Assessment
Volume 16 Number 3
September 2009 301-311
© 2009 SAGE Publications
10.1177/1073191109333760
http://asmnt.sagepub.com
hosted at
http://online.sagepub.com

Assessing the Universal Structure of Personality in Early Adolescence

The NEO-PI-R and NEO-PI-3 in 24 Cultures

Filip De Fruyt
Marleen De Bolle
Ghent University
Robert R. McCrae
Antonio Terracciano
Paul T. Costa Jr.
National Institute on Aging
Collaborators of the Adolescent Personality Profiles of Cultures Project

The structure and psychometric characteristics of the NEO Personality Inventory–3 (NEO-PI-3), a more readable version of the Revised NEO Personality Inventory (NEO-PI-R), are examined and compared with NEO-PI-R characteristics using data from college student observer ratings of 5,109 adolescents aged 12 to 17 years from 24 cultures. Replacement items in the PI-3 showed on average stronger item–total correlations and slightly improved facet reliabilities compared with the NEO-PI-R in both English- and non-English-speaking samples. NEO-PI-3 replacement items did not substantially affect scale means compared with the original scales. Analyses across and within cultures confirmed the intended factor structure of both versions when used to describe young adolescents. The authors discuss implications of these cross-cultural findings for the advancement of studies in adolescence and personality development across the lifespan.

Keywords: *adolescence; Five-Factor Model; cross-cultural; personality; observer ratings*

There is now substantial consensus on the structure of personality in adulthood, with the demonstration of the replicability of the dimensions of the Five-Factor Model (FFM) in self-ratings and observer ratings on the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) across a broad range of languages and cultures (McCrae & Costa, 1997; McCrae, Costa, del Pilar, Rolland, & Parker, 1998; McCrae, Terracciano, et al., 2005a). Moreover, there is also evidence that the NEO-PI-R can be validly administered to adolescents from 12 to 18 (De Fruyt, Mervielde, Hoekstra, & Rolland, 2000; McCrae et al., 2002), though not all NEO-PI-R items are considered optimal to assess personality in adolescents. The readability of some items for adolescents has been questioned, and other items are thought to refer to behavior that is infrequently observed in adolescents (De Fruyt et al., 2000).

Recently, the NEO-PI-3 (McCrae, Costa, & Martin, 2005) has been proposed as a more readable version of the NEO-PI-R, introducing 37 replacements for items that were considered problematic for use with adolescents. In an American sample, this modified instrument kept its intended factor structure and showed slightly increased readability, scale reliabilities, and cross-observer agreement (McCrae, Costa, et al., 2005). The present study addresses two questions: (a) Do the new NEO-PI-3 items maintain or improve the psychometric properties of the NEO-PI-R when translated into other languages? (b) Is the FFM structure of personality found across cultures[1] in adolescents aged 12 to 17? These are among the first questions that must be answered before the NEO-PI-3 can be recommended for use in early adolescence around the world.

Although the NEO-PI-R has long been used in college students, most of whom might be considered

late adolescents or emerging adults (Arnett, 1999), the factor structure of the NEO-PI-R has been replicated in early adolescents in only two cultures: the United States (Baker & Victor, 2003; Costa, McCrae, & Martin, 2008) and Belgium (De Fruyt et al., 2000). There is, however, a much larger body of data suggesting that the FFM factors in some form can be found in both children and adolescents (Digman, 1963; Goldberg, 2001; John, Caspi, Robins, Moffitt, & Stouthamer-Loeber, 1994; Mervielde, Buyst, & De Fruyt, 1995; Mervielde & De Fruyt, 2002). These studies, using a variety of inventories (e.g., adjective lists, Q-sort ratings), and using parents (Mervielde & De Fruyt, 2002) or teachers (Digman, 1963) as informants, convincingly demonstrated that the Big Five can be used to describe personality differences in childhood and young adolescence. Whereas specific item sets have sometimes been compiled for the assessment of younger children's personality to capture potential developmental differences (Mervielde & De Fruyt, 2002), adolescent personality has more frequently been assessed using measures that were primarily developed for adults, such as the NEO-PI-R or the brief NEO Five-Factor Inventory (Allik, Laidra, Realo, & Pullmann, 2004). Administering inventories designed for adults to adolescents enables researchers to conduct longitudinal studies using the same item sets across a broad time interval, but has the disadvantage that some items may be less well understood, especially when adolescents are asked to provide self-reports.

McCrae, Costa, et al. (2005) therefore identified 37 NEO-PI-R items (from 19 facets) for replacement in a more readable version of the instrument, the NEO-PI-3. Of these, 25 were selected because they included words that some high school–age respondents did not understand, such as *lackadaisical* and *panhandler*; 12 other items were selected because they showed poor (<.30) item–total correlations in both adolescent and adult samples. Keying was kept

Authors' Note: Collaborators of the Adolescent Personality Profiles of Cultures Project include the following: Maria E. Aguilar-Vafaie (Department of Psychology, Tarbiat Modares University, Tehran, Islamic Republic of Iran); Chang-kyu Ahn (Department of Education, Pusan National University, Pusan, South Korea); Hyun-nie Ahn (Department of Psychology, Ewha Womans University, Seoul, South Korea); Lidia Alcalay (Escuela de Psicologia, Pontificia Universidad Catolica de Chile, Santiago, Chile); Jüri Allik (Department of Psychology, University of Tartu, Tartu, Estonia); Tatyana V. Avdeyeva (Graduate School of Professional Psychology, University of St. Thomas, St. Paul, MN); Marek Blatný (Institute of Psychology, Academy of Science of the Czech Republic, Prague, Czech Republic); Denis Bratko (Department of Psychology, Faculty of Philosophy, Zagreb, Croatia); Marina Brunner-Sciarra (Universidad Peruana Cayetano Heredia, Lima, Peru); Thomas R. Cain (Rutgers University, New Brunswick, NJ); Niyada Chittcharat (Srinakharinwirot University, Bangkok, Thailand); Jarret T. Crawford (The College of New Jersey, Ewing, NJ); Ryan Fehr (Department of Psychology, University of Maryland, College Park, MD); Emília Ficková (Institute of Experimental Psychology, Slovak Academy of Sciences, Bratislava, Slovak Republic); Michele J. Gelfand (Department of Psychology, University of Maryland, College Park, MD); Sami Gulgoz (College of Arts and Sciences, Koç University, Istanbul, Turkey); Martina Hřebíčková (Institute of Psychology, Academy of Science of the Czech Republic, Prague, Czech Republic); Lee Jussim (Rutgers University, New Brunswick, NJ); Waldemar Klinkosz (The John Paul II Catholic University of Lublin, Lublin, Poland); Goran Knežević (Department of Psychology, Faculty of Philosophy, Belgrade University, Belgrade, Serbia); Nora Leibovich de Figueroa (Instituto de Investigaciones, Facultad de Psicologia, University of Buenos Aires, Buenos Aires, Argentina); Margarida P. Lima (University of Coimbra, Coimbra, Portugal); Corinna E. Löckenhoff (National Institute on Aging, Bethesda, MD); Thomas A. Martin (Susquehanna University, Selinsgrove, PA); Iris Marušić (Institute for Social Research, Zagreb, Croatia); Khairul Anwar Mastor (Universiti Kebangsaan Malaysia, Bangi Selangor, Malaysia); Katsuharu Nakazato (Faculty of Social Welfare, Iwate Prefectural University, Iwate, Japan); Florence Nansubuga (Institute of Psychology, Makerere University, Kampala, Uganda); Jose Porrata (San Juan, Puerto Rico); Danka Purić (Department of Psychology, Faculty of Philosophy, Belgrade University, Belgrade, Serbia); Anu Realo (Department of Psychology, University of Tartu, Tartu, Estonia); Norma Reátegui (Universidad Peruana Cayetano Heredia, Lima, Peru); Jean-Pierre Rolland (Université Paris Ouest Nanterre La Défense, Paris, France); Vanina Schmidt (Instituto de Investigaciones, Facultad de Psicologia, University of Buenos Aires, Buenos Aires, Argentina); Andrzej Sekowski (The John Paul II Catholic University of Lublin, Lublin, Poland); Jane Shakespeare-Finch (School of Psychology & Counselling, Queensland University of Technology, Brisbane, Australia); Yoshiko Shimonaka (Department of Human Studies, Bunkyo Gakuin University, Saitama, Japan); Franco Simonetti (Escuela de Psicologia, Pontificia Universidad Catolica de Chile, Santiago, Chile); Jerzy Siuta (Institute of Psychology, Jagiellonian University, Krakow, Poland); Barbara Szmigielska (Institute of Psychology, Jagiellonian University, Krakow, Poland); Vitanya Vanno (Srinakharinwirot University, Bangkok, Thailand); Lei Wang (Department of Psychology, Peking University, Beijing, China); Michelle Yik (The Hong Kong University of Science and Technology, Hong Kong).

Robert R. McCrae and Paul T. Costa, Jr., receive royalties from the NEO-PI-R. This research was supported in part by the NIH Intramural Research Program, National Institute on Aging. The Czech contribution was supported by grant 406/07/1561 from the Grant Agency of the Czech Republic. Address correspondence to Filip De Fruyt, Department of Developmental, Personality, and Social Psychology, H. Dunantlaan 2, B-9000 Gent, Belgium; e-mail: Filip.DeFruyt@ugent.be.

Table 1
Description of Samples and Raters

Culture	Language	*N*	Percentage Female	Mean Age	Percentage Native Citizens	Familiarity With Target[a]
Argentina	Spanish	220	88.1	14.78	99.0	2.84
Australia	English	209	76.6	14.80	96.1	3.35
Chile	Spanish	175	46.8	14.82	97.4	2.64
People's Republic of China	Chinese	188	73.6	14.77	99.4	2.56
Croatia	Croatian	203	74.5	14.84	93.6	2.59
Czech Republic	Czech	209	61.3	15.00	94.2	2.63
Estonia	Estonian	196	59.2	15.17	100.0	2.67
France[b]	French	270	—	—	—	2.85
Hong Kong	Chinese	183	41.3	14.94	86.0	2.29
Islamic Republic of Iran	Farsi	234	73.1	14.82	99.5	2.78
Japan	Japanese	314	81.3	15.14	100.0	1.88
Malaysia	Malay	251	66.8	15.01	100.0	3.06
Peru	Spanish	194	61.0	15.06	98.2	2.99
Poland	Polish	356	85.2	14.92	100.0	2.73
Portugal	Portuguese	204	87.1	14.78	92.9	2.91
Puerto Rico	Spanish	142	58.7	14.86	91.3	3.18
Russia	Russian	200	62.8	14.67	98.9	3.10
Serbia	Serbian	202	42.6	14.97	83.1	2.75
Slovak Republic	Slovakian	216	51.6	14.95	100.0	2.80
South Korea	Korean	153	41.8	14.91	99.3	2.36
Thailand	Thai	180	70.4	14.95	97.5	2.84
Turkey	Turkish	177	59.1	14.72	100.0	2.86
Uganda[c]	English	200	53.9	14.83	96.4	3.17
United States of America	English	634	60.5	14.81	87.8	2.81

a. Rated on a scale from 0 to 4, aggregated across three items assessing familiarity.
b. No descriptive data on the raters were available for France.
c. Administered in respondents' second language.

identical to the NEO-PI-R original. Self-reports (Form S) on the NEO-PI-R and the replacement items were subsequently collected on a sample of 500 American adolescents aged 14 to 20, supplemented with peer ratings (Form R) on targets of the same age range. The median item–total correlations across these 37 replacements increased from .28 in the NEO-PI-R to .37 in the NEO-PI-3 for Form S and from .30 to .42 for Form R. The median coefficient alpha across the 30 facets increased slightly for both self-ratings and peer ratings, whereas cross-observer agreement and the factor structure remained virtually the same. Costa et al. (2008) replicated these findings for the NEO-PI-3 in 449 American boys and girls aged 12 to 13, concluding that the NEO-PI-3 is a useful instrument for middle school–aged children.

However, these NEO-PI-3 findings were obtained from American samples of adolescents. NEO-PI-3 replacement items were primarily chosen to be more readable in English, and it remains to be seen whether they function well in translation. This article reports comparisons of Form R of the NEO-PI-R with Form R of the NEO-PI-3 in 18 different languages using undergraduates as raters and adolescents as targets.

The availability of age- and language-appropriate measures of specific traits allows a consideration of another question: Is the structure of personality traits the same in early adolescence as in adulthood, and is this structure universal, found in Asian, African, and South American cultures as well as American and European cultures? Data from 24 cultures are used to address this issue.

Method

Cultures

In March 2006, collaborators from the Personality Profiles of Cultures Project (PPOC; McCrae, Terracciano, et al., 2005a; McCrae, Terracciano, et al., 2005b) and other interested investigators were invited to join the Adolescent Personality Profiles of

Cultures Project (APPOC). One part of this project focused on the cross-cultural examination of the NEO-PI-3 for assessing adolescent personality; another (not dealt with in the present article) examined national character ratings of different age and sex groups. Data on the NEO-PI-3 were obtained from 27 different teams from 24 different cultures. An overview of the collaborating cultures is provided in Table 1. Ratings from multiple sites were available for the United States (three collaborating sites) and Poland (two collaborating sites), though data were grouped per country for the present analyses. Most cultures correspond to nations, but Hong Kong has been distinguished from the People's Republic of China, and Puerto Rico has been treated as an independent culture.

Participants, Procedures, and Targets

Research participants recruited to provide ratings of adolescents' personality were usually undergraduate students enrolled in a psychology major. A description of their characteristics is provided in Table 1. Raters ($N = 5,510$) were on average 21.3 years old ($SD = 3.9$; ranging from 15 to 66 years), with 64.7% females (ranging from 41.3% for Hong Kong to 88.1% for Argentina), reflecting the large proportion of females in undergraduate psychology programs. Overall, 95.6% were native-born citizens, ranging from 83.1% (Serbia) to 100% (Estonia, Japan, Malaysia, Poland, Slovak Republic, and Turkey). Participation was voluntary and anonymous and had been ruled exempt from institutional review board review in the United States.

Collaborators were requested to collect observer ratings of 50 boys and 50 girls aged 12 to 14 years and 50 boys and 50 girls aged 15 to 17 years. Questionnaires for each target category were randomly distributed among research participants. Data collection was done in groups, usually in undergraduate classes, often under the supervision of an APPOC collaborator. If needed, the collaborators answered questions about the meaning of words or items. If respondents did not understand an item, they were asked to leave it blank.

Participants received the following general instructions:

> This is a study of personality across cultures. We are interested in how people view others and rate their personality traits, and we will be comparing your responses to those of college students in other countries. Please think of a *boy* [*girl*] aged *12-14* [*15-17*] whom you know well. He [She] should be someone who is a native-born citizen of your country. He [She] can be a relative or a friend or neighbor—someone you like or someone you don't like.

The combined sample available for analysis across age and gender groups included ratings on 1,355 younger (12-14 years) adolescent females (24.6%), 1,314 younger adolescent males (23.8%), 1,440 older (15-17 years) adolescent females (26.1%), and 1,401 older adolescent males (25.4%). However, raters were asked to write down the age of the target individual they had in mind when providing ratings, and 215 of these ratings were for individuals outside the intended age range; these cases were dropped. The age distribution within the age categories shows a substantially larger proportion of target individuals toward the higher end of the targeted age categories, that is, for the group of 12 to 14 years, 9.9%, 14.1%, and 24.3% of the total sample were 12, 13, and 14 years, respectively, whereas for the age group 15 to 17, 8.7%, 15.0%, and 27.9% were 15, 16, and 17 years old, respectively.

On average, raters reported that they had known their targets for 9.2 years ($SD = 5.6$ years). Raters were also asked to indicate whether they described a relative (50.7%), a friend (25.3%), or an acquaintance (22.7%) and to rate their familiarity with the target on 5-point scales from *strongly disagree* (0) to *strongly agree* (4). Statements were translated, and an independent backtranslation was approved by the first author. A "familiarity with target" score was computed, aggregating scores across the items, "I know this person well," "I see this person often," and "I have interacted with this person in many contexts." This composite scale had a Cronbach α of .77 and a mean of 2.76 ($SD = .93$), suggesting strong familiarity with most targets. Familiarity ratings per culture are reported in the last column of Table 1.

Measures

NEO-PI-R/NEO-PI-3. The NEO-PI-R (Costa & McCrae, 1992) is among the most frequently used inventories to assess the FFM and its dimensions of Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. The inventory has 30 facets, organized under the five domains, and includes 240 items (8 items per facet), presented with a 5-point Likert response scale. The NEO-PI-3 (McCrae, Costa, et al., 2005) substitutes 37 items chosen to be more readable or more appropriate for assessment of the intended facet. In the present study,

replacement items were administered immediately after the 240 NEO-PI-R items. A transparent validity item (asking about honesty and accuracy) modeled on the NEO-PI-R validity check (Costa & McCrae, 1992) was added at the end.

Translations of the observer rating version (Form R) of the NEO-PI-R were available in all cultures, and collaborators were asked to translate the 37 replacement NEO-PI-3 Form R items into their native language and to obtain an independent backtranslation. These backtranslations were examined by the test author (RRM) and revised as needed.

The NEO-PI-R manual specifies that protocols are considered invalid if more than 40 items are missing, if respondents deny that they have responded honestly and accurately (the validity check item), or if there are repetitive strings of responses beyond specified cutoffs. Application of these rules to the 240 NEO-PI-R items led to the elimination of an additional 186 protocols, and a final sample size of 5,109.

Sample data quality index. McCrae, Terracciano, et al. (2005a) argued that an assessment of the quality of the data in each sample should precede analyses of the cross-cultural replicability of a measure, so deviations from strict replication can be interpreted in light of independent information on data quality. They proposed a data quality index for the NEO-PI-R, including six different indicators. A similar index was created for the present study, based on four indicators: the number of unscreened (i.e., valid plus invalid) protocols with more than 40 missing NEO-PI-R items, the mean score on the validity check item for the unscreened protocols, the proportion of missing items within the screened protocols (subsequently replaced by the neutral response), and the proportion of screened protocols with indications of yea- or nay-saying. Three items used in the PPOC data quality index that concerned presence or absence of problems noted by the administrator, whether participants were administered the inventory in their native language, and whether a published or an unpublished NEO-PI-R version was used showed restricted variance in this study and were discarded. The four data-quality indicators used here intercorrelated positively, showing a Cronbach alpha coefficient of .68. Each of these indicators was transformed into a rank score, and the mean rank was used as an indicator of quality for the sample. Item-level data were not available from Estonia because of variations in the questionnaire format, so a quality index could not be calculated for Estonia.

Results

Comparison of NEO-PI-R and NEO-PI-3

Corrected item–facet total correlations for NEO-PI-R and NEO-PI-3 replacement items and their facets' reliabilities are reported in Table 2 for the total sample and for English and non-English language administrations. Only the American and Australian data were included in the English analyses, because English was a second language for most Ugandans. Because most NEO-PI-3 replacement items were written to simplify the English wording, it might be hypothesized that the psychometric characteristics would be most affected in samples that were administered the English version.

In the English-speaking samples, the median item–total correlation increased from .35 (NEO-PI-R) to .47 (NEO-PI-3), and in non-English samples from .29 to .39, showing that replacements on average improved the item–total correlations across languages, with a minor extra gain for the English-speaking samples. A few replacement items produced slightly lower item–total correlations than their NEO-PI-R counterparts (e.g., N5 Item 21, .29 vs. .24), and neither the original E4 Item 17 nor its replacement performed well in these samples, but in general the NEO-PI-3 items appear to represent a modest improvement over the originals, even when translated into other languages.

As could be expected from the small number of changes, internal consistencies were only slightly affected at the domain level (αs = .81-.96), with all NEO-PI-3 values being the same or slightly higher than their NEO-PI-R counterparts. Table 2 compares internal consistencies for the 19 facets with different NEO-PI-R and NEO-PI-3 versions; in the total sample, 16 of these (84.2%) are as high or higher in the NEO-PI-3. The median alphas increase somewhat in both English- and non-English-speaking samples as a result of the item replacements. A substantial increase from .54 to .69 was observed for A6: Tender-Mindedness, with four replacement items. However, even after incorporating replacement items, the Cronbach alpha coefficients for O4: Actions and O6: Values facets remained very low, with coefficients of .48 and .34, respectively.

Mean differences. The NEO-PI-R has already been widely used in cross-cultural studies; is it appropriate to compare means from those studies to means based on the NEO-PI-3? To examine the effect of the

Table 2
Corrected Item–Total Correlations and Reliabilities for NEO-PI-3 Replacement Items

		Item–Total Correlation						Facet Reliability					
		Total Sample		English		Non-English		Total Sample		English		Non-English	
Facet	PI-R Item	PI-R	PI-3	PI-R	PI-3	PI-R	PI-3	PI-R	PI-3	PI-R	PI-3	PI-R	PI-3
N1	121	.26	.46	.27	.53	.26	.44	.69	.73	.77	.80	.67	.71
N4	16	.39	.42	.44	.59	.38	.39	.59	.64	.69	.76	.58	.61
N4	136	.35	.38	.52	.53	.31	.34						
N4	226	.22	.44	.19	.57	.23	.41						
N5	21	.29	.24	.39	.36	.28	.21	.62	.65	.64	.70	.62	.64
N5	81	.26	.42	.09	.45	.29	.42						
E3	42	.34	.29	.44	.33	.33	.28	.70	.69	.77	.76	.68	.67
E4	17	.09	.08	−.05	−.02	.11	.10	.62	.64	.59	.62	.62	.64
E4	47	.35	.43	.34	.42	.35	.43						
E5	202	.31	.43	.35	.47	.29	.42	.64	.67	.69	.71	.63	.66
E6	57	.40	.45	.42	.50	.39	.44	.73	.73	.77	.80	.72	.71
E6	147	.33	.29	.40	.53	.32	.25						
O2	8	.30	.55	.36	.63	.29	.54	.70	.74	.75	.78	.69	.74
O4	78	.13	.28	.03	.32	.13	.27	.43	.48	.43	.54	.42	.47
O4	168	.21	.25	.20	.33	.21	.23						
O6	118	.13	.16	.23	.28	.11	.14	.29	.34	.40	.45	.27	.31
O6	148	.13	.18	.25	.29	.11	.16						
O6	238	.11	.16	.12	.21	.11	.14						
A1	4	.37	.34	.52	.35	.35	.34	.75	.75	.83	.82	.73	.73
A2	219	.33	.32	.30	.27	.34	.32	.71	.70	.77	.77	.70	.69
A4	229	.45	.44	.55	.47	.43	.44	.73	.72	.80	.79	.71	.71
A5	114	.33	.44	.45	.58	.31	.41	.74	.75	.83	.84	.71	.73
A6	29	.20	.37	.25	.49	.19	.34	.54	.69	.62	.77	.53	.67
A6	59	.09	.48	.19	.57	.07	.47						
A6	119	.32	.51	.25	.58	.34	.50						
A6	149	.35	.31	.38	.40	.34	.30						
C1	5	.45	.52	.49	.53	.45	.53	.74	.78	.79	.82	.73	.77
C1	35	.28	.39	.39	.42	.27	.38						
C1	125	.29	.56	.41	.63	.27	.55						
C1	185	.51	.42	.50	.50	.52	.40						
C2	10	.25	.58	.21	.60	.25	.57	.73	.80	.74	.82	.73	.80
C2	70	.45	.58	.40	.62	.46	.57						
C2	160	.23	.30	.21	.25	.24	.31						
C3	105	.30	.36	.31	.45	.29	.34	.77	.78	.79	.81	.76	.77
C3	165	.40	.39	.42	.39	.39	.39						
C3	225	.48	.50	.50	.53	.47	.50						
C4	20	.26	.39	.00	.59	.31	.36	.77	.80	.79	.86	.77	.78
Median		.30	.39	.35	.47	.29	.39	.70	.72	.77	.78	.69	.71

Note: Facet scale labels given in Table 4. Cronbach alphas for the other facets, total sample, were N2, .76; N3, .73; N6, .74; E1, .76; E2, .74; O1, .68; O3, .60; O5, .79; A3, .76; C5, .83; C6, .80.

replacement items on scale means, NEO-PI-R and NEO-PI-3 domain and facet means were systematically compared. The results of a series of repeated-measures analyses of variance with accompanying effect sizes (Cohen, 1988) are reported in Table 3. For the majority of the proposed replacements, differences between the two versions were small to negligible, with the largest effects in the English samples found for O4: Actions and A6: Tender-Mindedness, both showing a mean of about one-quarter standard deviation higher for the NEO-PI-3. Version differences were generally smaller in the non-English-speaking sample, although this may conceivably mask larger effects in different directions in different languages.

Table 3
Mean-Level Differences Between PI-3 and PI-R Domain and Facet Scales

	Total Sample			Samples	Non-English Samples
	M				
	PI-3	PI-R	*d*	*d*	*d*
N	96.29	96.68	−.02***	−.04***	−.01***
E	116.27	116.06	.01***	.02***	.00*
O	104.61	103.91	.04***	.05***	.04***
A	102.30	102.98	−.03***	−.03***	−.03***
C	96.49	97.06	−.02***	.03***	−.03***
N1	16.33	16.25	.02***	.02**	.01**
N4	15.41	15.80	−.09***	−.18***	−.06***
N5	17.97	18.05	−.02**	−.03*	−.01*
E3	16.15	16.07	.02***	.05***	.01*
E4	17.84	17.87	*ns*	.05***	−.02**
E5	20.36	20.01	.07***	.05***	.08***
E6	20.92	21.14	−.04***	−.03***	−.05***
O2	16.57	16.83	−.05***	−.04***	−.05***
O4	15.65	14.74	.25***	.29***	.24***
O6	17.76	17.71	*ns*	*ns*	*ns*
A1	17.79	18.29	−.11***	−.10***	−.11***
A2	16.78	17.04	−.05***	−.07***	−.05***
A4	14.10	14.50	−.08***	−.12***	−.07***
A5	15.28	15.32	−.01*	−.06***	*ns*
A6	18.55	18.02	.13***	.30***	.09***
C1	17.44	17.61	−.03***	*ns*	−.04***
C2	14.90	15.16	−.04***	−.06***	−.04***
C3	16.92	17.34	−.08***	*ns*	−.09***
C4	17.42	17.15	.05***	.20***	.02***

Note: Facet scale labels given in Table 4; *ns* = nonsignificant difference; *d* = effect size.
*$p < .05$. **$p < .01$. ***$p < .001$.

NEO-PI-3 Factor Structure in Early Adolescence

To examine the factor structure of the NEO-PI-3 in early adolescence, we conducted a series of factor analyses. In each, five principal components were extracted and rotated toward the normative adult Form S NEO-PI-R structure (Costa & McCrae, 1992); factor, variable, and total congruence coefficients with this target structure (McCrae, Zonderman, Costa, Bond, & Paunonen, 1996) were then computed. Note that this is a very stringent test of replicability, because observer ratings of early adolescents in 24 cultures on the NEO-PI-3 are compared with self-reports of adult Americans on the NEO-PI-R; in previous studies, the two forms have shown a similar structure, although with Form R data, Agreeableness and Conscientiousness factors typically account for slightly more of the common variance (e.g., McCrae, Terracciano, et al., 2005a). Nevertheless, in the present sample the structure was closely replicated, with all factor congruence coefficients above .95 and all variable congruence coefficients above .87.

The effect of age on factor structure is of particular interest here, so this analysis was repeated within subgroups of targets aged 12 to 14 and 15 to 17 years. In these analyses, we standardized data within culture before combining the data across cultures. This procedure eliminates culture-level effects from the analyses of individual-level data. However, analyses of raw data yielded nearly identical results (cf. McCrae, Terracciano, et al., 2005a). For the original NEO-PI-R, congruence coefficients ranged from .95 (Conscientiousness) to .98 (Neuroticism) for the younger adolescents and from .96 (Openness to Experience, Agreeableness, and Conscientiousness) to .98 (Neuroticism) for the older adolescents. As Table 4 shows in detail for the NEO-PI-3, the same structure is found in both groups when principal components are rotated toward the American adult Form S structure, although the fit is slightly better in the older group. Despite the low coefficient alphas for Openness to Actions and Values, these variables clearly load on the Openness factor in both younger and older groups. Indeed, the only notable variation from the standard adult structure is the large secondary loading of O5: Ideas on the Conscientiousness factor.

It is possible that all these results are driven by the large number of targets from European and American cultures and that meaningful variations in structure would be seen in non-Western cultures. We therefore conducted analyses within each culture. Results are shown in Table 5. The first column reports the data quality index for the sample, and samples are arranged in increasing order of quality. Most of the lower quality data are from non-European cultures, where personality testing is probably less familiar to respondents. Note, however, that these are rankings of relative quality; all samples show reasonably high quality data in an absolute sense.

The next five data columns report Cronbach alphas for the NEO-PI-3 domains. These all were above .70, except for Openness to Experience in Puerto Rico, Uganda, and Malaysia. Coefficient alphas were strongly and significantly correlated with data quality across samples ($rs = .60$-$.74$; $N = 23$; $ps < .01$). These results suggest that, in general, lower indicators of internal consistency reflect test-taking artifacts rather

Table 4
Factor Loadings for Observer-Rated NEO-PI-3 Within-Cultures Standardized Facet Scales

	Age: 12-14 Years						Age: 15-17 Years					
NEO-PI-R Facet Scale	N	E	O	A	C	VCC	N	E	O	A	C	VCC
N1: Anxiety	**.80**	−.10	.03	.11	.07	.95	**.81**	−.12	.03	.10	.09	.95
N2: Angry Hostility	**.59**	−.02	−.02	**−.58**	−.14	.99	**.62**	−.01	−.07	**−.56**	−.13	.99
N3: Depression	**.76**	−.26	.02	.08	−.14	.96	**.78**	−.23	.02	.09	−.13	.97
N4: Self-Consciousness	**.71**	−.32	−.10	.15	−.05	.96	**.69**	−.33	−.11	.14	−.06	.96
N5: Impulsiveness	**.48**	.33	.05	−.36	−.39	.98	**.50**	.31	.06	−.37	−.39	.98
N6: Vulnerability	**.67**	−.09	−.13	−.08	**−.45**	.98	**.70**	−.06	−.11	−.04	**−.44**	.99
E1: Warmth	−.20	**.75**	.19	.34	.12	.99	−.20	**.73**	.18	.36	.13	.99
E2: Gregariousness	−.19	**.76**	.03	−.02	−.09	.99	−.17	**.77**	.06	.03	−.11	.99
E3: Assertiveness	−.32	**.49**	.17	**−.43**	.30	.99	−.29	**.50**	.18	**−.42**	.33	.99
E4: Activity	−.09	**.64**	.14	−.35	.15	.91	−.08	**.64**	.13	−.30	.24	.95
E5: Excitement Seeking	−.04	**.57**	.23	−.33	−.26	.94	−.09	**.56**	.18	−.33	−.27	.94
E6: Positive Emotions	−.13	**.76**	.25	.10	.00	.98	−.16	**.73**	.26	.11	.06	.98
O1: Fantasy	.20	.31	**.53**	.02	−.28	.95	.18	.22	**.58**	.02	−.30	.98
O2: Aesthetics	.15	.10	**.61**	.18	.30	.96	.16	.11	**.70**	.14	.23	.99
O3: Feelings	.27	**.45**	**.52**	.08	.19	.98	.25	**.48**	**.52**	.06	.21	.97
O4: Actions	−.32	.20	**.54**	−.08	−.05	.96	−.24	.31	**.52**	−.09	.03	.96
O5: Ideas	−.07	−.04	**.58**	.10	**.51**	.84	−.09	−.09	**.61**	.07	**.48**	.88
O6: Values	−.19	.03	**.54**	.09	−.17	.95	−.20	.02	**.56**	.11	−.18	.94
A1: Trust	−.17	.43	−.01	**.59**	.09	.90	−.20	.39	.02	**.62**	.04	.93
A2: Straightforwardness	−.10	−.05	−.04	**.73**	.17	.98	−.11	−.03	−.06	**.73**	.19	.98
A3: Altruism	−.15	.38	.16	**.68**	.29	.94	−.16	.38	.11	**.70**	.29	.95
A4: Compliance	−.23	−.18	−.01	**.76**	.15	.97	−.23	−.11	−.02	**.78**	.09	.99
A5: Modesty	.02	−.28	−.10	**.67**	−.04	.94	.06	−.26	−.11	**.68**	−.03	.95
A6: Tender-Mindedness	.08	.28	.27	**.55**	.31	.88	.13	.26	.20	**.58**	.31	.90
C1: Competence	−.32	.06	.18	.21	**.76**	.95	−.32	.06	.18	.18	**.78**	.96
C2: Order	−.08	−.10	−.03	.14	**.75**	.94	−.06	−.10	−.01	.12	**.72**	.93
C3: Dutifulness	−.13	.00	.09	.38	**.79**	.99	−.10	.03	.05	.37	**.80**	.98
C4: Achievement Striving	−.14	.13	.19	.03	**.83**	.97	−.14	.17	.17	.00	**.83**	.98
C5: Self-Discipline	−.25	−.04	.06	.23	**.83**	.93	−.25	.00	.11	.19	**.83**	.94
C6: Deliberation	−.22	−.29	.03	.32	**.72**	.99	−.20	−.28	.03	.32	**.71**	.99
Factor Congruence	.98	.96	.95	.95	.95	.96	.98	.96	.96	.96	.96	.96

Note: N = 5,109. These are principal components rotated to the American normative target NEO-PI-R. Loadings greater than .40 in absolute magnitude are given in boldface. NEO-PI-3 = NEO Personality Inventory–3; N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness; VCC = variable congruence coefficient.

than cultural differences in the coherence of the factors themselves.

The last six columns of Table 5 report factor and total congruence coefficients between the NEO-PI-3 in each culture and the American adult self-report NEO-PI-R structure (Costa & McCrae, 1992). All factor congruence coefficients for the Neuroticism, Extraversion, Agreeableness, and Conscientiousness factors were above the .85 threshold that is considered indicative of factor replicability (Lorenzo-Seva & ten Berge, 2006). Congruence coefficients for Openness to Experience were below .85 for Uganda (.61), Malaysia (.44), South Korea (.78), Peru (.50), Puerto Rico (.82), China (.80), and Slovak Republic (.78). Across samples, the largest of the correlations of data quality with factor congruences was for Openness, $r = .55$, $p < .01$, suggesting perhaps that this factor is particularly susceptible to test-taking artifacts. The correlation of total congruence with data quality was $r = .63$, $p < .001$.

We also examined variable congruence coefficients within each culture. The most problematic facets were Activity; Openness to Actions, Ideas, Values, and Tender-Mindedness, which all had variable congruences less than .86 (i.e., less than 95% of rotations from random data; see McCrae et al., 1996) in more than a quarter of the samples. Inspection of the factor loadings in these samples suggested that, in comparison with

Table 5
Quality, Reliability, and NEO-PI-3 Factor Replicability of the Samples

		Internal Consistency					Congruence Coefficients[b]					
Culture	Quality[a]	N	E	O	A	C	N	E	O	A	C	Total
Estonia	—	.93	.96	.91	.94	.97	.95	.93	.94	.90	.94	.93
Uganda	2.88	.80	.78	**.65**	.83	.92	.92	.86	**.61**	.92	.92	.86
Malaysia	5.00	.84	.83	**.45**	.80	.93	.93	.89	**.44**	.89	.90	**.84**
Serbia	5.25	.86	.81	.80	.89	.95	.96	.96	.90	.95	.91	.93
South Korea	7.00	.88	.92	.88	.91	.95	.89	.86	**.78**	.95	.90	.88
Peru	7.38	.81	.86	.72	.85	.91	.95	.90	**.50**	.93	.92	.86
Japan	7.50	.88	.91	.76	.90	.94	.95	.94	.87	.93	.88	.91
Chile	7.75	.90	.87	.85	.91	.95	.96	.93	.93	.96	.92	.94
Puerto Rico	7.75	.85	.81	**.65**	.83	.94	.97	.92	**.82**	.86	.87	.89
United States of America	11.25	.92	.92	.85	.93	.95	.97	.96	.94	.94	.96	.95
Argentina	12.25	.90	.90	.84	.92	.95	.96	.93	.91	.93	.91	.93
Portugal	12.25	.86	.89	.85	.90	.93	.96	.94	.93	.94	.91	.93
Islamic Republic of Iran	13.63	.93	.91	.80	.91	.96	.94	.93	.89	.95	.94	.93
Thailand	14.13	.90	.87	.75	.87	.94	.94	.95	.87	.90	.92	.92
France	14.38	.91	.91	.88	.92	.96	.97	.95	.94	.95	.97	.96
Russia	14.50	.86	.91	.80	.89	.95	.92	.93	.88	.94	.93	.92
Hong Kong	14.88	.92	.90	.79	.91	.96	.96	.94	.89	.97	.95	.94
Australia	15.38	.95	.93	.87	.95	.97	.97	.93	.96	.93	.95	.95
Czech Republic	16.00	.92	.93	.89	.91	.96	.95	.91	.94	.94	.92	.93
Poland	16.38	.91	.92	.85	.91	.95	.96	.94	.95	.95	.94	.95
Turkey	16.75	.90	.91	.88	.90	.95	.96	.94	.88	.95	.93	.93
China	17.38	.90	.91	.85	.89	.94	.94	.92	**.80**	.93	.91	.91
Croatia	17.75	.92	.92	.90	.92	.97	.97	.94	.97	.93	.94	.94
Slovak Republic	18.63	.91	.88	.83	.91	.96	.96	.94	**.78**	.91	.93	.91

Note: Alphas less than .70 and congruence coefficients less than .85 are given in boldface. NEO-PI-3 = NEO Personality Inventory–3; N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness.

a. Higher scores indicate higher data quality.

b. Factor and total congruence coefficients comparing Procrustes-rotated principal components to the normative American self-report NEO-PI-R structure (Costa & McCrae, 1992).

the American adult structure, secondary loadings on the Conscientiousness factor were generally low for Activity but high for Openness to Ideas.

Discussion

The present study examined the psychometric characteristics of NEO-PI-R and NEO-PI-3 observer ratings of adolescents aged 12 to 17 years in 18 different languages from 24 cultures. Ratings were gathered using uniform guidelines for international data collection and standardized instructions for participants.

There are two major findings with implications for personality assessment in early adolescence. The first is that the NEO-PI-3 works as well or slightly better than the NEO-PI-R, even in translation; the second is that the structure of personality in observer ratings of adolescents from around the world closely resembles that found in cross-cultural self-ratings and observer ratings of adults (McCrae, Terracciano, et al., 2005a) and self-ratings and peer ratings of American adolescents (Costa et al., 2008; McCrae, Costa, et al., 2005). The structure is nearly identical in younger and older adolescents and in a variety of cultures, including African, Asian, European, and North American and South American cultures.

The present data show that not all NEO-PI-3 replacement items do better than their NEO-PI-R counterparts. The replacement for Item 17 was not significantly correlated with E4: Activity and it had also the lowest item–total correlation in the study of McCrae, Costa, et al. (2005) in both self-reports and observer ratings. Furthermore, the present findings certainly do not constitute a complete construct validation of NEO-PI-3 scales for adolescents worldwide. Factor structure provides a useful

beginning, but further research is needed, including studies of retest reliability, cross-observer agreement, convergence with alternative personality measures, and validity in the prediction of meaningful criteria, such as psychopathology, drug use, academic performance, and career choice. In the present study, raters were undergraduates who may be presumed to have better vocabularies than raters from the general population, so the present results may give an indication of the upper limit of the psychometric values that can be achieved with the NEO-PI-3—although they may also minimize the difference between the two NEO versions. One could argue that the more readable NEO-PI-3 should have particular utility when raters have more limited literacy. However, the present results also suggest that the existing translations of the NEO-PI-R are serviceable for the assessment of personality in early adolescents by college-age or adult raters. Future research should determine whether translations of either or both versions provide valid personality data when extended to self-reports made by boys and girls themselves.

Despite similarities in factor structure, there are also some minor deviations, especially with respect to the replicability of the Openness to Experience factor in Malaysia, Peru, Uganda, South Korea, Slovak Republic, China, and Puerto Rico. Openness factors were clearly identified in most of these cultures when adults were rated (congruence coefficients = .82-.95; median = .89; McCrae, Terracciano, et al., 2005a), so the low values seen in the present research may represent true cultural/developmental differences, suggesting that the Openness to Experience factor may be difficult to identify in these cultures in adolescence. Alternatively, they may be attributable to differences in data quality among samples. Indeed, lower data quality may serve as an explanation for the deviant patterns observed in Uganda, Malaysia, South Korea, and Peru, but the data from China and the Slovak Republic received some of the highest data quality scores. Particular attention should be paid in future studies to the assessment of Openness to Values, which shows very low internal consistency across all samples.

Another variation from the familiar adult structure concerns secondary loadings on the Conscientiousness factor. In American adults, E4: Activity has a substantial secondary loading (.42) on this factor, whereas O5: Ideas (.16) does not. As Table 4 shows, this pattern is reversed in ratings of adolescents, especially those aged 12-14 years. Conscientiousness appears to be less energetic and more intellectual in these targets. This may reflect the perception that good students are conscientious, intellectually curious, and behaviorally docile. The trend is particularly marked in Asian cultures (cf. McCrae et al., 1994), which put a special emphasis on academic achievement: The largest loadings of Openness to Ideas on Conscientiousness are found in Japan, Thailand, Puerto Rico, South Korea, and Malaysia. Whether the same clustering of traits would be seen in self-reports of early adolescents remains to be seen, although American data suggest otherwise (McCrae, Costa, et al., 2005).

In sum, the present work provides promising cross-cultural evidence for the NEO-PI-3 as an assessment tool for personality description in adolescence. At the individual level, the similarity of factor structures in adolescent targets suggests that scores on the NEO-PI-3 are likely to provide useful descriptions of personality traits for this age group in a wide variety of cultures. At the culture level, evidence of the comparability of means from NEO-PI-R and NEO-PI-3 scales (see Table 3) suggests that aggregate scores from either version are directly comparable with those found in the extensive database that has been accumulated with the NEO-PI-R from young to late adulthood extending into old age (Terracciano, Costa, & McCrae, 2006). Such evidence opens new perspectives on the study of mean-level personality changes across international age cohorts, complementing current cross-cultural work in adulthood (McCrae et al., 1999; McCrae et al., 2004; McCrae, Terracciano, et al., 2005a) and meta-analytic work (Roberts, Walton, & Viechtbauer, 2006).

Acknowledgments

The authors are indebted to the following persons for their help with the data collection: Thomas R Cain (Rutgers University, New Brunswick, NJ), Ana Butković (Department of Psychology, Faculty of Philosophy, Zagreb, Croatia), Sylvie Kouřilová (Institute of Psychology, Academy of Science of the Czech Republic, Prague, Czech Republic), Valery E. Oryiol, Ivan G. Senin, Vera V. Onufrieva (Yaroslavl State University, Yaroslavl, Russia), A. Maglio, I. Injoque Ricle, G. Blum, A. Calero, L. Cuenya, V. Pedrón, M. J. Torres Costa, and D. Vion (Instituto de

Investigaciones, Facultad de Psicologia, University of Buenos Aires, Buenos Aires, Argentina), Hamira Alavi (Theology Department, Al-Zahra University, Tehran, Islamic Republic of Iran), Kristina Burgetova (Graduate School of Professional Psychology, University of St. Thomas, St. Paul, MN), Shuo Chen, Irene Lee, and Cindy Lo (The Hong Kong University of Science and Technology, Hong Kong), Javier Paredes (Universidad Peruana Cayetano Heredia, Lima, Peru).

Note

1. Consistent with earlier articles on the Personality Profile of Cultures project, the term *cultures* is loosely used throughout this article to refer to nations or to subgroups within nations. Of course, nations often do not have a single culture, and our samples do not necessarily reflect all cultural variation in each nation.

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.

Arnett, J. J. (1999). Adolescent storm and stress, reconsidered. *The American Psychologist, 54*, 317-326.

Baker, S. R., & Victor, J. B. (2003, August). *Adolescent self-reports of personality and temperament: NEO-PI-R and TPQ.* Paper presented at the XIth European Conference on Developmental Psychology, Milan, Italy.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* San Diego, CA: Academic Press.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory and Five-Factor Inventory professional manual.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., McCrae, R. R., & Martin, T. A. (2008). Incipient adult personality: The NEO-PI-3 in middle-school-aged children. *British Journal of Developmental Psychology, 26*, 71-89.

De Fruyt, F., Mervielde, I., Hoekstra, H. A., & Rolland, J. P. (2000). Assessing adolescents' personality with the NEO-PI-R. *Assessment, 7*, 329-345.

Digman, J. M. (1963). Principal dimensions of child personality as inferred from teachers' judgements. *Child Development, 34*, 43-60.

Goldberg, L. R. (2001). Analyses of Digman's child-personality data: Derivation of Big Five factor scores from each of six samples. *Journal of Personality, 69*, 709-743.

John, O. P., Caspi, A., Robins, R. W., Moffitt, T. E., & Stouthamer-Loeber, M. (1994). The Little 5: Exploring the nomological network of the Five-Factor Model of personality in adolescent boys. *Child Development, 65*, 160-178.

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*, 57-64.

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *The American Psychologist, 52*, 509-516.

McCrae, R. R., Costa, P. T., Jr., de Lima, M. P., Simões, A., Ostendorf, F., Angleitner, A., et al. (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., Jr., del Pilar, G. H., Rolland, J. P., & Parker, W. D. (1998). Cross-cultural assessment of the Five-Factor Model: The Revised NEO Personality Inventory. *Journal of Cross-Cultural Psychology, 29*, 171-188.

McCrae, R. R., Costa, P. T., Jr., Hřebíčková, M., Urbánek, T., Martin, T. A., Oryol, V. E., et al. (2004). Age differences in personality traits across cultures: Self-report and observer perspectives. *European Journal of Personality, 18*, 143-157.

McCrae, R. R., Costa, P. T., Jr., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO Personality Inventory. *Journal of Personality Assessment, 84*, 261-270.

McCrae, R. R., Costa, P. T., Jr., Terracciano, A., Parker, W. D., Mills, C. J., De Fruyt, F., et al. (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., Terracciano, A., & Members of the Personality Profiles of Cultures Project. (2005a). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology, 88*, 547-561.

McCrae, R. R., Terracciano, A., & Members of the Personality Profiles of Cultures Project. (2005b). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology, 89*, 407-425.

McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., 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.

Mervielde, I., Buyst, V., & De Fruyt, F. (1995). The validity of the Big Five as a model for teachers' ratings of individual-differences among children aged 4-12 years. *Personality and Individual Differences, 18*, 525-534.

Mervielde, I., & De Fruyt, F. (2002). Assessing children's traits with the Hierarchical Personality Inventory for Children. In B. De Raad & M. Perugini (Eds.), *Big Five assessment* (pp. 129-146). Göttingen, Germany: Hogrefe & Huber.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132*, 1-25.

Terracciano, A., Costa, P. T., Jr., & McCrae, R. R. (2006). Personality plasticity after age 30. *Personality and Social Psychology Bulletin, 32*, 999-1009.