raad, & M. Perugini (Eds.), *Big Five assessment* (pp. 107–128). Göttingen: Hogrefe & Huber Publishers.

Barbaranelli, C., Caprara, G. V., Rabasca, A., & Pastorelli, C. (2003). A questionnaire for measuring the Big Five in late childhood. *Personality and Individual Differences, 34*, 645–664.

Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research.* Springfield, IL: Charles C Thomas.

Block, J. (1977). Advancing the psychology of personality: Paradigmatic shift or improving the quality of research? In D. Magnusson & N. S. Endler (Eds.), *Personality at the cross-roads: Current issues in interactional psychology* (pp. 37–64). Hillsdale, NJ: Erlbaum.

Bonvillain, N. (2001). *Native nations: Cultures and histories of Native North Americans.* Upper Saddle River, NJ: Prentice-Hall.

Borkenau, P., & Liebler, A. (1995). Observable attributes as manifestations and cues of personality and intelligence. *Journal of Personality, 63*, 1–25.

Bouchard, T. J., & Loehlin, J. C. (2001). Genes, evolution, and personality. *Behavior Genetics, 31*, 243–273.

Church, A. T. (2000). Culture and personality: Toward an integrated cultural trait psychology. *Journal of Personality, 68*, 651–703.

Church, A. T., & Katigbak, M. S. (2000). Trait psychology in the Philippines. *American Behavioral Scientist, 44*, 73–94.

Church, A. T., & Katigbak, M. S. (2002). The Five-Factor Model in the Philippines: Investigating trait structure and levels across cultures. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 129–154). New York: Kluwer Academic/Plenum Publishers.

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. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources Inc.

Costa, P. T., Jr., McCrae, R. R., & Jónsson, F. H. (2002). Validity and utility of the Revised NEO Personality Inventory: Examples from Europe. In B. De Raad & M. Perugini (Eds.), *Big Five assessment* (pp. 61–77). Göttingen, Germany: Hogrefe & Huber.

Costa, P. T., Jr., 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.

Cross, S. E., & Markus, H. R. (1999). The cultural constitution of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 378–396). New York: Guilford.

Draguns, J. R., Krylova, A. V., Oryol, V. E., Rukavishnikov, A. A., & Martin, T. A. (2000). Personality characteristics of the Nentsy in the Russian Arctic. *American Behavioral Scientist, 44*, 126–140.

Epstein, S. (1977). Traits are alive and well. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 83–98). Hillsdale, NJ: Erlbaum.

Funder, D. C. (1980). On seeing ourselves as others see us: Self-other agreement and discrepancy in personality ratings. *Journal of Personality, 48*, 473–493.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652–670.

Funder, D. C., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology, 69*, 656–672.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.

Hřebíčková, M. (1997). *Jazyk a osobnost: Pětifaktorová struktura popisu osobnosti [Language and personality: Five-Factor structure of personality description]*. Brno: Vydavatelství Masarykovy univerzity a Psychologický ústav AV ČR.

Hřebíčková, M., Urbánek, T. (2001). *NEO pětifaktorový osobnostní inventář podle NEO Five-Factor Inventory* P.T. Costy a R.R. McCraee [*NEO Five-Factor Personality Inventory according to NEO Five-Factor Inventory by* P.T. Costa and R.R. McCrae]. Praha: Testcentum.

Hřebíčková, M., Urbánek, T., Čermák, I., Szarota, P., Ficková, E., & Orliká-Adamovová, 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.

Jang, K. L., McCrae, R. R., Angleitner, A., Riemann, R., & Livesley, W. J. (1998). Heritability of facet-level traits in a cross-cultural twin sample: Support for a hierarchical model of personality. *Journal of Personality and Social Psychology, 74*, 1556–1565.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61*, 521–551.

Kağıtçıbaşı, Ç., & Berry, J. W. (1989). Cross-cultural psychology: Current research and trends. *Annual Review of Psychology, 40*, 493–531.

Konstabel, K. (1999). *A bilingual retest study of the Revised NEO Personality Inventory: A comparison of Estonian and Russian versions*. Unpublished Master's Thesis, University of Tartu, Estonia.

Konstabel, K., Realo, A., & Kallasmaa, T. (2002). Exploring the sources of variations in the structure of personality traits across cultures. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 29–52). New York: Kluwer Academic/Plenum Publishers.

Kurtz, J. E., & Sherker, J. L. (2003). Relationship quality, trait similarity, and self-other agreement on personality traits in college roommates. *Journal of Personality, 71*, 21–48.

Lanning, K. (1994). Dimensionality of observer ratings on the California Adult Q-Set. *Journal of Personality and Social Psychology, 67*, 151–160.

Mastor, K. A., Jin, P., & Cooper, M. (2000). Malay culture and personality: A Big Five perspective. *American Behavioral Scientist, 44*, 95–111.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224–253.

Martin, T. A., Costa, P. T., Jr., Oryol, V. E., Rukavishnikov, A. A., & Senin, I. G. (2002). Applications of the Russian NEO-PI-R. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 253–269). New York: Kluwer Academic/Plenum Publishers.

Martin, T. A., Draguns, J. G., Oryol, V. E., Senin, I. G., Rukavishnikov, A. A., & Klotz, M. L. (August, 1997). Development of a Russian-language NEO-PI-R. Poster session presented at the 105th annual meeting of the American Psychological Association, Chicago.

McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology, 43*, 293–303.

McCrae, R. R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparisons. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 105–125). New York: Kluwer Academic/Plenum Publishers.

McCrae, R. R., & Allik, J. (Eds.). (2002). *The Five-Factor Model of personality across cultures*. New York: Kluwer Academic/Plenum Publishers.

McCrae, R. R., & Costa, P. T., Jr. (1983). Joint factors in self-reports and ratings: Neuroticism, Extraversion, and Openness to Experience. *Personality and Individual Differences, 4*, 245–255.

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). Different points of view: Self-reports and ratings in the assessment of personality. In J. P. Forgas & M. J. Innes (Eds.), *Recent advances in social psychology: An international perspective* (pp. 429–439). Amsterdam: Elsevier.

McCrae, R. R., Costa, P. T., Jr., Lima, M. P., Simões, A., Ostendorf, F., Angleitner, A., Marušić, 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., Stone, S. V., Fagan, P. J., & Costa, P. T., Jr. (1998). Identifying causes of disagreement between self-reports and spouse ratings of personality. *Journal of Personality, 66*, 285–313.

Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology, 4*, 681–691.

Ostendorf, F. (1990). *Sprache und Personlichkeitsstruktur: Zur Validität des Fünf-Faktoren-Modells der Persönlichkeit [Language and personality structure: Toward the validation of the Five-Factor Model of personality]*. Regensburg: S. Roderer Verlag.

Ostendorf, F., & Angleitner, A. (in press). *NEO-Persönlichkeitsinventar, revidierte Form, NEO-PI-R nach Costa und McCrae* [Revised NEO Personality Inventory, NEO-PI-R of Costa and McCrae]. Göttingen, Germany: Hogrefe.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128*, 3–72.

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.

Paunonen, S. V., & Ashton, M. C. (2002). The nonverbal assessment of personality: The NPQ and the FF-NPQ. In B. De Raad, & M. Perugini (Eds.), *Big Five assessment* (pp. 171–194) Göttingen, Germany: Hogrefe & Huber.

Peabody, D., & De Raad, B. (2002). The substantive nature of psycholexical personality factors: A comparison across languages. *Journal of Personality and Social Psychology, 83*, 983–997.

Perugini, M., & Ercolani, A. P. (1998). Validity of the Five-Factor Personality Inventory (FFPI): An investigation in Italy. *European Journal of Psychological Assessment, 14*, 234–248.

Piedmont, R. L. (1994). Validation of the NEO-PI-R observer form for college students: Toward a paradigm for studying personality development. *Assessment, 1*, 259–268.

Piedmont, R. L., Bain, E., McCrae, R. R., & Costa, P. T., Jr. (2002). The applicability of the Five-Factor Model in a Sub-Saharan culture: The NEO-PI-R in Shona. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 155–173). New York: Kluwer Academic/Plenum Publishers.

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.

Rolland, J.-P. (2002). Cross-cultural generalizability of the Five-Factor Model of personality. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model of personality across cultures* (pp. 7–28). New York: Kluwer Academic/Plenum Publishers.

Saroglou, V., & Fiasse, L. (2003). Birth order, personality, and religion: A study among young adults from a three-sibling family. *Personality and Individual Differences, 35*, 19–29.

Saucier, G., Hampson, S. E., & Goldberg, L. R. (2000). Cross-language studies of lexical personality factors. In S. E. Hampson (Ed.), *Advances in personality psychology* (Vol. 1, pp. 1–36).

Shweder, R. A. (1975). How relevant is an individual difference theory of personality? *Journal of Personality, 43*, 455–484.

Simahodskaya, Z. (2000, August). *Russian revised NEO-PI-R: Concordant validity and relationship to acculturation.* Poster session presented at the 108th Annual Meeting of the American Psychological Association, Washington, DC.

Spirrison, C. L., & Choi, S. (1998). Psychometric properties of a Korean version of the Revised NEO Personality Inventory. *Psychological Reports, 83*, 263–274.

Szarota, P., Zawadski, B., & Strelau, J. (2002). Big Five domain and gender as determinants of rater agreement: A comparison based on self- and peer-rating on the Polish Adjective List. *Personality and Individual Differences, 33*, 1265–1277.

Woodruffe, C. (1985). Consensual validation of personality traits: Additional evidence and individual differences. *Journal of Personality and Social Psychology, 48*, 1240–1252.

Xiu, S., Wu, Z., Wu, Z., & Shui, C. (1996). Study of age differences on personality features of adults. *Psychological Science,* [in Chinese] 19, 1–5.

Yang, J., McCrae, R. R., Costa, P. T., Jr., Dai, X., Yao, S., Cai, T., & Gao, B. (1999). Cross-cultural personality assessment in psychiatric populations: The NEO-PI-R in the People's Republic of China. *Psychological Assessment, 11*, 359–368.

Zawadzki, B., Strelau, J., Szczepaniak, P., & Śliwińska, M. (1998). *Inwentarz Osobowości NEO-FFI Costy i McCrae (Adaptacja polska).* Warsaw: Pracownia Testów Psychologicznych.