Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures

Robert R. McCrae and Antonio Terracciano
National Institute on Aging, National Institutes of Health, Department of Health and Human Services

78 Members of the Personality Profiles of Cultures Project

To test hypotheses about the universality of personality traits, college students in 50 cultures identified an adult or college-aged man or woman whom they knew well and rated the 11,985 targets using the 3rd-person version of the Revised NEO Personality Inventory. Factor analyses within cultures showed that the normative American self-report structure was clearly replicated in most cultures and was recognizable in all. Sex differences replicated earlier self-report results, with the most pronounced differences in Western cultures. Cross-sectional age differences for 3 factors followed the pattern identified in self-reports, with moderate rates of change during college age and slower changes after age 40. With a few exceptions, these data support the hypothesis that features of personality traits are common to all human groups.

Strong claims have recently been made about the universality of personality traits. McCrae and Costa (1997) argued that the five-factor model (FFM) of personality is found in all cultures,[1] a hypothesis subsequently supported in a wider range of cultures (Rolland, 2002). McCrae et al. (1999) reported that cross-sectional age differences were similar in different cultures whose cohorts had experienced very different life histories, and Costa, Terracciano, and McCrae (2001) reported pancultural patterns of gender differences. McCrae, Costa, Martin, et al. (2004) provided data on cross-observer agreement suggesting that even in collectivistic cultures, where there is purportedly a greater emphasis on relationships than on traits, people accurately perceive their own and others' traits. These recurring regularities—despite differences in language, history, religion, and culture—suggest that personality

Robert R. McCrae and Antonio Terracciano, Gerontology Research Center, National Institute on Aging; the 78 contributing members of the Personality Profiles of Cultures Project.

The 78 members of the Personality Profiles of Cultures Project are listed in order of data submission in the Appendix.

Robert R. McCrae receives royalties from the Revised NEO Personality Inventory (NEO-PI-R). Portions of these data were presented at the Second World Congress on Women's Mental Health, Washington, DC, March 2004. Czech participation was supported 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. Sami Gülgöz's participation was supported by the Turkish Academy of Sciences. Burkinabè and French Swiss participation was supported by a grant from the Swiss National Science Foundation to J. Rossier. The data collection in Hong Kong was supported by RGC Direct Allocation Grants DAG02/03.HSS14 and DAG03/04.HSS14 awarded to M. Yik. Data collection in Malaysia was supported by UKM Fundamental Research Grant 11JD/015/2003.

We thank Dave Schmitt for contact information on potential collaborators, Bob Smith for authorizing use of the NEO-PI-R, and the translators whose earlier work on Form S of the NEO-PI-R made this project feasible. For assistance on this project we thank Herbert Biggs, Luciana de Almeida, Hudson W. Carvalho, Marco Montarroyos Calegaro, Andréia da Silva Bez, Zheng Li, Ana Butković, Ole Dreyer, Susy Ball, Anna Gramberg, Honathan Harrow, V. S. Bose, Suguna Kannan, K. Sarita, K. Madhavi, Lidwina Dominica R, Vina Bunyamin, Hiromi Imuta, Kenji Sugiyama, Midori Takayama, Rozita Kamis, Rosmaini Ismail, Anna Nedtwig, Zachary Smith, Aaron Wolen, Maya Tamir, Christie Napa Scollon, Valery E. Oryol, Ivan G. Senin, J. C. Munene, Silvo Kozelj, Manca Jakic, Simona Zbačnik, Nadia Messoulam, Facundo Abal, Fernanda Molina, Daiana Bion, Sebastián Mosquera, Ludmila Firpo, Lorena Etcheverry, Fernando Vera, Catherine Currell, Richard Chan, Christopher Paik, Herbert H. Freudenthaler, Andreas Fink, and Cornelia Hohenbichler.

German, Russian, and Czech data were taken from earlier studies (McCrae, Costa, Hřebíčková, et al., 2004; Ostendorf & Angleitner, 2004), and portions of the Brazilian, Lebanese, and Thai data are also reported in chapters in McCrae (in press); McCrae, Terracciano, and Khoury (in press); and Costa and McCrae (in press), respectively.

Correspondence concerning this article should be addressed to Robert R. McCrae, Gerontology Research Center, National Institute on Aging, National Institutes of Health, Department of Health and Human Services, Box #03, 5600 Nathan Shock Drive, Baltimore, MD 21224-6825. E-mail: mccraej@grc.nia.nih.gov

[1] Throughout this article, we use the term *cultures* loosely to refer to nations or subgroups within nations. We are aware that nations do not have monolithic cultures and that our samples do not necessarily reflect the full cultural diversity seen within nations.

Journal of Personality and Social Psychology, 2005, Vol. 88, No. 3, 547–561
In the public domain DOI: 10.1037/0022-3514.88.3.547

traits are basic features of the human species (Allik & McCrae, 2002). The present study offers new tests of these hypotheses of universality.

One obvious limitation to prior claims is that cultures and subcultures have not been exhaustively studied. Only a few African—and no Arabic—cultures have been included in previous studies using the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992a) or other measures of the FFM (e.g., Heaven, Connors, & Stones, 1994). No preliterate cultures have been examined; in fact, most studies have used college student samples, whose members may be relatively Westernized.

A second limitation is that most studies have relied exclusively on self-report methods, leaving the possibility that method artifacts may be responsible for some or all of the findings. Observer ratings form an alternative method of personality measurement, known to be convergent but not wholly redundant with self-reports (McCrae, Costa, Martin, et al., 2004). In American studies, observer ratings typically yield similar conclusions about structure and about age and gender differences (e.g., Costa & McCrae, 1992b), but this is not invariably the case in cross-cultural research. For example, Extraversion and Openness to Experience both appeared to decline cross-sectionally in German adults when self-reports were analyzed but not when observer ratings were analyzed (McCrae et al., 2000). In a Czech sample, age associations found in self-reports were replicated in observer ratings for Extraversion and Openness but not for Neuroticism or Agreeableness (McCrae, Costa, Hřebíčková, et al., 2004).

The NEO-PI-R offers two versions: a self-report Form S and an observer rating Form R, with the same items rephrased in the third person. The factor structure of Form R in American samples closely resembles that of Form S (e.g., Piedmont, 1994), and the same is true in German-, Russian-, and Czech-language versions (McCrae, Costa, Martin, et al., 2004; Ostendorf & Angleitner, 2004). However, there appear to be no published studies of the factor structure of observer rating measures of the FFM in non-Western cultures. The present article includes data from more than a dozen.

Past and Present Designs

Most previous cross-cultural studies of the FFM were based on secondary analyses of data collected for a variety of purposes (Costa et al., 2001; McCrae, 2002; Rolland, 2002). Samples varied in size and composition (although only normal volunteer data were used) and in the time period of data collection. In many cases, only summary statistics were available, and demographic data were generally not available. No attempt was made to assess or control the quality of the data. It is testimony to the robustness of the underlying effects that clear regularities emerged despite these limitations.

In the present study, we collected data from college students who were asked to identify an individual from one of four target groups—college-aged men, college-aged women, adult men, and adult women—and provide ratings of that target on Form R of the NEO-PI-R. Because a uniform approach was taken to data collection, results are more likely to be comparable across cultures (cf. Schwartz, 1992). Samples are similar in size, age and sex of targets, and time period in which data were collected. In addition, item-level data and basic demographics are available for each sample.

The use of college student raters also offers advantages. College students are not, in general, representative of their national populations, and this is particularly likely to be true in less affluent cultures. However, this fact is less problematic in observer rating studies than in self-report studies: Raters could choose anyone they knew well as a target, yielding a wider age and educational range than would normally be obtained in self-report studies. For example, about 11% of the targets in the present study had fewer than 9 years of education.

In addition, college students may be more familiar and comfortable with questionnaire methods than members of the general population, especially in non-Western cultures (cf. Marsella, Dubanoski, Hamada, & Morse, 2000), yielding more meaningful data. However, even college students may differ across cultures in test-taking experience and attitudes; in particular, cultural differences in acquiescence have been noted (Smith, 2004). In the present study, we attempted to assess the quality of the data in each sample, to compare cultures on data quality, and to take quality into account in interpreting results. It must be stressed that quality is primarily a function of the fit between an imported Western personality measure and the experiences and attitudes of each culture; poorer data quality should not be seen as evidence of problems with either the instrument or the respondents but rather of their mismatch.

Finally, the use of observer ratings permits an analysis of certain aspects of person perception and assessment. When self-reports are examined, target and rater are completely confounded, making it impossible to know whether ratings are a function of the person being rated or the person making the ratings. It is possible, for example, that women everywhere score higher on Neuroticism not because they are less emotionally stable but merely because they are better able to perceive negative affect (cf. Feldman Barrett, Lane, Sechrest, & Schwartz, 2000; Terracciano, Merritt, Zonderman, & Evans, 2003) or more willing to attribute it to a target than men are. In the present design, both men and women rate men and women, so it is possible to estimate sex differences in rating styles or biases.

Method

Cultures

We recruited collaborators from a wide range of cultures, subject to the requirement that prospective participants would be fluent in English or one of the other languages for which an authorized NEO-PI-R translation was available. Collaborators were primarily individuals who had previously used the NEO-PI-R in their own research or who had been members of another multinational study (Schmitt et al., 2003). To increase representation of African and Arabic cultures, we searched the Internet and PsycINFO for personality psychologists from those areas. Data gathered so far are from 50 cultures representing six continents, using translations into Indo-European, Hamito-Semitic, Sino-Tibetan, Daic, Uralic, Malayo-Polynesian, Dravidian, and Altaic languages. American and Brazilian data were gathered from several sites. German, Russian, and Czech data were obtained by selecting targets of the intended ages from existing observer rating data (McCrae, Costa, Martin, et al., 2004; Ostendorf & Angleitner, 2004). Cultures and the language in which the NEO-PI-R was administered are given in the first two columns of Table 1. As noted in the table, 22 of these cultures have not previously been studied using the NEO-PI-R.

Table 1
Characteristics of the Samples

Culture	Language	*N*	Raters		Targets		
			% male	Mean age (years)	% male	Mean age (years)	Education[a]
Argentina[b]	Spanish[c]	204	4.9	21.6	51.0	35.3	2.4
Australia[b]	English[c]	206	31.1	21.2	51.0	34.2	2.4
Austria	German[c]	158	15.2	22.9	50.6	37.3	—
Belgium	Flemish[c]	247	18.2	18.7	51.4	33.7	2.7
Botswana[b]	English	186	31.2	21.2	51.1	36.0	2.2
Brazil[b]	Portuguese[c]	597	29.3	24.4	51.3	34.8	2.3
Burkina Faso	French	207	69.3	24.8	49.3	34.1	2.0
Canada	English[c]	133	26.3	20.0	26.3	29.9	2.6
Chile[b]	Spanish[c]	194	23.7	20.1	48.5	34.9	2.9
China	Chinese	177	47.5	21.4	51.4	33.8	2.4
Croatia	Croatian	191	33.5	20.5	50.8	34.6	2.5
Czech Republic	Czech[c]	400	39.0	41.7	42.5	45.5	2.2
Denmark	Danish[c]	153	15.8	24.9	47.1	38.5	2.4
Estonia	Estonian	298	32.9	19.4	53.7	33.7	2.6
Ethiopia[b]	English	197	87.2	21.8	50.3	34.6	1.9
France	French[c]	274	49.3	—	49.3	37.3	2.5
Germany	German[c]	593	34.7	40.1	26.3	41.9	—
Hong Kong	Chinese	207	47.3	20.1	50.7	34.4	2.3
Iceland[b]	Icelandic	199	50.3	26.0	49.7	35.6	2.5
India	Telugu	185	49.2	21.0	50.3	33.1	2.2
Indonesia	Indonesian	196	38.8	19.7	49.5	35.8	2.7
Italy	Italian	195	48.7	23.4	49.2	35.9	2.4
Japan	Japanese[c]	191	49.2	19.5	50.3	34.1	2.6
Kuwait[b]	Arabic	468	45.4	20.1	46.6	35.2	2.5
Lebanon[b]	English	200	38.5	18.9	41.0	37.3	2.7
Malaysia	Malay	289	22.1	22.0	43.9	37.0	2.3
Malta[b]	English	202	26.7	20.9	50.0	35.6	2.6
Mexico[b]	Spanish[c]	173	43.0	23.9	53.8	34.5	2.6
Morocco[b]	English	171	41.4	21.4	57.3	33.0	2.3
New Zealand[b]	English[c]	200	27.5	19.9	50.5	34.5	2.5
Nigeria[b]	English	184	46.7	28.3	48.9	34.2	2.5
Peru	Spanish	154	28.6	21.3	43.5	35.2	2.6
Philippines	Filipino	197	25.9	18.3	50.3	33.3	2.7
Poland[b]	Polish[c]	197	49.7	22.1	49.7	34.9	2.6
Portugal	Portuguese[c]	198	20.7	21.7	51.0	34.8	2.3
Puerto Rico[b]	Spanish[c]	160	33.1	—	53.1	33.9	2.8
Russia	Russian	320	48.4	36.1	48.1	36.7	—
Serbia	Serbian	200	20.5	21.5	50.5	34.7	2.6
Slovakia[b]	Slovakian	198	50.5	20.1	49.5	33.5	2.3
Slovenia[b]	Slovene	209	42.7	22.8	47.4	35.6	2.5
South Korea	Korean[c]	196	50.5	22.0	51.5	34.0	2.6
Spain	Spanish[c]	200	18.0	21.0	50.0	35.7	2.3
Switzerland	German[c]	214	27.4	29.4	45.8	37.7	2.6
Switzerland	French[c]	265	21.7	21.3	48.3	36.5	2.3
Thailand[b]	Thai	209	33.5	19.6	50.7	34.7	2.7
Turkey	Turkish[c]	208	44.2	19.3	50.0	34.1	2.6
Uganda[b]	English	166	44.6	25.8	50.0	35.3	2.3
United Kingdom: England[b]	English[c]	194	40.7	23.1	49.0	37.2	2.7
United Kingdom: Northern Ireland[b]	English[c]	106	12.3	21.3	49.1	33.3	2.6
United States	English[c]	919	30.9	20.5	49.6	34.1	2.6

Note. Dashes in cells indicate data were not available. NEO-PI-R = Revised NEO Personality Inventory.
[a] Mean level, where 0 = no education, 1 = primary, 2 = secondary, 3 = college. [b] Not included in previous cross-cultural comparisons using the NEO-PI-R. [c] Used a published version of the NEO-PI-R in participants' native language.

Instrument

The NEO-PI-R is a 240-item measure of the FFM. It contains 30 eight-item facet scales, six for each of the five basic personality factors, Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). Responses are made on a 5-point Likert scale, from *strongly disagree* to *strongly agree.* The factors can be estimated by domain scores, which sum the relevant 6 facets, or more precisely by factor scores, which are a weighted combination of all 30 facets (Costa & McCrae, 1992a, Table 2). Two parallel forms have been developed: Form S for self-reports and Form R for observer ratings, in which the items have been rephrased in the third person. Evidence on the

reliability and validity of the English version are presented in the NEO-PI-R manual (Costa & McCrae, 1992a). Although the NEO-PI-R does not include a social desirability scale (Piedmont, McCrae, Riemann, & Angleitner, 2000), it does provide some checks for protocol validity, and protocols deemed invalid have substantially lower retest stability (Carter et al., 2001).

Form S of the NEO-PI-R has been translated into over 30 languages. In almost all cases, translations were done by bilingual psychologists native to the culture. An independent back-translation was reviewed by the test authors, and modifications were made as needed. In some cases, the translations have been extensively validated and published (e.g., Hoekstra, Ormel, & De Fruyt, 1996; Shimonaka, Nakazato, Gondo, & Takayama, 1999); in other cases, the translations can be considered research instruments. For the present study, collaborators modified the first-person version to create a third-person version. They also translated the instructions, which were reviewed in back-translation by Robert R. McCrae and Antonio Terracciano. Revisions were made based on these reviews.

Participants, Targets, and Procedures

Participants were college students[2] who volunteered to participate anonymously in a study of personality across cultures. The composition of the samples by sex and the mean age of the raters are given in the third and fourth columns of Table 1. The great majority of raters were native-born citizens of their countries and generally reflected the ethnic make-up of their countries.[3]

Raters were randomly assigned to one of four target conditions[4] and were asked for ratings of college-aged women, college-aged men, adult (over 40) men, or adult women. For the college-aged targets, raters were asked the following:

> Please think of a *woman* [*man*] *aged 18–21* whom you know well. She [he] should be someone who is a native-born citizen of your country. She [he] can be a relative or a friend or neighbor—someone you like, or someone you do not like. She [he] can be a college student, but she [he] need not be.

In the adult conditions, the age specified was *over age 40* to form a clear contrast to the college-aged targets. Raters were then asked to estimate the age and years of formal education (none, 1–8 years, 9–12 years, over 12 years) of the target and to provide demographic information on themselves before completing the NEO-PI-R. Data on the compositions of the target samples by sex, their mean age, and their degree of education are given in the last three columns of Table 1.

Data Quality Assessment

When instruments and methods developed and validated in one culture are exported to another, their psychometric properties may be affected. That might be due to real differences in psychological functioning, but it might also be due to culture-related artifacts. Subtleties of meaning may be lost in translation; response styles may vary across cultures; the task of completing a questionnaire may be unfamiliar and confusing. Ideally, an assessment of the quality of the data should be made before substantive results are considered. Deviations from strict replication can be discounted if there are independent indicators that the instrument itself is less than optimal in some cultural contexts.

The NEO-PI-R manual (Costa & McCrae, 1992b) specifies that protocols with more than 40 missing responses are considered invalid. In addition, repetitive responses (e.g., more than 9 consecutive *disagree* responses or 10 consecutive *neutral* responses), which are rare in volunteer samples, are considered evidence of random responding. Cases considered invalid by either of these criteria were eliminated. However, we also considered that the frequency of valid responses in a sample probably reflected the quality of data in that administration in general, and we used the percentage of valid protocols in the unscreened sample (ranging from 85.1% to 100%) as a first indicator of data quality.

Acquiescence can be estimated by counting the number of *agree* and *strongly agree* responses to all items. Because NEO-PI-R scales are balanced in keying, the net effect of acquiescent responding is limited, and acquiescence does not invalidate a protocol. However, it is a possible indicator of poorer quality data. Using the cutoff scores in the manual, we calculated the frequency of acquiescent ($\geq$150 *agree* or *strongly agree* responses) or nay-saying ($\leq$50 *agree* or *strongly agree* responses) protocols in each unscreened sample (from 0% to 21.5%) as a second (reversed) index of data quality.

Where fewer than 40 items are missing, missing data are treated by substituting a neutral response. Before making that substitution, we counted the number of missing responses and used the sample mean (from 0 to 11.4 items) as a third (reversed) indicator of data quality.

We considered it likely that fewer problems would occur when raters completed the questionnaire in their native language or when the samples as a whole were judged by our collaborators as being fluent in the second language that was used. Our fourth indicator of data quality was scored 2 for native language, 1 for very fluent in the second language, and 0 for somewhat fluent in the second language. Although many of the unpublished NEO-PI-R translations are excellent, it is probably fair to assume that published versions are further along in psychometric development than most unpublished versions. All samples that were tested in a second language used a published version (English or French); for samples in which the native language was used, our fifth indicator was scored 1 for published and 0 for unpublished translations (see Table 1 note).

Finally, we asked collaborators directly if there were any problems. The most common problem mentioned was the length of the questionnaires and the time required to complete them. Presence or absence of a problem was our sixth indicator of data quality.

Results and Discussion

Data Quality and Internal Consistency

The six indicators of data quality were modestly intercorrelated (rank-order *r*s = .09–.66; coefficient α = .76; all indices were significantly related to at least two other indices), so we expressed each as a rank score and used the mean of the six indicators as an overall measure of data quality. This value is reported in the second column of Table 2, and the entries are listed in descending order. By and large, the entries at the top of the list are from affluent, mostly Western nations, whereas those at the bottom are from underdeveloped nations. In part, this is probably due to the availability of translations in most European languages but relatively few African languages. In part, it probably also reflects the fact that the NEO-PI-R was developed within the Western tradition of psychological measurement, and completing it is perhaps a more meaningful task for Westerners.

In item analyses, we examined the corrected item/domain correlations in the full sample and in each culture. In the full sample,

[2] In Germany, Russia, and the Czech Republic, where existing data were used, raters were usually spouses or same-age peers of the targets. Nonstudent raters were also rarely included in the new data collection.

[3] Exceptions were Russia and Malaysia, where samples were almost entirely composed of ethnic Russians and Malays, respectively.

[4] Because of a misunderstanding, participants in Uganda and France were asked to complete all four versions, which many raters found burdensome.

Table 2
Quality, Reliability, and Factor Replicability of the Samples

Culture	Quality	Internal consistency					Congruence coefficients[a]					Total
		N	E	O	A	C	N	E	O	A	C	
Germany	37.9	.91	.89	.88	.91	.93	.98	.96	.96	.97	.97	.97
Spain	37.8	.92	.93	.93	.93	.95	.97	.93	.94	.95	.93	.94
French Switzerland	37.0	.94	.92	.91	.92	.92	.95	.94	.95	.96	.96	.95
Denmark	35.9	.93	.91	.92	.94	.95	.95	.92	.96	.94	.96	.94
France	35.6	.93	.90	.91	.93	.93	.97	.96	.96	.95	.95	.96
German Switzerland	34.7	.93	.89	.91	.93	.94	.97	.97	.96	.95	.96	.96
Chile	33.4	.92	.91	.93	.92	.94	.96	.95	.97	.95	.94	.95
New Zealand	33.3	.92	.91	.90	.94	.95	.97	.93	.95	.95	.93	.94
Belgium	33.3	.92	.91	.90	.94	.94	.97	.93	.95	.95	.96	.95
Portugal	32.9	.90	.89	.89	.93	.94	.96	.97	.93	.95	.93	.94
Turkey	32.3	.90	.93	.90	.92	.95	.95	.96	.94	.96	.95	.95
Poland	31.7	.89	.90	.91	.92	.94	.97	.94	.93	.95	.93	.94
Serbia	31.6	.90	.90	.90	.94	.95	.97	.94	.97	.94	.94	.95
Malta	31.6	.91	.90	.90	.94	.94	.98	.95	.93	.97	.94	.95
Czech Republic	31.0	.90	.90	.92	.93	.94	.95	.96	.96	.95	.96	.95
Estonia	30.7	.92	.95	.91	.94	.96	.96	.97	.94	.92	.97	.95
United Kingdom: Northern Ireland	30.5	.92	.90	.90	.95	.96	.95	.94	.90	.97	.94	.94
Slovakia	30.4	.90	.84	.87	.93	.94	.96	.94	.95	.94	.95	.94
Iceland	29.8	.90	.91	.88	.92	.95	.97	.94	.96	.96	.97	.95
Austria	29.1	.93	.90	.92	.93	.94	.95	.92	.93	.96	.95	.94
United Kingdom: England	28.8	.92	.91	.90	.95	.94	.97	.92	.97	.95	.94	.95
Canada	27.9	.85	.90	.87	.92	.94	.97	.93	.93	.95	.95	.94
Australia	27.3	.92	.90	.88	.94	.95	.97	.95	.96	.95	.96	.95
Japan	26.9	.90	.91	.87	.93	.92	.96	.96	.91	.93	.95	.94
South Korea	26.7	.86	.90	.89	.92	.95	.97	.94	.89	.93	.95	.93
Hong Kong	26.3	.92	.87	.88	.93	.93	.96	.93	.92	.96	.95	.94
Brazil	26.3	.90	.90	.88	.91	.94	.97	.96	.94	.95	.96	.96
Italy	25.8	.89	.87	.91	.90	.94	.95	.94	.96	.96	.95	.95
United States	25.7	.91	.91	.88	.93	.94	.97	.96	.96	.96	.97	.96
Thailand	25.0	.88	.87	.75	.89	.93	.94	.92	**.83**	.95	.93	.92
Indonesia	22.8	.88	.83	.71	.85	.91	.94	.94	**.84**	.94	.96	.93
Argentina	22.8	.84	.89	.85	.91	.92	.96	.96	.93	.93	.94	.94
Burkina Faso	21.6	.84	.85	.73	.91	.94	.96	.92	.85	.94	.91	.92
Kuwait	19.3	.87	.84	.75	.88	.92	.97	.95	.86	.95	.95	.94
Mexico	18.9	.87	.87	.80	.85	.92	.96	.95	.89	.95	.95	.94
Philippines	18.3	.81	.84	.77	.89	.93	.97	.92	.89	.94	.93	.93
Croatia	17.7	.90	.90	.88	.92	.95	.96	.96	.95	.95	.96	.95
Russia	16.6	.88	.90	.85	.89	.93	.94	.94	.94	.95	.95	.94
China	16.3	.87	.85	.83	.87	.90	.93	.93	.90	.95	.94	.93
India	16.1	.77	.80	**.59**	.83	.88	.93	.87	**.80**	.91	.92	.89
Peru	15.8	.86	.87	.75	.85	.91	.96	.92	.88	.97	.92	.93
Slovenia	13.8	.90	.89	.90	.91	.93	.98	.97	.96	.95	.96	.96
Malaysia	13.5	.80	.78	**.59**	.85	.91	.92	**.80**	**.82**	.94	.93	.90
Botswana	13.5	.75	.82	**.61**	.89	.92	.88	**.82**	**.53**	.90	.89	.82
Nigeria	13.2	**.61**	.73	**.25**	**.63**	.78	**.76**	**.66**	**.56**	.88	**.65**	.71
Puerto Rico	12.9	.89	.86	.81	.86	.90	.95	.94	.93	.94	.96	.95
Ethiopia	10.9	.71	.70	**.60**	.76	.87	.89	.85	**.82**	.93	.96	.90
Lebanon	10.0	.84	.85	.85	.91	.94	.96	.95	.88	.95	.95	.93
Uganda	6.0	.73	.77	**.68**	.81	.89	.93	.88	**.84**	.91	.95	.90
Morocco	5.5	**.54**	**.57**	**.58**	**.66**	.82	.91	.85	**.66**	.89	.90	.85

Note. Alphas less than .70 and congruence coefficients less than .85 are given in boldface type. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness.
[a] These are factor and total congruence coefficients comparing five Procrustes-rotated principal components in each sample with the American normative self-report structure (Costa & McCrae, 1992a).

these correlations were positive for 239 items. Item 17, "I have a leisurely style in work and play" (reversed), was the exception; in some cultures it was a good indicator of Extraversion, but in most it appeared to assess Introversion. Within cultures, 394 of the 12,000 (i.e., 240 × 50) corrected item/domain correlations (3.3%) were negative. These tended to occur for the same items across cultures (such as Item 17) and for the same subset of cultures, leading to lower internal consistency. Although poorly performing items might be treated as missing, we retained them in the present study. The third through seventh columns of Table 2 report coefficient alpha for the 48-item domain scales. In general, these are quite high, with median values of .90, .90, .88, .92, and .94 for N,

E, O, A, and C, respectively. Nevertheless, there are some instances of low alphas (12 of 250, or 4.8%, are lower than .70), especially for O—a domain that has also shown problematic reliability in self-report data in Malaysia and Zimbabwe (Mastor, Jin, & Cooper, 2000; Piedmont, Bain, McCrae, & Costa, 2002). The value of .25 in Nigeria is particularly notable, suggesting the possibility that O is not a meaningful dimension in that culture. However, an alternative interpretation is that low alphas reflect only poor data quality. That hypothesis is supported by rank-order correlations of column 2, Quality, with columns 3–7 (rs = .63–.82, $p < .001$). Careless or acquiescent responding, fatigue, or failure to understand the nuances of language can have serious consequences for item-level analyses. When aggregated into facet scales, however, some of this error may be reduced.

Factor Structure

The first substantive question addressed here is the universality of the FFM in observer ratings. An analysis combining raw data from each sample would confound the covariation of traits across individuals with covariation across cultures (Bond, 2001). We therefore standardized data within each culture (so that the means of all facets in each culture were transformed to 0, the standard deviations to 1.0), and factored the 30 facet scales.[5] The first six eigenvalues were 6.67, 4.40, 3.51, 2.43, 1.46, and 0.84, unmistakably suggesting a five-factor solution. After varimax rotation, the expected structure of the FFM was clearly replicated, with factor congruence coefficients ranging from .96 to .98. The principal difference between this matrix and the normative self-report matrix is that the Form R factors account for more total variance than Form S factors (61.6% vs. 56.9%), and the A and C factors account for a larger percentage of the common variance (23.9% and 26.4%) in observer ratings than in self-reports (19.8% and 22.2%). That phenomenon had already been noted in American Form R data (Costa & McCrae, 1992a).

Although the varimax-rotated structure is almost identical to the American self-report normative structure, comparisons of factor structures are most direct when orthogonal Procrustes rotation is used to align factors maximally with the target (McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). Table 3 reports the factor structure for the total sample and gives variable, factor, and total congruence coefficients. In this study, E3: Assertiveness has a somewhat stronger (negative) loading on A than on E, but all other facets have their primary loading on the intended factor, and secondary loadings (such as N2: Angry Hostility on A and O3: Feelings on E) are also replicated, as attested to by the large variable congruence coefficients. The FFM structure was also replicated within each of the four age-and-sex target groups, with factor congruence coefficients after Procrustes rotation ranging from .96 to .98.[6]

Although it is clear from Table 3 that the FFM does in fact represent the structure of observer-rated personality traits across cultures, it is possible that there is a minority of cultures in which the structure is not found. Factor analyses with Procrustes rotation were therefore conducted in each sample separately; results are summarized as factor and total congruence coefficients in the last six columns of Table 2. By the .85 criterion of factor replicability (Haven & ten Berge, 1977), 94.4% of the factors are replications of the American normative Form S structure. Statistically, there is evidence that the FFM is replicable in all the cultures considered here: With one exception (O in Botswana), all factor congruence coefficients are greater than 95% of chance rotations (McCrae et

Table 3
Factor Loadings for Observer-Rated NEO-PI-R Facet Scales in the Combined Sample

	Factor					
NEO-PI-R facet	N	E	O	A	C	VCC
N1: Anxiety	**.81**	−.07	−.05	.06	.12	.95[a]
N2: Angry Hostility	**.59**	−.05	−.12	**−.57**	.00	.98[a]
N3: Depression	**.78**	−.18	.02	.08	−.17	.98[a]
N4: Self-Consciousness	**.68**	−.23	−.06	.18	−.04	.97[a]
N5: Impulsiveness	**.41**	.33	.08	−.34	−.36	.97[a]
N6: Vulnerability	**.71**	−.08	−.11	−.06	**−.43**	.99[a]
E1: Warmth	−.14	**.72**	.15	**.42**	.11	.99[a]
E2: Gregariousness	−.12	**.76**	.04	.05	−.15	.98[a]
E3: Assertiveness	−.31	**.40**	.10	**−.44**	.38	.97[a]
E4: Activity	−.03	**.58**	.08	−.26	.34	.98[a]
E5: Excitement Seeking	−.07	**.54**	.31	−.26	−.29	.89[b]
E6: Positive Emotions	−.14	**.70**	.32	.17	.02	.97[a]
O1: Fantasy	.15	.27	**.60**	.00	−.32	.97[a]
O2: Aesthetics	.16	.14	**.75**	.14	.10	.99[a]
O3: Feelings	.28	**.46**	**.51**	.10	.15	.98[a]
O4: Actions	−.14	.24	**.55**	−.02	−.13	.98[a]
O5: Ideas	−.14	−.05	**.72**	.00	.31	.97[a]
O6: Values	−.20	.11	**.49**	.19	−.07	.88[b]
A1: Trust	−.22	.38	.09	**.62**	.01	.96[a]
A2: Straightforwardness	−.10	−.08	−.05	**.72**	.21	.99[a]
A3: Altruism	−.09	.38	.08	**.70**	.27	.96[a]
A4: Compliance	−.20	−.05	.02	**.79**	.00	.99[a]
A5: Modesty	.04	−.11	−.09	**.73**	.04	.94[a]
A6: Tender-Mindedness	.09	.27	.21	**.63**	.15	.97[a]
C1: Competence	−.31	.07	.16	.07	**.77**	.97[a]
C2: Order	.03	−.07	−.09	.03	**.72**	.97[a]
C3: Dutifulness	−.11	−.01	−.04	.33	**.79**	.99[a]
C4: Achievement Striving	−.12	.14	.11	−.09	**.80**	.99[a]
C5: Self-Discipline	−.23	.03	−.03	.14	**.83**	.97[a]
C6: Deliberation	−.24	−.26	.01	.31	**.67**	.99[a]
Congruence[c]	.98[a]	.97[a]	.97[a]	.97[a]	.97[a]	.97[a]

Note. N = 11,985. These are principal components rotated to the American normative target (Costa & McCrae, 1992a). Loadings greater than .40 in absolute magnitude are given in boldface type. NEO-PI-R = Revised NEO Personality Inventory; N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; VCC = variable congruence coefficient.
[a] Congruence higher than that of 99% of rotations from random data. [b] Congruence higher than that of 95% of rotations from random data. [c] Factor/total congruence coefficient with target matrix.

[5] In fact, an analysis of raw scores produced almost identical results.

[6] The normative Form S structure is used as a target because it has been the target in all previous cross-cultural comparisons and because no large-scale American Form R structure has been published. The largest available (N = 908) Form R factor structure comes from Czech data (McCrae, Costa, Martin, et al., 2004), some of which are used in this study. Congruence coefficients between the present total sample and the Czech varimax structure are .98, .98, .99, .99, and .98 for N, E, O, A, and C, respectively.

al., 1996), and all 50 total congruence coefficients are greater than 99% of chance rotations.

However, it is also clear that factor solutions in several cultures are far from perfect replications of the American normative structure. Particularly striking are the low congruences in Botswana and Nigeria. The three other Black African cultures—Burkina Faso, Ethiopia, and Uganda—had clearer replications, but not so clear as those found in most European cultures. Although these countries differ dramatically in language, religion, and customs, Okeke, Draguns, Skeku, and Allen (1999) have argued that they share certain features, such as close bonds within the family and a traumatic history of European colonialism, that might lead to a common personality structure. We therefore considered the possibility that there is some distinctive African personality structure that differs appreciably from the FFM found elsewhere in the world. Alternatively, it may be that these imperfect replications are due to problems in the data that stem from the use of a Western questionnaire that may not be wholly appropriate in this cultural context. The latter interpretation appears more plausible for three reasons. First, the magnitude of the total congruence coefficient is strongly associated with our index of data quality (rank-order $r = .60$, $p < .001$). Second, we found no evidence that the five African cultures resembled each other more than they did the normative structure: Total congruence coefficients between the 10 pairs of African cultures ranged from .71 to .91, with a median of .85 (for similar findings with self-reports, cf. Rossier, Dahourou, & McCrae, in press). Finally, if weak results are due to random error, then increasing the sample size should improve the fit. We therefore factored the combined data ($N = 940$) from the five Black African cultures. After Procrustes rotation, congruence coefficients with the normative structure were .96, .91, .88, .95, and .96 for N, E, O, A, and C, respectively. Thus, Africans appear to share the common FFM (although, of course, they may also have additional, emic aspects of personality that set them apart from non-Africans).

This project included the first NEO-PI-R studies of Arabic cultures. In Lebanon and Morocco, respondents used the English version, and both had very low quality scores. Factor replication was weak in the Moroccan sample but good in the Lebanese sample. Of most interest are the Kuwaiti data, which report the first use of an Arabic translation. That sample showed a fair replication of the O factor and clear replications of N, E, A, and C factors.

Sex Differences in Targets

Raters were instructed to describe a man or woman aged "18 to 21" or "over age 40" and were later asked to specify the age (or estimated age) of the target. About 6.4% of targets fell outside the prescribed age ranges or were missing estimated age, so for analyses of age and gender groups, we excluded them. We compared women with men on the factors[7] and facets of the NEO-PI-R, using within-culture z scores. Analyses of the five factors show that women score higher than men on all five factors (ds = 0.49, 0.15, 0.07, 0.32, and 0.14 for N, E, O, A, and C, respectively), especially N and A. These results closely resemble findings in self-report analyses (Costa et al., 2001).

More detailed analyses on the individual facets are summarized in Table 4 for college-aged and adult subsamples. For comparison,

Table 4
Mean z-Score Differences (d) Between Women and Men on NEO-PI-R Facets in Self-Reports and Observer Ratings

	Self-reports[a]		Observer ratings	
NEO-PI-R facet	College aged	Adult	College aged	Adult
N1: Anxiety	.32***	.43***	.42***	.54***
N2: Angry Hostility	.16***	.19***	.15***	−.02
N3: Depression	.17**	.29***	.19***	.29***
N4: Self-Consciousness	.22***	.23***	.28***	.31***
N5: Impulsiveness	.16**	.11*	−.01	−.11***
N6: Vulnerability	.28***	.36***	.29***	.34***
E1: Warmth	.24***	.23***	.11***	.29***
E2: Gregariousness	.20***	.14***	.15***	.26***
E3: Assertiveness	−.10*	−.27***	−.07*	−.24***
E4: Activity	.04	.11*	.07*	.16***
E5: Excitement Seeking	−.18***	−.38***	−.17***	−.25***
E6: Positive Emotions	.27***	.16***	.17***	.26***
O1: Fantasy	.12**	.06	.06*	.10***
O2: Aesthetics	.40***	.35***	.26***	.31***
O3: Feelings	.33***	.31***	.26***	.42***
O4: Actions	.11**	.17**	.07*	.21***
O5: Ideas	−.17***	−.16*	−.19***	−.31***
O6: Values	.15**	.01	−.02	.09***
A1: Trust	.10*	.17***	.08**	.16***
A2: Straightforwardness	.34***	.32***	.09**	.17***
A3: Altruism	.25***	.25***	.10***	.33***
A4: Compliance	.03	.17***	.01	.17***
A5: Modesty	.22***	.22***	.19***	.26***
A6: Tender-Mindedness	.26***	.28***	.19***	.39***
C1: Competence	−.09	−.10	−.03	−.17***
C2: Order	.09	.10**	.19***	.24***
C3: Dutifulness	.18***	.13*	.13***	.09***
C4: Achievement Striving	.06	−.04	.14***	−.12***
C5: Self-Discipline	.09*	.04	.14***	.05*
C6: Deliberation	−.04	−.06	.10***	−.02

Note. For self-reports, Ns = 10,952 college age, 10,690 adults; for observer ratings, Ns = 5,095 college age, 6,128 adults. NEO-PI-R = Revised NEO Personality Inventory; N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness.
[a] From Costa, Terracciano, and McCrae, 2001.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4 also reproduces data from Costa et al. (2001), which examined self-reports. It is clear that the pattern of sex differences in observer ratings of personality is very similar to what had been seen before, despite a substantial difference in the cultures examined and a different method of measurement. The rank-order correlations between the four columns in Table 4 range from .72 to .88 (all ps $< .01$). The most notable difference between results from the two methods concerns N5: Impulsiveness, which is higher for women in self-reports and for men in observer ratings.

[7] In this and all subsequent analyses in this article, factor scores are calculated by applying factor scoring weights from the NEO-PI-R manual (Costa & McCrae, 1992a) to facet scores standardized within culture. The correlation of these factor scores with factor scores derived directly from Table 3 ranged from .98 to .99.

This difference had previously been found in analyses of Russian data (McCrae, Costa, Martin, et al., 2004). It is also noteworthy that sex differences in A2: Straightforwardness are larger when self-reports are analyzed, whereas differences in C2: Order are larger when observer ratings are analyzed.

There are a few instances of age effects on sex differences that are replicated across method. Adult women scored higher than adult men in E4: Activity and A4: Compliance, whereas college-aged women scored higher than college-aged men in C5: Self-Discipline. Perhaps the most interesting pattern is found for C4: Achievement Striving, in which adult men were rated higher than adult women, but college-aged men were rated lower than college-aged women (the same trend was seen in self-reports, although it did not reach significance). This finding suggests a role reversal across generations, perhaps reflecting increased vocational aspirations in young women around the world or diminished aspirations in older women with commitments to family.

These subtleties aside, the chief message of Table 4 is the universality of sex differences across methods, age groups, and cultures. In particular, men are rated as being higher than women in E3: Assertiveness, E5: Excitement Seeking, and O5: Ideas. Women are rated as higher on many traits, especially N1: Anxiety, N6: Vulnerability, O2: Esthetics, O3: Feelings, and A6: Tender-Mindedness. Most of these effects, however, are small, with only a single instance of more than .50 standard deviation difference.

We next examined cultural variations in sex differences.[8] Following Costa et al. (2001), we created four indices on which men and women could be compared. Two of these were the N and A factors, in which women tended to score higher on all facets. However, sex differences vary by facet within the other domains; for example, women are typically higher in E1: Warmth but lower in E3: Assertiveness. We therefore created a composite to represent sex differences within the E domain, defining Feminine Extraversion/Introversion (F-Ex/In) as (E1: Warmth + E2: Gregariousness − E3: Assertiveness − E5: Excitement Seeking + E6: Positive Emotions)/5. Similarly, we created a Feminine Openness/Closedness composite (F-Op/Cl) as (O2: Esthetics + O3: Feelings + O4: Actions − O5: Ideas)/4. For the present study we also included a fifth composite, Feminine Conscientiousness/Unconscientiousness (F-Co/Un), defined as (C2: Dutifulness + C3: Order − C1: Competence)/3. On all these composites, women are hypothesized to score higher than men.[9]

As shown in Table 5, the directions of the effects are uniform across cultures, with only six negative values (2.4%). Overall, the magnitude of sex differences is relatively small, none reaching a full standard deviation. However, there were systematic cultural differences in the magnitude of sex differences. As in self-reports, the magnitude of each of these indices of sex differences was correlated across cultures: Those cultures in which sex differences in one domain were pronounced tended also to have large sex differences in other domains. The rank-order correlations among the five columns in Table 5 ranged from .16 to .79, with all correlations except that between F-Op/Cl and F-Co/Un significant at $p < .01$. The cultures are listed in ascending order of the overall magnitude of gender differences (the sum of the five columns). As in self-report data, Asian and African cultures generally show the smallest sex differences, whereas European and American cultures show the largest.

Best and Williams (1994) proposed that the magnitude of gender differences might be understood by correlating, at the level of cultures, differences in masculinity/femininity with cultural comparison variables, such as gross domestic product per capita (GDP) and Hofstede's (2001) dimensions of culture. Costa et al. (2001) had hypothesized that Hofstede's Masculinity should be related to more pronounced sex differences, but no significant relation was found in that study. Given that the basic contrast appears to be between European and non-European cultures, it is likely that magnitude of sex differences will be related to a host of culture-level variables that differentiate Europe from Asia and Africa, including GDP and Hofstede's Individualism and Power Distance (the acceptance of a hierarchical social order). In the present data, the rank-order correlations between overall sex differences and Hofstede's variables were significant for Individualism ($r = .59$, $N = 47$, $p < .001$), Power Distance ($r = –.61$, $N = 47$, $p < .001$), and Masculinity ($r = .29$, $N = 47$, $p < .01$). Gender differentiation was also related to GDP ($r = .75$, $N = 49$, $p < .001$). Rich, individualistic, egalitarian, and masculine cultures have marked sex differences in personality.

The overall magnitude of sex differences may also be related to cultural differences in mate preference variables. Using evolutionary arguments, Buss (1989) predicted and showed that men and women have different goals in selecting mates: Women value earning capacity and industriousness, whereas men look for youth, physical attractiveness, and chastity. Buss also noted that there are cultural differences in the importance that men and women place on these qualities. In fact, these cultural differences are systematic, and a factor analysis of Buss's 10 variables across cultures shows a single general factor. The rank-order correlation of this factor score with our overall sex difference score was –.85 ($N = 21$, $p < .001$). It appears that where personality differences between the sexes are marked, qualities that promote reproduction are less important in mate selection. Perhaps evolutionary psychologists could suggest an explanation for this phenomenon.

One nonevolutionary explanation is artifact. Rates of acquiescence tend to be higher in non-Western nations (Smith, 2004), and the scores that Buss (1989) reported were based on single ratings with no control for acquiescence. At the same time, it is possible that gender differences are obscured across all factors in cultures in which data quality is relatively poor—again, predominantly non-Western nations. The rank-order correlation between quality (see Table 2) and the overall magnitude of sex differences in Table 5 is .71 ($p < .001$). We return to this hypothesis below in evaluating cultural variation in age differences.

Costa et al. (2001) argued that the most plausible reason for cultural variability in the magnitude of sex differences was attri-

[8] Canada was omitted from these analyses because no adult men were included. In Canadian data, college-aged women scored higher than college-aged men in N; adult women scored higher than college-aged women in A and C ($p < .05$).

[9] In Costa et al. (2001), gender differences on these indices are incorrectly described as z scores when in fact they are composites of z-scored facets and have standard deviations less than 1.00. For consistency with previous results, we continue this metric for gender differences. The rank-order correlation between total gender difference based on this scoring and total gender difference based on z-standardized composite scores is .99.

Table 5
Mean Differences Between Women and Men in 50 Cultures on Revised NEO Personality Inventory Factors or Composites

Culture	N	A	F-Ex/In	F-Op/Cl	F-Co/Un
Nigeria	.00	.00	−.04	.00	.00
India	.13	.13	−.05	−.03	.07
Botswana	.08	.06	−.01	.13	.03
Ethiopia	.13	−.02	.01	.15	.05
Russia	.25*	−.02	.14**	.19**	.11*
Puerto Rico	.23	.11	.09	.20*	.05
Uganda	.30	.23	.13	.00	.08
Morocco	.16	.26	.20*	.08	.11
Mexico	.33*	.17	.10	.26**	.05
Croatia	.49**	.03	.17**	.25***	.10
Indonesia	.40**	.22	.15*	.21**	.08
Peru	.57***	.16	.06	.21*	.13
Malaysia	.36**	.39**	.17**	.13*	.17**
Kuwait	.48***	.31**	.19***	.09	.16**
Thailand	.41**	.31*	.15*	.13*	.24***
Philippines	.52***	.32*	.11	.25**	.04
Serbia	.54***	.10	.19**	.34***	.11
China	.56***	.33*	.19*	.13	.12
Brazil	.59***	.25**	.11**	.26***	.15***
Chile	.48**	.27	.21**	.23**	.18**
Poland	.42**	.29*	.22***	.27***	.18**
Portugal	.45**	.31*	.19**	.11	.33***
Italy	.55***	.22	.22**	.23**	.19**
Argentina	.20	.46**	.39***	.30***	.10
Malta	.67***	.24	.21**	.14	.20**
Japan	.70***	.11	.21**	.38***	.09
Lebanon	.39**	.34*	.23***	.34***	.19**
United States	.59***	.29***	.23***	.30***	.09**
France	.48***	.39**	.17**	.33***	.21***
Turkey	.46**	.49**	.25***	.35***	.06
Hong Kong	.51***	.41**	.22***	.22**	.26***
Burkina Faso	.59***	.46**	.25***	.20**	.12*
Slovakia	.68***	.29*	.16*	.23***	.30***
South Korea	.48**	.50***	.26***	.40***	.08
Slovenia	.29	.53***	.34***	.45***	.11
Germany	.54***	.50***	.28***	.32***	.13**
Estonia	.56***	.49***	.25***	.39***	.15**
Iceland	.57***	.45**	.20**	.37***	.22***
Belgium	.52***	.47***	.33***	.36***	.14*
Spain	.64***	.45**	.26***	.28***	.20**
United Kingdom: Northern Ireland	.66**	.43*	.19*	.30**	.25**
Denmark	.70***	.35*	.23**	.36***	.20**
Australia	.76***	.42**	.19**	.36***	.17**
New Zealand	.54***	.50***	.33***	.40***	.11
French Switzerland	.79***	.42**	.28***	.22**	.33***
German Switzerland	.80***	.49**	.37***	.42***	.10
Austria	.67***	.54**	.40***	.42***	.21*
Czech Republic	.70***	.64***	.34***	.48***	.34***
United Kingdom: England	.78***	.84***	.43***	.28***	.20**

Note. N = Neuroticism, A = Agreeableness, F-Ex/In = Feminine Extraversion/Introversion, F-Op/Cl = Feminine Openness/Closedness, F-Co/Un = Feminine Conscientiousness/Unconscientiousness.
* $p < .05$. ** $p < .01$. *** $p < .001$.

bution of characteristics to roles. In cultures with traditional gender stereotypes, sex-typical behavior is perceived as a reflection of role requirements rather than individual traits and correspondingly discounted in forming an impression of the individual. This argument presumes that most or all cultures share the same gender stereotypes, an assumption that appears to be warranted (Williams & Best, 1990). It also implies that Western observers would perceive sex differences in personality even in non-Western targets, which is a testable hypothesis.

In a final analysis of sex differences in targets, we examined agreement across cultures on sex differences at the facet level. We calculated *d* scores for each facet for the 49 cultures (excluding Canada) and intercorrelated the cultures across the 30 facets. We factored these correlations, which indicate the similarity of facet-level gender differences between pairs of cultures, and found a large first factor. With one exception, all cultures had positive loadings on this factor, ranging from .36 for Morocco to .92 for Australia, which showed the most prototypic pattern of gender

differences. The single exception was Nigeria (–.20), where none of the gender differences was significant.

Sex Differences in Raters

Target categories were randomly assigned to raters, so it is possible to examine results in terms of the sex of the rater: Did women differ systematically from men in the mean levels of the traits they ascribed to targets? We conducted *t* tests of factors and facets standardized within culture, comparing male and female raters separately for male and female targets. These analyses led to two general conclusions. First, the magnitude of rater biases was very low: Of 70 comparisons (35 scores each for male and female targets), only 14 were as large as .10 standard deviations, and none was larger than .18 standard deviations. Such small differences cannot account for the observed sex differences in self-reports (see Table 4, columns 2–3); thus, when women rate themselves as higher in N, it is more probably due to real differences in N than to a bias by women in rating N.

Second, the 14 largest effects all suggest that women are more lenient than men in describing others, especially other women. Female raters described both men and women as more straightforward and altruistic than did male raters. When rating women, female raters described them as being less anxious, less self-conscious, and less vulnerable and warmer, more gregarious, more open to ideas and values, and more competent than did male raters. All these findings are consistent with the observation that women in general are more agreeable than men and that agreeable raters make more lenient ratings of others (Bernardin, Cooke, & Villanova, 2000).

Age Differences

In the full sample, the college-aged group ranged in age from 18–21 years (M = 19.8); the adult group was aged 40–98 years (M = 49.9). Working with z scores standardized within culture, and separately for men and women, we calculated the mean differences between adult and college-aged samples for the factors and facets. Results are given in Table 6. Because varying age groups were used in previous self-report studies, it is not possible to conduct a quantitative comparison of effect sizes from those studies with the present results. The first column of Table 6 therefore summarizes previous research in terms of the net number of significant effects in 12 cultures. For example, E4: Activity had significant negative correlations with age in four cultures and significant positive correlations in two; the net effect is thus listed as "Down (2)."

In general, age differences in both men and women replicate self-report results. When the second column is coded from −12 to +12, rank-order correlations between the four data columns in Table 6 range from .84 to .99 (all ps $< .01$). The largest age differences are found for Conscientiousness and its facets, which increase with age, and for E5: Excitement Seeking, E6: Positive Emotions, and O1: Fantasy, which decline with age.

More noteworthy is the very limited effect of age on observer-rated N and A, which typically show effects comparable in magnitude to those of E, O, and C in self-report studies (e.g., McCrae et al., 1999). A closer examination of the facets shows the reason for these muted effects. Although N1: Anxiety and N2: Angry Hostility typically decline with age in self-reports (like the other N facets), in these observer ratings they increase—a finding previously reported in Czech data (McCrae, Costa, Hřebíčková, et al., 2004). In the A domain, most of the effects are small, with observer-rated A1: Trust lower in adults than in college-aged targets. College student raters appear to regard their elders as higher in negative emotionality and mistrust than adults view themselves. In future studies it would be useful to gather ratings from adults as well to see if these findings are attributable to age of rater.

In addition to the contrast of college-aged and adult groups, we also conducted regressions predicting personality factors from age

Table 6
Mean z-Score Differences (d) Between Adult and College-Aged Targets on NEO-PI-R Facets

NEO-PI-R factor/facet	Self-report trend[a]	Observer ratings		
		Men	Women	Total
N	Down (7)	−.08**	−.01	−.04*
E	Down (10)	−.48***	−.35***	−.41***
O	Down (11)	−.36***	−.34***	−.35***
A	Up (12)	−.01	.22***	.11***
C	Up (12)	.68***	.56***	.62***
N1: Anxiety	Down (5)	.02	.14***	.09***
N2: Angry Hostility	Down (6)	.15***	−.02	.06**
N3: Depression	Down (7)	−.23***	−.13***	−.17***
N4: Self-Consciousness	Down (7)	−.18***	−.15***	−.16***
N5: Impulsiveness	Down (12)	−.31***	−.42***	−.37***
N6: Vulnerability	Down (9)	−.31***	−.25***	−.27***
E1: Warmth	Down (2)	−.20***	−.01	−.10***
E2: Gregariousness	Down (8)	−.44***	−.32***	−.38***
E3: Assertiveness	Down (4)	.36***	.19***	.27***
E4: Activity	Down (2)	−.02	.07**	.03
E5: Excitement Seeking	Down (11)	−.79***	−.88***	−.84***
E6: Positive Emotions	Down (10)	−.48***	−.38***	−.43***
O1: Fantasy	Down (12)	−.61***	−.58***	−.59***
O2: Aesthetics	Down (9)	−.35***	−.31***	−.32***
O3: Feelings	Down (12)	−.38***	−.22***	−.29***
O4: Actions	Down (10)	−.41***	−.27***	−.33***
O5: Ideas	Down (10)	−.02	−.14***	−.09***
O6: Values	Down (11)	−.24***	−.13***	−.18***
A1: Trust	Up (9)	−.11***	−.03	−.07**
A2: Straightforwardness	Up (12)	.21***	.29***	.25***
A3: Altruism	Up (8)	.07*	.30***	.19***
A4: Compliance	Up (11)	.04	.20***	.13***
A5: Modesty	Up (12)	.13***	.20***	.17***
A6: Tender-Mindedness	Up (11)	−.04	.16***	.07***
C1: Competence	Up (10)	.49***	.36***	.42***
C2: Order	Up (9)	.42***	.47***	.45***
C3: Dutifulness	Up (12)	.61***	.57***	.59***
C4: Achievement Striving	Up (8)	.42***	.17***	.29***
C5: Self-Discipline	Up (12)	.61***	.52***	.57***
C6: Deliberation	Up (12)	.53***	.41***	.46***

Note. *N*s = 5,338 men, 5,885 women. NEO-PI-R = Revised NEO Personality Inventory; N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness.
[a] Trend shows the direction and net number of significant age associations in self-report data from 12 cultures: Germany, Italy, Portugal, Croatia, South Korea, the United States, Russia, Japan, Estonia, China, Zimbabwe, and the Czech Republic (McCrae & Costa, in press).
* $p < .05$. ** $p < .01$. *** $p < .001$.

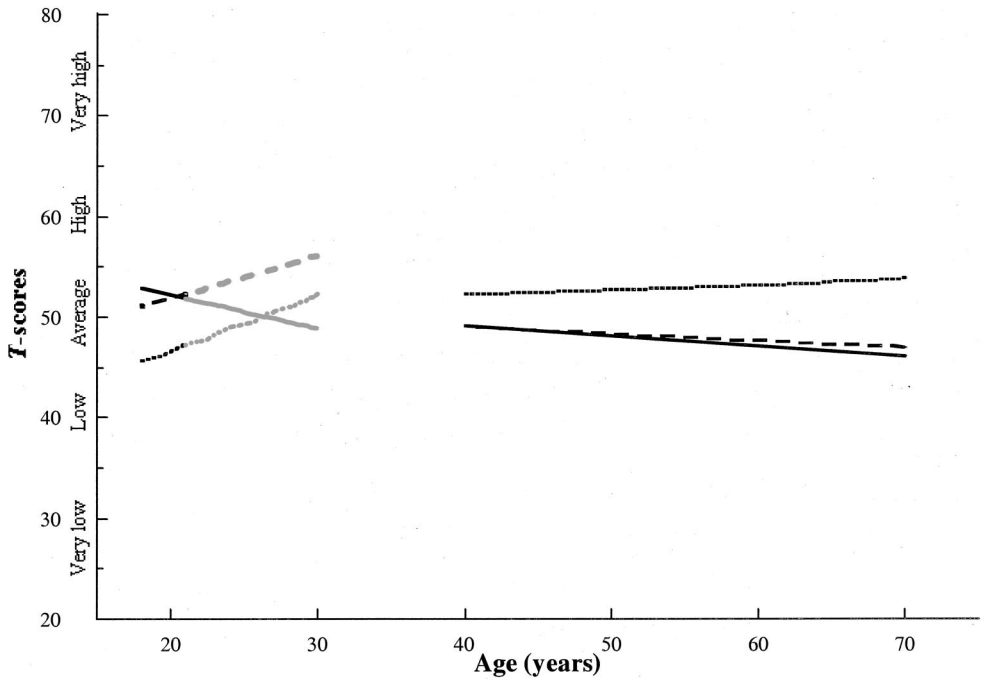

Figure 1. Linear regression lines predicting Extraversion (solid lines), Openness (dashed lines), and Conscientiousness (dotted lines) from age. The scale of the vertical axis was chosen because it is the range used in the interpretation of individual profiles (Costa & McCrae, 1992a); age group trends are seen relative to the range of individual scores. Lines on the left are based on data from 18- to 21-year-olds, extrapolated to age 30 (gray line segments) to illustrate the slope. Lines on the right are based on data from 40- to 90-year-olds.

within each of these age groups. Robins, Fraley, Roberts, and Trzesniewski (2001) reported cross-sectional increases in O, A, and C and a decline in N during the college years. In our college-aged sample, we replicated the increases in O and C but not the effects for A or N. Further, there was a significant decline in E over this age period, although it amounted to only about 1 *T*-score point. Costa and McCrae (2002) predicted that after age 30, N, E, and O would decline, whereas A and C would not change. In the present data, E and O declined significantly, but N did not. As hypothesized, A did not change, but C increased with age.

The magnitude of age changes is also of interest. As shown in Figure 1, the rate of cross-sectional change from ages 18 to 21 was substantially larger than that seen after age 40. These trends are consistent with Costa and McCrae's (2002) view that after age 30, personality changes are very modest. Except for O, however, they differ in details: Costa and McCrae did not anticipate that the decline in E would begin during college age, nor did they predict an increase in C after age 30.

Because most facets follow the same age trend as the factor they define, cultural variations in age differences can be addressed at the factor level. Table 7 reports effect sizes (*d*) and statistical tests for age group differences in all cultures. Effects for E, O, and C do appear to be pancultural, with E and O higher and C lower among college-aged targets in almost every culture. The picture is much less clear for N and A. Only six cultures show the hypothesized decline of N with age, and in two cultures—Estonia and Slovakia—adults scored significantly higher than college-aged targets. Ten of the 12 significant age effects for A show the expected increase with age, but adults were viewed as more disagreeable than college-aged targets in Japan and Portugal. There are no obvious explanations for these anomalies.

Just as there are sex roles to which behaviors may be attributed, there are also age roles, and we might hypothesize that more traditional cultures would attribute differences in behavior of younger and older adults to these roles, thus diminishing perceived age differences in all aspects of personality. However, the data in Table 7 do not show cross-domain consistency: When values for age differences in A and C are reflected so that college-aged targets tend to score higher on all factors, the rank-order correlations of the five columns in Table 7 range from –.45 to +.45 (*Mdn* = –.12). For example, New Zealand shows a large effect for C, a moderate effect for E, but no age difference for O. It thus does not appear to be possible to generalize about cultural variations in the magnitude of perceived age differences in personality traits,[10] so the cultures in Table 7 are listed in alphabetical order.

When the magnitude of age differences (adult minus college age) is correlated with other culture-level variables for each factor, a few significant findings emerge (rank-order *r*s = .32 to .47). Cultures showing larger age differences in E and smaller age differences in C score higher in Power Distance and lower in Individualism and GDP. Differences in O and A are positively related to Individualism. Data quality was associated with the

[10] Reanalyses of self-report data for the five domains in McCrae (2002) give similar results, with rank-order correlations ranging from –.24 to .56 (*Mdn* = .14).

Table 7
Mean z-Score Differences (d) Between Adult and College-Aged Targets in 42 Cultures on Revised NEO Personality Inventory Factors

Culture	N	E	O	A	C
Argentina	.15	−.40**	−.30*	.36*	.84***
Australia	−.12	−.27	.03	.41**	.72***
Austria	−.02	−.77***	−.72***	−.01	.53**
Belgium	.06	−.86***	−.18	−.03	.86***
Botswana	−.38*	−.05	.02	−.15	.41*
Brazil	−.01	−.52***	−.35***	.30**	.74***
Burkina Faso	−.08	−.08	−.49***	.41**	.80***
Chile	−.10	−.64***	−.39**	−.22	.85***
China	.04	−.26	−.62***	.07	.34*
Croatia	.20	−.57***	−.25	.04	.66***
Czech Republic	.06	−.48***	−.79***	.03	.33*
Denmark	.33	−.48**	−.15	.10	.91***
Estonia	.37**	−.55***	−.25*	.22	.86***
Ethiopia	−.10	−.06	−.79***	−.03	.15
France	−.08	−.60***	−.12	.00	.75***
French Switzerland	.03	−.52***	−.19	.09	.77***
German Switzerland	−.21	−.64*	.03	−.13	.41
Germany	−.32***	−.40***	−.51***	.29**	.50***
Hong Kong	.20	−.95***	−.90***	−.08	.57***
Iceland	−.29*	−.50***	−.08	−.09	.84***
India	.01	−.42**	−.03	−.15	.32*
Indonesia	−.15	−.29*	−.45**	−.16	.26
Italy	.04	−.57***	−.29*	.12	.77***
Japan	−.03	−.40**	−.52***	−.29*	.49**
Kuwait	−.27*	−.27*	−.60***	.31**	.66***
Lebanon	.11	−.59***	−.15	.19	.81***
Malaysia	−.46***	−.18	−.41**	.40**	.55***
Malta	.06	−.43**	−.53***	.16	.83***
Mexico	.10	−.12	−.55***	−.13	.54**
Morocco	−.23	.09	−.10	−.03	.43*
New Zealand	.16	−.61**	.00	.14	.94***
Nigeria	.00	.10	−.13	.10	.05
Peru	.26	−.50**	−.36*	−.06	.73***
Philippines	−.19	−.44**	−.43**	−.22	.52***
Poland	−.13	−.63***	−.67***	−.19	.38**
Portugal	.27	−.78***	−.60***	−.29*	.46**
Puerto Rico	−.20	−.08	−.37*	.07	.63***
Russia	.12	−.42***	−.34**	.62***	.22
Serbia	.00	−.63***	−.18	.08	.63***
Slovakia	.42**	−.51***	−.61***	.00	.82***
Slovenia	−.14	−.36*	−.60***	.20	.52***
South Korea	−.06	−.13	−.59***	−.15	.50***
Spain	−.01	−.53***	−.43**	.11	.87***
Thailand	−.13	−.44**	−.22	−.01	.51***
Turkey	−.18	−.49***	−.28*	.20	.89***
United Kingdom: England	.02	−.65***	.10	.34*	.98***
United Kingdom: Northern Ireland	.25	−.61**	−.27	.12	.87***
Uganda	−.16	−.10	−.56***	−.07	.28
United States	−.17*	−.38***	−.27***	.24***	.82***

Note. N = Neuroticism; E = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness.
* $p < .05$. ** $p < .01$. *** $p < .001$.

magnitude of age differences in E and C (rank-order *r*s = –.65 and .48, *p*s < .001), but not N, O, or A. If data quality were a complete explanation for cultural variation in the magnitude of age differences, it should be related to all five factors. It seems likely that variations in quality contribute to, but do not fully account for, the observed age effects, and the same is likely true for sex effects.

If attributional processes account for the diminished sex differences in traditional cultures, why do they not also lead to diminished age differences? The attribution argument assumes that gender stereotypes are the same everywhere and are basically accurate. Under these conditions, behavior can plausibly be attributed either to traits or to roles, and it is more likely to be attributed to roles in traditional societies that emphasize sex roles. In the case of age differences, it is possible that age stereotypes are not as widely shared as gender stereotypes or not as accurate. Some evidence has been presented for the cross-cultural generalizability of age stereotypes (Harwood et al., 2001), but accuracy is questionable. There is, for example, a common belief that older adults are prone to depression, but that is not supported by epidemiological evidence (Copeland et al., 1999).

It is also possible that attribution effects are present but are dwarfed by other causes of cultural variation in age differences. Different societies have had very different recent histories, and these may have left imprints on successive birth cohorts. Their social history may explain why today's adult New Zealanders are seen as so much more conscientious and so little less open than today's college-aged New Zealanders. Perhaps the most fruitful way to approach these questions is by seeking common characteristics of cultures that share similar levels of age differences for each factor: What features of history or culture do Belgium, Hong Kong, and Portugal share that might account for large age differences in E? Are these features lacking in Botswana, Malaysia, and South Korea, where differences in E are small?

Education Effects

We examined the associations of personality traits with target's education in the adult sample. (Russia, Germany, and Austria did not provide data on education.) Correlations with N, E, O, A, and C factors were –.10, –.03, .22, .01 (*ns*), and .11, respectively (N = 5,394). These are very similar to correlations reported for American Form R data (–.10, .07, .22, .06, .10; Costa & McCrae, 1992a), which in turn are close to the Form S findings. Similar results were found for men and women. In the present data, the facets most strongly related to education were O5: Ideas (r = .24) and C1: Competence (r = .17). Effects were similar across cultures; the strongest correlations were seen for O, which was positively related to education in 45 of 47 cultures, significantly so in 33 of them. It appears that education is systematically related to personality traits across a wide variety of cultures but that the effects are generally rather small.

Conclusion

In the mid-19th century, the German "father of ethnography," Adolf Bastian, proposed the idea of the "psychic unity of mankind" (Koepping, 1983). His fundamental notion, a progressive one at the time, was that all human beings were a single species and must therefore share all basic cognitive and psychological characteristics. More recent anthropologists have been unwilling to root psychology so deeply in biology and have argued that culture shapes psychology (Shweder & Sullivan, 1993). The present data, which largely confirm recent findings of universality in trait psychology in a new sample of cultures using a different method of measurement, give strong support to Bastian's hypothesis of psychic unity and could be interpreted as evidence of the biological basis of personality traits (Allik & McCrae, 2002).

The present article goes beyond replication in several respects. It proposes an independent index of data quality for cross-cultural comparisons and shows that quality—generally a reflection of the fit between the assessment instrument and the cultural background and experience of the sample—could explain cultural variation in internal consistency and factor replicability. It provides new data on the FFM in Arabic and Black African cultures that have rarely been studied before. It provides cross-cultural evidence of gender differences in person perception, showing that women are more positive than men in their assessments of others, especially other women. It demonstrates that there are differences in the perception of age differences in Neuroticism and Agreeableness by the self and external observers (cf. McCrae, Costa, Hřebíčková, et al., 2004), and points to a new puzzle: Why are perceived sex differences in personality traits consistently attenuated in traditional cultures whereas perceived age differences are not?

There are, of course, limitations in the present study. The participants within each culture are samples of convenience, and most raters were college aged, so adult perspectives on personality are lacking. Perhaps most significant was the use of a single questionnaire, the NEO-PI-R, as a tool for assessing personality. The fixed items of this instrument preclude the discovery of emic dimensions of personality that might be found in some cultures (cf. Cheung & Leung, 1998), and the questionnaire format, requiring decontextualized trait assessments, may be difficult for individuals accustomed to describing people within the context of interpersonal relationships (see Church, 2000). Although standard questionnaires can be used in college student samples around the world, future research might seek alternate assessment methods (e.g., structured interviews; Trull & Widiger, 1997) that might be more appropriate in collectivistic cultures.

This article focused on comparisons of individuals within cultures to test the universality of trait psychology. These are analyses at the transcultural level (McCrae, 2000). However, there is also a recent literature on intercultural comparisons, relating the mean levels of personality traits in a culture to cultural values (Hofstede & McCrae, 2004) and geographic proximity (Allik & McCrae, 2004). Because the present article demonstrates the pancultural viability of observer assessments of personality, aggregated data from the present study can now be used to address these culture-level questions.

References

Allik, J., & McCrae, R. R. (2002). A five-factor theory perspective. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 303–322). New York: Kluwer Academic/Plenum Publishers.

Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology, 35,* 13–28.

Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and Agreeableness as predictors of rating leniency. *Journal of Applied Psychology, 83,* 232–234.

Best, D. L., & Williams, J. E. (1994). Masculinity/femininity in the self and ideal self descriptions of university students in fourteen countries. In A.-M. Bouvy & F. J. R. van de Vijver (Eds.), *Journeys into cross-cultural psychology* (pp. 297–306). Lisse, the Netherlands: Swets & Zeitlinger.

Bond, M. H. (2001). Surveying the foundations: Approaches to measuring group, organizational, and national variation. In M. Erez & U. Kleinbeck (Eds.), *Work motivation in the context of a globalizing economy* (pp. 395–412). Mahwah, NJ: Erlbaum.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12,* 1–49.

Carter, J. A., Herbst, J. H., Stoller, K. B., King, V. L., Kidorf, M. S., Costa, P. T., Jr., & Brooner, R. K. (2001). Short-term stability of NEO-PI-R personality trait scores in opioid-dependent outpatients. *Psychology of Addictive Behaviors, 15,* 255–260.

Cheung, F. M., & Leung, K. (1998). Indigenous personality measures: Chinese examples. *Journal of Cross-Cultural Psychology, 29,* 233–248.

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

Copeland, J. R. M., Beekman, A. T. F., Dewey, M. E., Jordan, A., Lawlor, B. A., Linden, M., et al. (1999). Cross-cultural comparison of depressive

symptoms in Europe does not support stereotypes of ageing. *British Journal of Psychiatry, 174,* 322–329.

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

Costa, P. T., Jr., & McCrae, R. R. (1992b). Trait psychology comes of age. In T. B. Sonderegger (Ed.), *Nebraska Symposium on Motivation 1991: Psychology and Aging* (pp. 169–204). Lincoln: University of Nebraska Press.

Costa, P. T., Jr., & McCrae, R. R. (2002). Looking backward: Changes in the mean levels of personality traits from 80 to 12. In D. Cervone & W. Mischel (Eds.), *Advances in personality science* (pp. 219–237). New York: Guilford Press.

Costa, P. T., Jr., & McCrae, R. R. (in press). Trait and factor theories. In J. C. Thomas & D. L. Segal (Eds.), *Comprehensive handbook of personality and psychopathology* (Vol. 1). New York: Wiley.

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.

Feldman Barrett, L., Lane, R. D., Sechrest, L., & Schwartz, G. E. (2000). Sex differences in emotional awareness. *Personality and Social Psychology Bulletin, 26,* 1027–1035.

Harwood, J., Giles, H., McCann, R. M., Cai, D., Somera, L. P., Ng, S. H., et al. (2001). Older adults' trait ratings of three age-groups around the Pacific rim. *Journal of Cross-Cultural Gerontology, 16,* 157–171.

Haven, S., & ten Berge, J. M. F. (1977). *Tucker's coefficient of congruence as a measure of factorial invariance: An empirical study* (Heymans Bulletin 290 EX). Groningen, the Netherlands: University of Groningen.

Heaven, P. C. L., Connors, J., & Stones, C. R. (1994). Three or five personality dimensions? An analysis of natural language terms in two cultures. *Personality and Individual Differences, 17,* 181–189.

Hoekstra, H. A., Ormel, J., & De Fruyt, F. (1996). *Handleiding NEO Persoonlijkheids-vragenlijsten NEO-PI-R en NEO-FFI* [Manual for NEO personality inventories NEO-PI-R and NEO-FFI]. Lisse, the Netherlands: Swets & Zeitlinger.

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

Hofstede, G., & McCrae, R. R. (2004). Personality and culture revisited: Linking traits and dimensions of culture. *Cross-Cultural Research, 38,* 52–88.

Koepping, K.-P. (1983). *Adolf Bastian and the psychic unity of mankind: The foundations of anthropology in nineteenth century Germany.* St. Lucia, Queensland, Australia: University of Queensland Press.

Marsella, A. J., Dubanoski, J., Hamada, W. C., & Morse, H. (2000). The measurement of personality across cultures: Historical, conceptual, and methodological issues and considerations. *American Behavioral Scientist, 44,* 41–62.

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

McCrae, R. R. (2000). Trait psychology and the revival of personality and culture studies. *American Behavioral Scientist, 44,* 10–31.

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. (in press). What is personality? In R. Colom & C. Flores-Mendoza (Eds.), *Introduction to the psychology of individual differences* [in Portuguese]. Porto Alegre, Brazil: ArtMed Publishers.

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

McCrae, R. R., & Costa, P. T., Jr. (in press). Cross-cultural perspectives on adult personality trait development. In D. Mroczek & T. Little (Eds.), *Handbook of personality development.* Hillsdale, NJ: Erlbaum.

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., 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., Martin, T. A., Oryol, V. E., Rukavishnikov, A. A., Senin, I. G., et al. (2004). Consensual validation of personality traits across cultures. *Journal of Research in Personality, 38,* 179–201.

McCrae, R. R., Costa, P. T., Jr., Ostendorf, F., Angleitner, A., Hřebíčková, M., Avia, M. D., et al. (2000). Nature over nurture: Temperament, personality, and lifespan development. *Journal of Personality and Social Psychology, 78,* 173–186.

McCrae, R. R., Terracciano, A., & Khoury, B. (in press). Dolce far niente: The positive psychology of personality stability and invariance. In A. Ong & M. van Dulmen (Eds.), *Handbook of methods in positive psychology.* New York: Oxford University Press.

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.

Okeke, B. I., Draguns, J. G., Skeku, B., & Allen, W. (1999). Culture, self, and personality in Africa. In Y.-T. Lee, C. R. McCauley, & J. G. Draguns (Eds.), *Personality and person perception across cultures* (pp. 139–162). Mahwah, NJ: Erlbaum.

Ostendorf, F., & Angleitner, A. (2004). *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.

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.

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales in volunteer samples: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78,* 582–593.

Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality, 69,* 617–640.

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.

Rossier, J., Dahourou, D., & McCrae, R. R. (in press). Structural and mean level analyses of the five-factor model and locus of control: Further evidence from Africa. *Journal of Cross-Cultural Psychology.*

Schmitt, D. P., & 118 Members of the International Sexuality Description Project. (2003). Universal sex differences in the desire for sexual variety: Tests from 52 nations, 6 continents, and 13 islands. *Journal of Personality and Social Psychology, 85,* 85–104.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1–65). New York: Academic Press.

Shimonaka, Y., Nakazato, K., Gondo, Y., & Takayama, M. (1999). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual for the Japanese version* [in Japanese]. Tokyo: Tokyo Shinri.

Shweder, R. A., & Sullivan, M. A. (1993). Cultural psychology: Who needs it? *Annual Review of Psychology, 44,* 497–523.

Smith, P. B. (2004). Acquiescent response bias as an aspect of cross-cultural communication style. *Journal of Cross-Cultural Psychology, 35,* 50–61.

Terracciano, A., Merritt, M., Zonderman, A. B., & Evans, M. K. (2003). Personality traits and sex differences in emotion recognition among African Americans and Caucasians. In P. Ekman, J. J. Campos, R. J. Davidson, & F. B. M. de Waal (Eds.), *Annals of the New York Academy of Sciences: Vol. 1000. Emotions inside out: 130 Years after Darwin's* The Expression of the Emotions in Man and Animals (pp. 309–312). New York: New York Academy of Sciences.

Trull, T. J., & Widiger, T. A. (1997). *Structured Interview for the Five-Factor Model of Personality (SIFFM): Professional manual.* Odessa, FL: Psychological Assessment Resources.

Williams, J. E., & Best, D. L. (1990). *Sex and psyche: Gender and self viewed cross-culturally.* Newbury Park, CA: Sage.

Appendix

Contributing Members of the Personality Profiles of Cultures Project

The 78 coauthors of this article, listed in order of data submission, are Brigitte Khoury, American University of Beirut Medical Center, Beirut, Lebanon; Florence Nansubuga, Makerere University, Kampala, Uganda; Goran Kneževič, University of Belgrade, Belgrade, Yugoslavia; Dragana Djuric Jocic, Institute for Psychiatry, Belgrade, Yugoslavia; Hyun-nie Ahn and Chang-kyu Ahn, Pusan National University, Pusan, South Korea; Filip De Fruyt, Ghent University, Ghent, Belgium; Sami Gülgöz, Koç University, Istanbul, Turkey; Willibald Ruch, University of Zürich, Zürich, Switzerland; M. Arif Ghayur, Al Akhawayn University, Ifrane, Morocco (now at San Diego State University); Maria D. Avia and Maria L. Sánchez-Bernardos, Universidad Complutense de Madrid, Madrid, Spain; Jérôme Rossier, University of Lausanne, Lausanne, Switzerland; Donatien Dahourou, University of Ouagadougou, Ouagadougou, Burkina Faso; Ronald Fischer, Victoria University of Wellington, Wellington, New Zealand; Jane Shakespeare-Finch, Queensland University of Technology, Brisbane, Australia; Michelle Yik, Hong Kong University of Science and Technology, Hong Kong, People's Republic of China; Peter B. Smith, University of Sussex, Brighton, United Kingdom; Alois Angleitner and Fritz Ostendorf, Universität Bielefeld, Bielefeld, Germany; Magdalena S. Halim, Atma Jaya Indonesia Catholic University, Jakarta, Indonesia; Martina Hřebíčková, Academy of Sciences of the Czech Republic, Brno, Czech Republic; Thomas A. Martin, Susquehanna University; Tilahun Sineshaw, Ramapo College of New Jersey; Andrzej Sekowski and Waldemar Klinkosz, Catholic University of Lublin, Lublin, Poland; Garry Prentice and Margaret McRorie, Queen's University, Belfast, Northern Ireland; Carmen Flores-Mendoza, Universidade Federal De Minas Gerais, Belo Horizonte, Brasil; Yoshiko Shimonaka, Bunkyo Gakuin University, Irumagun, Saitama, Japan; Katsuharu Nakazato, Iwate Prefectural University, Takizawa, Japan; Khairul A. Mastor, Universiti Kebangsaan Malaysia, Selangor Darul Ehsan, Malaysia; Claudio Barbaranelli, University of Rome "La Sapienza," Rome, Italy; Lidia Alcalay and Franco Simonetti, Pontificia Universidad Católica de Chile, Santiago, Chile; V. S. Pramila, Andhra University, Visakhapatnam, India; Ruth Falzon, Mary Ann Lauri, Mary Ann Borg Cunen, and Sandra Scicluna Calleja, University of Malta, Msida, Malta; Margarida Pedroso de Lima, Universidade de Coimbra, Coimbra, Portugal; Denis Bratko, University of Zagreb, and Iris Marušić, Center for Educational Research and Development, Zagreb, Croatia; Jüri Allik and Anu Realo, University of Tartu, Tartu, Estonia; Ahmed M. Abdel Khalek and Badr M. Alansari, Kuwait University, Safat, Kuwait; Gregorio E. H. del Pilar, University of the Philippines, Quezon City, the Philippines; A. Oluyinka Ojedokun, University of Ibadan, Ibadan, Nigeria; Margaret Munyae, University of Botswana, Gaborone, Botswana; Louisa Budzinski and Shigehiro Oishi, University of Minnesota; Ed Diener, University of Illinois at Urbana-Champaign; Niyada Chittcharat, Srinakharinwirot University, Bangkok, Thailand; Lei Wang, Peking University, Beijing, People's Republic of China; Andrew Beer and John Humrichouse, University of Iowa; Erik Lykke Mortensen and Hans Henrik Jensen, University of Copenhagen, Copenhagen, Denmark; Friðrik H. Jónsson, University of Iceland, Reykjavík, Iceland; Emília Ficková and Lucia Adamovová, Slovak Academy of Sciences, Bratislava, Slovakia; Velko S. Rus, University of Ljubljana, Nuska Podobnik, Psychiatric Hospital Idria, and Agata Zupancic, Ministry of Health of the Republic of Slovenia, Ljubljana, Slovenia; Rolando Diaz-Loving, Universidad Nacional Autónoma de México, Mexico City, Mexico; Nora B. Leibovich and V. Schmidt, University of Buenos Aires and National Research Council, Buenos Aires, Argentina; Norma Reátegui and Marina Brunner-Sciarra, Universidad Peruana Cayetano Heredia, Lima, Peru; Lindsay E. Ayearst and Krista K. Trobst, York University, Toronto, Ontario, Canada; David Matsumoto, San Francisco State University; Aljoscha Neubauer, Karl-Franzens-University Graz, Graz, Austria; José Porrata, University of Puerto Rico, San Juan, Puerto Rico; Jean-Pierre Rolland, Jean-Michel Petot, and Nathalie Camart, Paris 10-Nanterre University, Paris, France.

Received April 9, 2004
Revision received September 29, 2004
Accepted October 7, 2004 ■