Journal of Personality and Social Psychology
2005, Vol. 89, No. 3, 407–425

DOI: 10.1037/0022-3514.89.3.407

Personality Profiles of Cultures: Aggregate Personality Traits

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

79 Members of the Personality Profiles of Cultures Project

Personality profiles of cultures can be operationalized as the mean trait levels of culture members. College students from 51 cultures rated an individual from their country whom they knew well (N = 12,156). Aggregate scores on Revised NEO Personality Inventory (NEO-PI-R) scales generalized across age and sex groups, approximated the individual-level 5-factor model, and correlated with aggregate self-report personality scores and other culture-level variables. Results were not attributable to national differences in economic development or to acquiescence. Geographical differences in scale variances and mean levels were replicated, with Europeans and Americans generally scoring higher in Extraversion than Asians and Africans. Findings support the rough scalar equivalence of NEO-PI-R factors and facets across cultures and suggest that aggregate personality profiles provide insight into cultural differences.

Keywords: personality, five-factor model, cross-cultural, culture-level analyses

There is enormous appeal in the idea that cultures have distinctive personalities. Ruth Benedict's (1934) classic description of the Southwestern American Indian Pueblo culture as Apollonian—sober, conventional, cooperative, and orderly—seems apt and insightful. Yet one need not have the trained observational skills of an anthropologist to make such judgments: Laypersons of all nationalities readily attribute psychological characteristics to their own group and others (Peabody, 1985). Contemporary personality psychologists have occasionally attempted to characterize nations in terms of mean trait levels (Lynn & Martin, 1995).

However, these characterizations can be problematic on ethical, conceptual, and empirical grounds. Ethically, the attribution of psychological characteristics to ethnic or racial groups has been used as a rationale for some of the ugliest events in history, and as Pinker (2002) detailed in *The Blank Slate,* the possible misuse of findings on group differences has led many social scientists to deny categorically the existence of real psychological differences among groups. However, Pinker argued cogently that

> the problem is not with the possibility that people might differ from one another, which is a factual question that could turn out one way or the other. The problem is with the line of reasoning that says that if people do turn out to be different, then discrimination, oppression, or genocide would be OK after all. (Pinker, 2002, p. 141)

Provided that they reject this faulty reasoning, psychologists can ethically study possible cultural differences in personality. They should do so responsibly, which means carefully qualifying their conclusions and reminding readers that a range of individual differences can always be found within each culture

Robert R. McCrae and Antonio Terracciano, Gerontology Research Center, National Institute on Aging, National Institutes of Health, Department of Health and Human Services; the 79 contributing members of the Personality Profiles of Cultures Project.

The 79 contributing members of the Personality Profiles of Cultures Project are listed alphabetically by country in the Appendix.

Robert R. McCrae receives royalties from the Revised NEO Personality Inventory. German, Russian, and Czech data were taken from earlier studies (McCrae et al., 2004; Ostendorf & Angleitner, 2004), and portions of the Thai, Brazilian, and Lebanese data have also been reported in chapters in Costa and McCrae (in press), McCrae (in press), and McCrae, Terracciano, and Khoury (in press). Portions of these data were presented at the Second World Congress on Women's Mental Health, Washington, DC, March 2004. This research was supported in part by the Intramural Research Program of the National Institutes of Health, National Institute on Aging. Czech participation was supported by Grant 406/01/1507 from the Grant Agency of the Czech Republic and is related to Research Plan AV AV0Z0250504 of the Institute of Psychology, Academy of Sciences of the Czech Republic. S. Gülgöz's participation was supported by the Turkish Academy of Sciences. Burkinabé 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 awarded to Khairul A. Mastor.

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 Butkovič, Ole Dreyer, Susy Ball, Anna Gramberg, Honathan Harrow, V. S. Bose, Suguna Kannan, K. Sarita, K. Madhavi, Lidwina R. Dominica, 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, Sigrun Birna Sigurdardottir, Veronika Najzrova, 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, Cornelia Hohenbichler, Fatemeh Bayat, and Mahmoud Heydari.

Correspondence concerning this article should be addressed to Robert R. McCrae, Box #03, Gerontology Research Center, 5600 Nathan Shock Drive, Baltimore, MD 21224-6825. E-mail: mccraej@grc.nia.nih.gov

(McCrae, 2004). However, with suitable caution, it might be argued that research on this topic is ethically necessary because accurate assessments of cultural differences in personality—if any—are needed to help psychologists become "aware of and respect cultural, individual, and role differences," as required by their ethical principles (American Psychological Association, 2002, p. 1063).

The conceptual problems in characterizing the personality of a culture stem from the fact that cultures occupy a different level of analysis than persons, and it cannot be assumed that the same constructs are applicable to both. For example, it is known that anxiety, hostility, and depression covary among individuals to define a Neuroticism factor (N; Watson & Clark, 1984), but are anxious cultures also usually hostile and depressed cultures? If not, the concept of N would not be applicable to cultures. Hofstede (2001) has referred to the assumption that individual-level constructs are necessarily applicable to cultures as the *reverse ecological fallacy*. More profoundly, social scientists have long debated whether any aspect of psychology is relevant to an understanding of social groups or whether groups must be understood entirely in their own terms (Kroeber, 1917).

Empirically, the status of concepts such as national character is mixed. For example, later anthropologists have contested the accuracy of Benedict's (1934) description of the Pueblo (see Barnouw, 1985). National stereotypes are surely subject to ethnocentric and xenophobic biases, although Peabody (1985) argued that such biases have probably been exaggerated. Characterizations of cultures based on mean trait ratings have shown convergence across instruments in some comparisons (McCrae, 2002) but not in others (Poortinga, van de Vijver, & van Hemert, 2002). Church and Katigbak (2002) found agreement between American and Filipino judges on typical Filipino traits, but these judgments did not match observed mean profiles. The Personality Profiles of Cultures Project was designed to help resolve these issues by gathering new data on aggregated personality traits and perceptions of national character and relating them to features of culture.

Conceptualizing Personality in Cultures

There are at least three ways in which the personality of a culture might be conceptualized, which we call *ethos, national character,* and *aggregate personality*. Ethos, at a superorganic level (Kroeber, 1917), refers to traitlike characteristics used to describe the institutions and customs of the culture, such as its folktales, political organization, child rearing practices, and religious beliefs. Afghanistan under the Taliban might have been characterized as *closed to experience* because music was banned and Islamic orthodoxy was rigidly enforced. This personality-as-ethos does not imply anything directly about the personality traits of members of the culture: Afghans under Taliban rule might have been—some doubtless were—highly open to experience. Dimensions of ethos are sometimes inferred from the values of individual culture members (Hofstede, 2001; Inglehart, 1997), but they might be abstracted directly from features of culture, such as economic systems or health statistics (cf. Georgas & Berry, 1995).

National character refers to personality traits that are perceived to be prototypical of members of a culture. If this is to be a useful scientific construct, it must be shown that the characteristics are more descriptive than evaluative (Peabody, 1985) and that they are shared by knowledgeable judges both within and outside the culture (Church & Katigbak, 2002). Although national character is in some sense related to the traits of culture members, it does not necessarily represent a modal personality (Du Bois, 1944). Americans, for example, might think that the prototypical Texan has the personality characteristics of a cowboy, although there are relatively few cowboys still living in Texas and other Texans may not share their traits.

Aggregate personality, the focus of interest in the present article, characterizes cultures in terms of the assessed mean personality trait levels of culture members. Thus, "Norway is an extraverted culture" means, in this sense, that the average level of Extraversion (E) is high in Norway compared with other cultures. In this formulation, the whole culture is represented by the mean of its parts—the culture members—just as the wealth of a nation's citizens is reflected in per capita income.

For psychologists at least, aggregate personality is the most conveniently assessed of these three culture-level personality profiles. Standard measures of personality traits can be administered to a representative sample from each culture to be compared, and mean profiles can be computed. In one sense, this is precisely like comparing other groups, such as patients with different personality disorders (Morey et al., 2002). Yet methodologists have long noted that cross-cultural comparisons pose special challenges (McCrae, 2001; van de Vijver & Leung, 1997). They require, first, that it be demonstrated that the same constructs exist in each culture; next, that measuring instruments maintain construct validity in all cultures to be compared; and finally, that scales show scalar equivalence—that is, that a raw score has the same absolute interpretation in each culture. If these requirements can be met, then comparisons of representative samples from different cultures should yield meaningful results.

Bottom-Up and Top-Down Approaches

The present research used a measure of the five-factor model of personality (FFM; Digman, 1990), and there is by now considerable evidence that FFM dimensions are in fact universally replicable (McCrae & Allik, 2002; Paunonen & Ashton, 1998) and that instruments such as the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) retain their validity in translation. The remaining, and most challenging, requirement for cross-cultural comparisons is some demonstration that the scales have scalar equivalence and thus can be quantitatively compared. Note that scalar equivalence is not an all-or-nothing property: Like construct validity, it is always a matter of degree, and like construct validity, it is best assessed by the convergence of multiple lines of evidence. There are two basic approaches to this problem, which might be called *bottom-up* and *top-down.*

The bottom-up approach uses individual-level analyses (in which the person is the unit of analysis) to show that psychometric properties have been retained in transferring a scale across cultures. Item-response theory (IRT) has been used to determine if the items in a scale operate equivalently across cultures (e.g., Huang, Church, & Katigbak, 1997). One problem with the IRT approach is that it focuses on individual items, whereas the constructs of interest are measured by scales that typically aggregate across a number of items. It is possible that none of the items in a translated scale is strictly equivalent to its counterpart in the original version,

but that the differences introduced are random in nature and cancel out, leaving comparable total scores. Analyses of differential test functioning (Raju, van der Linden, & Fleer, 1995) can address this possibility. A second problem with IRT analyses is that samples from two cultures might have identical distributions of item scores and thus no differential item or test functioning, but the scores from one sample might in fact be systematically inflated by self-presentation bias; failure to find differential item functioning thus does not necessarily imply comparability of scores.

A second bottom-up approach relies on testing bilinguals who can complete the instrument in two different languages. At least six studies (Gülgöz, 2002; Konstabel, 1999; McCrae, 2001) have compared different translations of the NEO-PI-R using this design. They have all shown strong correlations between versions, indicating preservation of the basic constructs, and small and scattered mean level differences. To the extent that these studies are generalizable, it appears that translation in itself does not have a major impact on the interpretation of raw scale scores.

Still, translation is only one of several possible sources of inequivalence, and bilingual retest studies do not address others. Members of different cultures may differ in response styles such as acquiescence, in standards of comparison, and in norms of self-presentation. All of these biases might affect their responses regardless of the language in which they take a test.

Cross-cultural methodologists have focused on these bottom-up approaches because most cross-cultural studies have been based on comparisons of two or a very few cultures; in these circumstances, mean differences might be due to almost anything, and the comparability of scores should be ascertained before comparisons are made. With the recent availability of data from large numbers of cultures, however, a completely different, top-down approach is now possible, one that obviates some of the limitations of bottom-up approaches. In the top-down approach, researchers use culture-level analyses (in which the culture is the unit of analysis) to validate aggregate scores across cultures. If differences between cultures in mean trait levels were merely a matter of response biases and random error introduced by translations, then the aggregate scores should be meaningless. However, if a pattern of construct validity can be established for aggregate culture-level scores, then the scores themselves must be meaningful, and comparison across cultures would be appropriate.

Construct validation of culture-level scores parallels construct validation of individual scores where reproducibility or reliability, factor structure replicability, and convergent and discriminant validity are typically assessed. Multimethod studies are particularly valuable because they minimize the possibility that results may reflect shared biases. Culture-level scores are reproducible if the same score means are obtained from different samples of respondents; they are generalizable if these groups represent different sections of the culture, such as men and women or adolescents and adults (McCrae, 2001). Culture-level scores show factorial validity if a factor analysis of aggregate variables yields meaningful factors (which might or might not parallel the factors found in individuals). Hofstede (2001) called this *ecological factor analysis* and used it to identify dimensions of culture. Finally, evidence of convergent and discriminant validity can be obtained by correlating aggregate scores with other culture-level variables. These might be alternative operationalizations of the same constructs (as when McCrae, 2001, correlated mean NEO-PI-R N scores with the mean Eysenck Personality Questionnaire [EPQ] N scores tabulated by Lynn & Martin, 1995, across a sample of 14 cultures) or other culture-level criteria, such as per capita gross domestic product (GDP) or national health statistics.

Interpreting Ecological (Culture-Level) Factor Analyses

One step in this process requires special attention. Although most cross-cultural researchers understand that factor structures found at the individual level may or may not be replicated when aggregate data are analyzed, ecological factor analysis is an unusual and somewhat mysterious procedure. Some readers are surprised when an individual factor structure is replicated in an ecological analysis (e.g., McCrae, 2002), but in fact, that is the statistically expectable result. When two variables covary, groups that happen for any reason to be high on one tend also to be high on the other; when group-level data are analyzed, these two variables still covary. Departures from this expectation are most informative because they suggest that the groups—in this case, cultures—contribute something not found on the individual level. This culture-level addition may be random or systematic.

Random influences might be substantive because of the idiosyncratic effects of each particular culture on each trait. For example, Mexican *simpatia* (a norm dictating an avoidance of interpersonal conflict; see Diaz-Loving & Draguns, 1999) might elevate levels of A4: Compliance in that culture without affecting other facets of Agreeableness (A), such as A1: Trust or A2: Straightforwardness. Random influences might also be artifactual: error contributed by translation, varying response styles, or cultural variations in the meaningfulness of individual items. These are precisely the features that threaten scalar equivalence, and if there are marked departures from scalar equivalence, ecological factor analysis might show a sharply degraded version of the individual-level structure.

However, cultural influences might also be systematic, superorganic contributions to personality traits that change the factor structure at the culture level. For example, individualistic cultures might configure traits somewhat differently than collectivistic cultures do.

As a basis for interpreting the ecological factor analyses reported here, we conducted simulations of these conditions and evaluated the resulting factor congruences with the normative individual-level structure. A first simulation randomly reassigned targets to "cultures" to assess whether such groupings in fact retained the individual-level structure. A second simulation added random values to the means of these cultures to assess the impact of cultural idiosyncrasy or scalar inequivalence on ecological factor structure. A final simulation modeled systematic variation between cultures by contrasting hypothetical thinking and feeling cultures.

Aggregate Personality Profiles in 51 Cultures

The present study built on previous findings of meaningful differences in aggregate personality profiles using the self-report version of the NEO-PI-R. McCrae (2001, 2002) reported secondary analyses of data collected by other researchers from 36 cultures

(or subcultures). He found that (a) mean scores for the five NEO-PI-R domains were generalizable across age and sex groups; (b) culture-level factor analysis replicated the individual-level factor structure, though with a broader E factor; (c) scale variances were related to geography, being consistently largest in European and American cultures; and (d) aggregate scores showed convergent and discriminant correlations with other culture-level measures of personality and with Hofstede's (2001) dimensions of culture (see also Hofstede & McCrae, 2004). All of these findings argue for the meaningfulness of aggregate personality scores. However, these scores did not match the intuitive assessments of a panel of expert cross-cultural judges (McCrae, 2001): Japan, for example, showed a low score for Conscientiousness (C) despite the widespread perception that the Japanese are an industrious people. Poortinga et al. (2002) concluded, in a review of cross-cultural differences in personality, that "the validity of such claims [of real differences in mean levels] has to remain tentative" (p. 298) and encouraged research on alternative explanations for apparent group differences, such as response biases like acquiescence.

The present study was designed to replicate and extend evidence on the validity of aggregate personality scores as indicators of the personality profiles of cultures. To minimize the possibility that replications were due to shared response biases, we used an alternative method of measurement—observer ratings—to assess personality. College students from 51 cultures (including African, Arab, and Latin American cultures underrepresented in earlier studies) provided ratings on a male or female adult or college-age acquaintance who was a native-born citizen of their country. Although the resulting samples are unlikely to be strictly representative of any culture's population as a whole, they do appear to be comparable across cultures.

Analyses at the individual level (McCrae, Terracciano, & 78 Members of the Personality Profiles of Cultures Project, 2005) showed that the basic structure of personality traits was universally replicable and that age and sex differences seen in self-report studies (Costa, Terracciano, & McCrae, 2001; McCrae et al., 1999) were also generally found in observer-rating data. However, there was also systematic variation in the quality of the data collected, with more reliable and valid results obtained in Western and Westernized cultures, whose members are more familiar with personality questionnaires.

McCrae (2002), who first noted cultural differences in trait variances, speculated that they might reflect the operation of acquiescent response biases on balanced scales, random error introduced by translations, or substantive differences in homogeneity of personality traits in different cultures, but he was unable to test these hypotheses with available data. In the present study, an aggregate measure of acquiescence was included, along with a measure of data quality, to examine associations of these artifacts with variations in scale variances.

We also assessed the generalizability of aggregate personality scores across men and women and college-age and adult subsamples and the interrater reliability of the aggregate scores, examined the culture-level factor structure of the NEO-PI-R, and correlated aggregate scores with a variety of culture-level criteria, including self-report personality scores, Hofstede's (2001) dimensions of culture, and Schwartz's (1994) cultural value orientations. Previous research was limited to comparisons on the factor level, but the availability of culture-level facet scores (McCrae, 2002) made it possible to examine the culture-level convergence for specific traits in the present study. To characterize cultures as a whole, we analyzed personality profiles for the five factors and 30 facets of the NEO-PI-R. These profile analyses are informative about the validity of scores in individual cultures. We also considered the effects of national wealth, aggregate acquiescence, and within-culture sampling on these cross-cultural comparisons.

If aggregate trait scores show a pattern of convergent and discriminant validity, one can legitimately turn to the substantive interpretation of scores. Yet, even with evidence of rough scalar equivalence, it would be unwise to place much confidence in the characterization of an entire culture on the basis of a single sample of convenience. However, in a reanalysis of aggregate self-report data from McCrae (2002), Allik and McCrae (2004) showed that there are geographic patterns that can be used to characterize broad regions based on correspondingly larger samples. Specifically, a multidimensional scaling (MDS) plot based on the profile of NEO-PI-R facet scales showed that European cultures differ systematically from Asian and African cultures, chiefly with respect to E and Openness to Experience (O) scores, on which Europeans score higher. Southern European cultures tend to score higher on N than Northern European cultures. In the present study, we attempted to replicate these patterns as a basis for a substantive interpretation of personality scores.

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. Data gathered are from 51 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 multiple sites. German, Russian, and Czech data were taken from existing observer-rating data (McCrae et al., 2004; Ostendorf & Angleitner, 2004).

Individual-level analyses for 50 of these cultures have been reported in McCrae, Terracciano, et al. (2005); in this article, we add data from 34 Canadians, including 22 adult male participants. In addition, data became available from the Islamic Republic of Iran (*N*s = 35 male, 38 female raters; 137 targets, all adults). Domain reliabilities in the Iranian sample were .92, .88, .84, .93, and .95 for N, E, O, A, and C, respectively. After targeted rotation, factor congruence coefficients comparing the Iranian structure with the American normative structure (Costa & McCrae, 1992) were .93, .93, .72, .93, and .95, with a total congruence coefficient of .90.

Participants, Targets, and Procedures

Except where existing data were used, participants were college students who volunteered to participate anonymously in a study of personality across cultures. More detail on the raters has been given in McCrae, Terracciano, et al. (2005). The great majority were native-born citizens of their country, and the samples generally reflected the ethnic make-up of their countries.

Raters were randomly assigned to one of four target conditions[1] asking for ratings of college-age women, college-age men, adult (over 40 years old) men, or adult women. For the college-age targets, raters were asked to

> 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-age targets. The original study design called for 50 targets in each category; obtained subsamples ranged from 22 to 305, with a total of $N = 12{,}156$ valid ratings.

Instrument

The NEO-PI-R is a 240-item measure of the FFM. It contains 30 eight-item facet scales, 6 for each of the five basic personality factors, N, E, O, A, and C. Responses are made on a five-point Likert-type scale, from *strongly disagree* to *strongly agree.* The factors can be estimated by domain scores, which sum the relevant six facets, or more precisely by factor scores, which are a weighted combination of all 30 facets (Costa & McCrae, 1992, 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 *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Professional Manual* (hereafter, the *Manual*; Costa & McCrae, 1992).

The mean level of acquiescence varies across cultures (Smith, 2004), so some measure would be useful as a control variable. Because NEO-PI-R scales are roughly balanced, a general index of acquiescent response bias was calculated by summing raw (unreflected) responses to the 240 NEO-PI-R items (McCrae, Herbst, & Costa, 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. Independent back-translations were reviewed by the test authors, and modifications were made as needed. 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 and revised.

Invalid protocols were screened out using the rules specified in the *Manual* for missing data and random responding. In addition, the quality of data in each sample as a whole was assessed by an index based on proportion of valid protocols, yea- and nay-saying, proportion of missing data, first language of the respondent, publication status of the translation, and a judgment by the test administrator regarding miscellaneous problems. This quality index was internally consistent ($\alpha = .76$) and correlated across samples with domain reliabilities (rank-order rs = .63–.82) and overall factor replicability (rank-order $r = .60$; McCrae, Terracciano, et al., 2005).

The quality index was based on ranking within the group of 50 cultures. To estimate quality in the Iranian sample and to provide a transportable index of quality for use in future studies, we used a multiple regression to predict the total quality index from its components in the original 50 cultures. Four predictors were significant: the percentage of the unscreened sample with valid protocols (VALID), the judgment that respondents had problems with the questionnaire (PROBLEM; 0 = *no,* 1 = *yes*), the percentage of the unscreened sample that exceeded the cutoffs for acquiescence or nay-saying (ACQUIES) specified in the *Manual* (Costa & McCrae, 1992), and the estimated fluency of the sample in the language in which the NEO-PI-R was administered (FLUENCY; 2 = *native,* 1 = *very fluent nonnative,* 0 = *somewhat fluent nonnative language*). The regression equation estimated quality index scores as

$$-33.08 + .61*\text{VALID} - 9.15*\text{PROBLEM} - .91*\text{ACQUIES} + 2.83*\text{FLUENCY},$$

with an R^2 of .85. Quality index scores ranged from 5.5 to 37.9 in the original 50 cultures, with scores above 25 generally associated with excellent psychometric properties. Estimated data quality for Iran was low (10.2) because of frequent invalid and acquiescent protocols and comments by several respondents that the task was too long or confusing. Nevertheless, psychometric properties were adequate in the screened Iranian sample.

Culture-Level Correlates

To validate aggregate personality scores, we correlated them with other culture-level variables. Most directly relevant were national means on personality scales from previous self-report studies, including the NEO-PI-R (McCrae, 2002; Rossier, Dahourou, & McCrae, 2005); the EPQ (Eysenck & Eysenck, 1975), as reported in two overlapping meta-analyses by Lynn and Martin (1995) and van Hemert, van de Vijver, Poortinga, and Georgas (2002); and the Locus of Control scale (Rotter, 1966; Smith, Trompenaars, & Dugan, 1995). In previous research (McCrae, 2001, 2002), EPQ data from India were omitted as outliers; in the present study, we substituted Indian data from Lodhi, Deo, and Belhekar (2002) in the EPQ analyses.

Several sets of dimensions have been proposed to reflect national levels of values and beliefs. Hofstede (2001) provided scores for five dimensions: power distance (acceptance of status differences), uncertainty avoidance (preference for rules and routines to reduce stress), individualism (emphasis of self over family or group), masculinity (egoistic vs. social work goals), and, for a subset of countries, long-term orientation (orientation toward future rewards). Schwartz (1994) assessed seven cultural value orientations—conservatism, affective autonomy, intellectual autonomy, hierarchy, mastery, egalitarian commitment, and harmony—in samples of teachers. Inglehart and Norris (2003) reported scores on two dimensions derived from responses to the World Values Survey: traditional versus secular–rational values and survival versus self-expression values. Leung and Bond (2004) reported scores for social axioms (general beliefs about the social world), including social cynicism, social complexity, reward for application, religiosity, and fate control. Smith, Dugan, and Trompenaars (1996) reported scores for attitudes of organizational employees: conservatism versus egalitarian commitment and loyal involvement versus utilitarian involvement. Finally, Diener, Diener, and Diener (1995) tabulated subjective well-being values for nations.

Three economic indicators for each country were obtained from Internet sources: per capita GDP (http://www.bartleby.com/151/fields/64.html), the Gini Index (a measure of the equitable distribution of wealth; www.bartleby.com/151/fields/68.html), and the Human Development Index (HDI; http://hdr.undp.org/reports/global/2002/en/indicator/indicator.cfm?File=indic_290_1_1.html).

Some judgment was required in matching cultures across these studies because cultures have been defined differently in different studies and national boundaries have changed in recent years. In general, the most specific matches available were used (e.g., Telugu-speaking Indians with Telugu-speaking Indians). Separate data for Northern Ireland were provided in some studies (Diener et al., 1995; Inglehart & Norris, 2003); otherwise, Northern Ireland was matched with the United Kingdom or Great Britain. Germany was matched with West Germany. Data from Czechoslovakia were paired with both the Czech Republic and Slovakia; data from Yugoslavia were paired with Croatia, Slovenia, and Serbia, except that McCrae's (2002) Yugoslavians were in fact Serbians and were

[1] In Uganda and France, raters described four targets varying in age and sex; in Iran, raters described two adult targets.

matched only to Serbia. Data from the Soviet Union were matched to Russia but not to Estonia. German and French Switzerland were distinguished where possible. For Schwartz's (1994) values, rural and urban Estonian samples were averaged. Burkina Faso and Nigeria were matched with Hofstede's (2001) West African region; Ethiopia, Uganda, and Botswana with East Africa; and Kuwait and Lebanon with Arab countries.

Replications With Self-Report Data

Previous studies (e.g., Leung & Bond, 2004; McCrae, 2002; Steel & Ones, 2002) have reported correlations between aggregate-level NEO-PI-R self-report data and other culture-level variables. For the present study, we recalculated these correlations using all available cultures and the matching rules noted above to assess replicability of culture-level associations across methods. Note that these are very conservative tests of replicability because the samples of cultures, although overlapping, are not the same in the two sets of analyses.[2]

Results

Generalizability, Reliability, and Standardization

Group-level analyses began with means from the four separate subsamples: college-age men, college-age women, adult men, and adult women.[3] To assess generalizability of culture-level scores across age groups, we correlated the mean raw domain scores for college-age subsamples with mean domain scores for adult subsamples matched on culture and sex (e.g., the college-age male subsample from Peru was paired with the adult male subsample from Peru). Correlations for N, E, O, A, and C were .66, .45, .51, .62, and .36, respectively (all ps $< .001$), suggesting that culture-level scores generalize at least minimally across these age groups. To assess generalizability across sex, we correlated mean raw domain scores for female subsamples with domain scores for male subsamples matched on culture and age group (e.g., the college-age male subsample from Peru was paired with the college-age female subsample from Peru). Correlations for N, E, O, A, and C were .55, .78, .75, .64, and .84, respectively (all ps $< .001$), suggesting generalizability across sexes.

All these generalizability coefficients underestimate the reliability of the aggregate scores; they are in essence uncorrected split-half correlations. A more accurate estimate of the reliability of the aggregate scores is given by the intraclass correlation, ICC(1, k), calculated as

$$\frac{\text{(between-cultures mean square} - \text{within-culture mean square)}}{\text{between-cultures mean square}}$$

derived from a one-way analysis of variance (see Shrout & Fleiss, 1979). Intraclass correlations usually apply to ratings given by a set of judges of the same target. Here, the targets were different individuals, but all were representatives of the same culture. ICC(1, k)s were .88, .91, .92, .91, and .89 for N, E, O, A, and C, respectively. As shown in the eighth column of Table 1, ICC(1, k)s for the 30 facets ranged from .81 to .97, with a median of .91. These very high values are understandable given that each of the 51 data points is based on an average of 238 ratings.

Age and sex differences at the aggregate level were examined by paired t tests on subsample means across all 202 subsamples. Older subsamples scored lower on N, E, and O and higher on A and C than younger subsamples (all ps $< .001$); female groups scored higher than male groups on all five factors (all ps $< .01$). To adjust for these differences, we standardized the 30 NEO-PI-R facet scores as T scores within the four age and sex groups across all 51 cultures, and all subsequent analyses used these facet T scores.[4] Factor scores were created using scoring weights given in the *Manual* (Costa & McCrae, 1992, Table 2, bottom panel), which is reasonable because the American structure was replicated in all the individual cultures (McCrae, Terracciano, et al., 2005).

Ecological Factor Analysis Simulations

As a baseline for interpreting the effects of cultural influences on ecological factor analyses, we conducted a set of simulations. First, the 12,156 cases were randomly reassigned to 202 "cultures" to parallel the 202 subsamples. This reassignment eliminated any real, systematic effects due to culture. A culture-level principal components analysis was then conducted on the means of the 30 facet scales in these randomly constituted cultures, and five factors were extracted and compared with the American normative factor structure (McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). The resulting structure was a near-perfect replication of the individual-level structure, with factor congruence coefficients ranging from .93 to .96 after varimax rotation. Thus, as expected from statistical reasoning, ecological factor analyses replicate individual-level results if there are no distinct cultural effects.

To simulate the effect of random cultural contributions to the factor structure, we created 30 random variables with an expected mean of 0 and standard deviation of 4 T-score points. These perturbations were added to the facet scores of the 202 cultures; the mean absolute change in facet scores was 3.2 T-score points. (Note that this is equivalent to adding constants to the scores of all members of a culture and would have no effect on the factor structure at the individual level within any culture.) These relatively modest—although pervasive—random changes had a pronounced effect on the factor structure: After varimax rotation, factor congruence coefficients for optimally matched factors ranged from .07 for E to .66 for C; even after targeted rotation (McCrae et al., 1996), congruence coefficients ranged only from .47 to .77. A second random simulation used the same random additions but divided by two and thus representing a mean absolute change of only 1.6 T-score points per facet. In this analysis, varimax factor congruence coefficients were .65, .80, .69, .81, and .87 for N, E, O, A, and C, respectively; they ranged from .69 to .90 after targeted rotation. It thus appears that even small deviations from scalar equivalence can degrade the factor structure if they are found for many facets.

Finally, to simulate the effect of systematic cultural contributions to ecological factor structures, we divided the 202 cultures

[2] The self-report correlations are available from Robert R. McCrae.

[3] There were no Iranian data for college-age targets, so the total number of subsamples was 202.

[4] Previous research had used U.S. age and sex norms to standardize data. However, there are no published college-age norms for Form R of the NEO-PI-R, and the use of U.S. norms might be considered ethnocentric. In preliminary analyses, data in the present study were also standardized using the U.S. data collected for this study, with very similar results. The international norms used in the present study are available from Robert R. McCrae.

Table 1
Culture-Level Factor Structure of NEO-PI-R Facet Scales After Targeted Rotation, Intraclass Reliability of Aggregates, and Cross-Instrument Correlations

	Procrustes-rotated principal component							
NEO-PI-R facet scale	N	E	O	A	C	VC[a]	ICC(1,*k*)	*r*[b]
N1: Anxiety	**.77**	.08	−.14	.07	.17	.93[d]	.90	.69***
N2: Angry Hostility	**.65**	−.09	−.17	**−.42**	−.10	.97[e]	.86	.39*
N3: Depression	**.52**	−.21	−.25	.17	**−.43**	.83	.89	.53**
N4: Self-Consciousness	.31	**−.40**	−.21	.35	−.16	.69	.91	.61***
N5: Impulsiveness	**.52**	**.50**	.18	−.19	−.27	.96[e]	.87	.63***
N6: Vulnerability	**.61**	**−.40**	−.16	−.05	−.37	.94[e]	.88	.58***
E1: Warmth	−.02	**.68**	.20	**.44**	.20	.99[e]	.94	.46**
E2: Gregariousness	−.37	**.64**	−.12	.17	−.18	.92[d]	.88	.35*
E3: Assertiveness	**−.50**	.32	−.03	−.29	.31	.90[d]	.81	.36*
E4: Activity	.04	**.45**	.28	.09	.35	.85	.88	.64***
E5: Excitement Seeking	−.22	**.41**	−.23	−.25	**−.48**	.64	.96	.48**
E6: Positive Emotions	−.26	**.72**	.19	.24	.12	.95[e]	.91	.52**
O1: Fantasy	.18	**.55**	**.59**	.03	−.18	.87[d]	.92	.58***
O2: Aesthetics	−.11	−.26	**.68**	.23	.18	.87[d]	.91	.48**
O3: Feelings	.03	**.50**	**.58**	.27	.21	.84	.94	.70***
O4: Actions	−.20	−.16	**.71**	.07	−.23	.83	.90	.44**
O5: Ideas	−.36	−.02	**.61**	.11	.21	.91[d]	.85	.63***
O6: Values	.13	**.52**	**.55**	.20	.07	.62	.97	.75***
A1: Trust	−.27	**.41**	.21	**.59**	.05	.96[e]	.91	.40*
A2: Straightforwardness	.16	**.45**	.08	**.56**	.22	.62	.94	.27
A3: Altruism	.09	**.63**	.19	**.43**	**.43**	.92[d]	.96	.72***
A4: Compliance	−.35	−.24	.13	**.73**	−.08	.94[e]	.88	.32
A5: Modesty	.35	.34	−.08	**.59**	.05	.76	.92	.63***
A6: Tender-Mindedness	.10	.16	.13	**.69**	.25	.93[d]	.93	.60***
C1: Competence	−.24	**.45**	.25	.13	**.70**	.92[d]	.94	.66***
C2: Order	−.17	−.28	−.17	.23	**.66**	.84	.84	.68***
C3: Dutifulness	.02	.21	.09	.35	**.85**	.92[d]	.94	−.02
C4: Achievement Striving	−.24	.11	−.07	.05	**.74**	.91[d]	.91	.69***
C5: Self-Discipline	−.17	.17	.05	.13	**.83**	.96[e]	.85	.09
C6: Deliberation	−.37	**−.44**	−.18	.24	**.55**	.96[e]	.85	.70***
Factor congruence[c]	.86[e]	.81[e]	.88[e]	.90[e]	.94[e]	.87[e]		
Factor comparability	.83	.94	.94	.90	.95			

Note. These are principal components from 202 subsamples targeted to the American normative factor structure. Loadings greater than .40 in absolute magnitude are given in boldface. NEO-PI-R = Revised NEO Personality Inventory; N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness; VC = variable congruence coefficient; ICC = intraclass correlation.
[a] Total congruence coefficient in the last row. [b] Correlations with aggregate self-report NEO-PI-R facet scores (McCrae, 2002; J. Rossier, personal communication, August 19, 2004), $N = 28$. [c] Congruence with American normative factor structure. [d] Congruence higher than that of 95% of rotations from random data. [e] Congruence higher than that of 99% of rotations from random data.
* $p < .05$, one-tailed. ** $p < .01$, one-tailed. *** $p < .001$, one-tailed.

into two groups. The first was hypothesized to consist of cultures that emphasized thinking over feeling; in these, 5 *T*-score points were added to O5: Ideas, and 5 points were subtracted from O3: Feelings. In the second group, hypothesized to emphasize feeling over thinking, 5 *T*-score points were added to O3: Feelings, and 5 points were subtracted from O5: Ideas. Factor congruence coefficients after varimax rotation were .93, .86, .14, .96, and .97 for N, E, O, A, and C, respectively, and could be increased to .96, .88, .61, .96, and .96 by targeted rotation. In that optimized rotation, five of the O facets had positive loadings on the O factor, whereas O3: Feelings loaded –.66. Systematic cultural contributions of this magnitude are thus clearly noticeable in ecological factor analyses.[5]

Ecological Factor Analysis

A culture-level principal components analysis was conducted on the means of the 30 facet scales in the 202 subsamples. Previous work at both the individual and cultural levels had suggested that five factors should be extracted; however, the first seven eigenvalues in the present analysis were 8.21, 4.23, 2.99, 2.39, 1.77, 1.55, and 0.98, and parallel analysis (Cota, Longman, Stewart, Holden, & Fekken, 1993) indicated that six factors should be

[5] All simulations were repeated twice with different initial randomizations and very similar results.

retained. Both five- and six-factor solutions were therefore examined.

The six-factor solution was evaluated by calculating comparability coefficients with the American normative self-report structure (Costa & McCrae, 1992)—that is, by correlating factor scores generated in this analysis with group means for the factor scores calculated at the individual level using scoring coefficients given in the *Manual*. Factors resembling E, O, A, and C could be roughly identified (factor comparabilities = .71–.96); the two remaining factors were related chiefly to N (comparabilities = .80 and .45). The first N factor had its largest loadings on N3: Depression, N4: Self-Consciousness, and N6: Vulnerability; the second was chiefly defined by N2: Angry Hostility and N5: Impulsiveness, as well as (low) A4: Compliance. The two aspects of N reflected in these factors call to mind Achenbach, McConaughy, and Howell's (1987) distinction between internalizing and externalizing disorders. However, a reanalysis of self-report data from McCrae (2002) extracting six factors (although only five were warranted by parallel analysis) found a single N factor, E and A factors, and three factors defined by O and C facets. Thus, the six-factor solution was not replicable across methods of measurement.

In a varimax rotation of five factors, only O and C were clearly replicated; N was divided into two factors as in the six-factor solution, and E and A were fused. However, in large part, the differences from the normative structure appear to be a matter of rotation: Table 1 reports the five-factor solution rotated to maximum similarity to the American normative self-report structure (McCrae et al., 1996). Although factor congruence was beyond chance for all five factors (see Table 1), only N, O, A, and C factors clearly replicated the American structure using Haven and ten Berge's (1977) criterion of congruence over .85. The remaining factor was defined by five of the six E facets and by O3: Feelings and A3: Altruism, which have secondary loadings on the E factor in individual-level analyses. Yet it also had large loadings for other facets that are not definers of the E factor in individual-level analyses, including N5: Impulsiveness, O1: Fantasy, and C1: Competence.

The same phenomenon was reported by McCrae (2002) in an analysis of aggregate self-report data from 36 cultures. The factor congruence coefficients between that culture-level structure and the structure in Table 1 were .81, .91, .87, .80, and .88 for N, E, O, A, and C, respectively, suggesting similar culture-level structures, especially for E. Finally, an analysis was conducted for 98 subsamples from cultures not included in McCrae's study; results closely resembled those in Table 1, with factor congruences with the normative self-report structure of .94, .76, .86, .86, and .93 for N, E, O, A, and C, respectively, after targeted rotation. The anomalies with the E factor thus replicated using a different method of personality assessment in a completely distinct sample of cultures. This appears to be a real culture-level contribution to the covariation of aggregate personality scores, which McCrae noted was related to cultural differences in individualism/collectivism—that is, individualism/collectivism was correlated chiefly with the facets that define the broad E factor. We return to this point in the Discussion.

Nevertheless, the overall structure clearly resembles the FFM. As simulations showed, this would not be the case if scalar inequivalences were widespread or large. Further evidence is provided by factor comparabilities, which relate factor scores in the same sample calculated with two different sets of scoring weights (from American normative self-reports and the present analysis). These values, reported in the last row of Table 1, were all high and argue that all five factors can be interpreted in terms of the familiar FFM.

Culture Means and Standard Deviations

To characterize each culture, we calculated overall mean factor and facet *T* scores based on the combined international norms. Columns 2 through 6 of Table 2 report the factor means for the 51 cultures. Inspection of the table shows that there is a fairly narrow range of values (7.5, 11.2, 12.3, 8.1, and 8.0 *T*-score points for N, E, O, A, and C, respectively). These ranges are consistently smaller than those seen in self-reports (10.8, 16.0, 15.1, 11.8, and 13.1 *T*-score points for N, E, O, A, and C, respectively; McCrae, 2002), suggesting that cultural differences in rated personality are smaller than differences in self-reported personality. This relative restriction of range may reduce correlations with other culture-level variables.

An examination of the table shows that Brazilians were rated highest in N, Northern Irish in E, Czechs in A, and German Swiss in both O and C. A more detailed description of substantive findings is postponed until evidence of the validity of these scores can be evaluated.

We also examined scale variability. For each of the 30 facets, standard deviations for college-age subsamples were compared with adult subsamples matched on culture and sex; correlations ranged from .19 to .73, of which 29 were significant ($p < .05$). Similar analyses showed generalizability across sex (rs = .31–.75, all ps $< .01$). As in analyses of self-report data (McCrae, 2002), scale variability appeared also to be generalizable across content domains: Cultures with smaller standard deviations on one facet tended to have smaller standard deviations on all the others. A factor analysis of standard deviations for the 30 facets across the 202 subsamples showed a single large factor accounting for 47% of the variance, with all facets loading .39 or higher. Each culture's characteristic variability was therefore computed as the mean standard deviation across all 30 facet scales.

Mean standard deviations for each culture are reported in column 7 of Table 2. When the data are sorted by that variable, as in McCrae (2002), the geographical organization of results is clear: Fifteen of the 19 Asian and African cultures are found among the 20 cultures with lowest variability. The values in column 7 are significantly correlated ($r = .68$, $N = 26$, $p < .001$) with mean standard deviations in self-reports (McCrae, 2002).

There are several possible substantive explanations for these findings. Nisbett, Peng, Choi, and Norenzayan (2001) argued that East Asians use dialectical reasoning, which might lead them to endorse indicators of both high and low poles of a trait, thus yielding average-level total scores and reducing variability of the scale. Lee and Ottati (1993) suggested that differences in perceived group heterogeneity may be veridical; members of collectivistic cultures (including Asians and Africans) may vary less in personality traits, perhaps because their behavior is shaped primarily by interpersonal relationships (Markus & Kitayama, 1991). Au (2000) discussed several possible reasons for intracultural variation, although his data for values and attitudes showed a different geographical pattern than that seen here for traits.

Table 2
Aggregate Factor T *Scores, Mean Facet Standard Deviation, Quality Index, Acquiescence, and Profile Agreement in 51 Cultures*

Culture	Factor					*SD*[a]	Quality index	Acquiescence score	ICC[b]
	N	E	O	A	C				
Americans (919)	48.1	52.2	50.4	49.1	48.8	9.9	25.7	504.6	.67***
Argentines (204)	51.3	52.3	46.1	50.6	50.0	10.3	22.8	497.1	
Australians (206)	48.6	53.8	50.7	50.0	47.5	10.1	27.3	522.2	
Austrians (158)	48.3	50.7	50.5	50.6	52.4	10.8	29.1	512.8	.04
Batswana (186)	48.9	46.7	47.7	48.0	46.8	8.8	13.5	514.8	
Belgians (247)	49.6	52.2	50.4	49.8	47.4	9.7	33.3	502.7	.74***
Brazilians (597)	53.7	52.1	49.0	50.3	51.5	10.3	26.3	517.7	
Burkinabé (207)	53.1	48.8	49.3	51.3	49.7	8.4	21.6	534.7	.88***
Canadians (167)	49.5	52.5	48.4	49.9	49.6	9.7	27.9	519.5	.42**
Chileans (194)	50.0	51.7	51.8	50.8	52.2	10.7	33.4	496.9	
Croatians (191)	49.3	50.9	49.1	48.4	50.3	10.1	17.7	514.0	.44**
Czechs (400)	51.4	48.1	50.4	54.2	51.5	11.1	31.0	536.1	.45**
Danes (153)	50.3	51.8	55.2	53.1	48.4	10.8	35.9	499.0	.31*
English (194)	50.1	53.7	53.5	50.2	48.1	10.3	28.8	512.5	
Estonians (298)	47.9	52.1	46.8	47.8	50.0	11.4	30.7	506.6	.43**
Ethiopians (197)	48.8	47.0	48.5	47.3	47.2	7.9	10.9	522.5	
Filipinos (197)	48.3	48.9	50.8	47.4	53.5	9.4	18.3	504.2	.45**
French (274)	52.7	48.0	51.4	51.3	48.4	10.3	35.6	496.9	.66***
French Swiss (265)	53.6	51.0	51.6	53.0	49.7	10.7	37.0	501.8	.79***
German Swiss (214)	47.5	48.3	58.4	54.0	53.5	10.0	34.7	496.0	.31*
Germans (593)	48.1	49.6	54.9	52.1	52.3	9.1	37.9	516.8	.25
Hong Kong Chinese (207)	50.5	46.2	47.3	46.9	49.6	9.2	26.3	522.9	.29
Icelanders (199)	48.6	51.5	51.2	52.0	49.3	9.7	29.8	508.8	
Indians (185)	50.1	48.5	48.8	51.7	52.3	8.3	16.1	555.1	.72***
Indonesians (196)	50.0	45.4	48.9	49.0	49.6	9.3	22.8	515.3	.34*
Iranians (137)	48.4	48.2	50.1	48.6	47.0	11.1	10.2	528.9	
Italians (195)	52.6	46.5	52.3	48.1	48.3	9.9	25.8	488.6	.44**
Japanese (191)	50.7	49.4	51.2	48.8	49.5	9.9	26.9	486.8	.49**
Kuwaitis (468)	51.9	52.9	47.6	51.0	52.6	8.7	19.3	542.2	
Lebanese (200)	50.0	51.2	48.1	46.4	50.5	9.2	10.0	519.4	
Malays (289)	51.8	48.3	47.5	51.7	53.0	7.9	13.5	521.3	.75***
Maltese (202)	53.1	50.5	48.5	49.4	51.6	10.6	31.6	518.8	
Mexicans (173)	46.2	47.8	50.2	47.5	50.7	9.5	18.9	493.9	
Moroccans (171)	50.5	44.8	48.5	46.1	45.5	7.6	5.5	516.7	
New Zealanders (200)	47.9	52.4	50.1	50.1	47.8	10.4	33.3	514.1	
Nigerians (184)	47.8	44.4	49.1	46.6	45.8	8.0	13.2	507.3	
Northern Irish (106)	50.1	55.6	47.5	52.4	47.4	10.4	30.5	516.4	
People's Republic Chinese (177)	46.5	46.6	50.1	48.6	48.0	8.9	16.2	517.4	.28
Peruvians (154)	48.5	50.1	48.9	48.5	48.7	8.7	15.8	501.2	.58***
Poles (197)	50.7	49.2	48.6	48.5	49.4	10.1	31.7	515.7	
Portuguese (198)	51.6	51.3	51.3	51.1	50.7	9.7	32.9	512.9	.51**
Puerto Ricans (160)	49.9	51.6	49.7	48.9	52.9	9.7	12.9	521.6	
Russians (320)	51.4	45.7	49.7	50.3	49.1	8.7	16.6	527.7	.51**
Serbians (200)	49.3	49.3	51.6	48.4	51.7	10.6	31.6	528.3	.66***
Slovaks (198)	49.2	49.7	48.2	50.6	48.6	9.2	30.4	508.5	
Slovenians (209)	50.6	49.5	48.8	49.0	52.3	10.1	13.8	515.0	
South Koreans (196)	48.4	50.7	50.9	50.3	48.3	9.6	26.7	494.0	.06
Spaniards (200)	49.7	50.4	48.8	51.4	51.3	10.9	37.8	500.2	.30
Thais (209)	48.9	49.6	48.5	49.6	48.9	10.2	25.0	521.0	
Turks (208)	51.4	53.0	48.2	51.0	51.4	10.3	32.3	517.8	.60***
Ugandans (166)	49.4	46.5	49.5	48.3	48.2	8.3	6.0	518.7	

Note. Cultures are labeled by culture members' name as a reminder that these data are the aggregate of individuals. Except for Iran, the quality index is taken from McCrae, Terracciano, and 78 Members of the Personality Profiles of Cultures Project (2005); acquiescence is the mean sum of all Revised NEO Personality Inventory (NEO-PI-R) raw-score responses before reflecting; *n*s for each culture are given in parentheses. N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness; ICC = intraclass correlation.

[a] Mean *T*-score-standardized standard deviation across 30 NEO-PI-R facet scales. [b] Intraclass correlation with standardized aggregate self-report NEO-PI-R facet scores ($N = 28$; McCrae, 2002; J. Rossier, personal communication, August 4, 2004).

* $p < .05$, one-tailed. ** $p < .01$, one-tailed. *** $p < .001$, one-tailed.

However, the standard deviations in Table 2 are also correlated with acquiescence ($r = -.28$, $N = 51$, $p < .05$) and especially with the quality index ($r = .66$, $N = 51$, $p < .001$). Acquiescence—itself a culturally salient response style (Smith, 2004)—reduces variance when applied to a balanced scale, as does random error. These correlations suggest that apparent differences in facet scale variance across cultures may have been due largely to artifacts of response style.

Within-Nation Variability

In five cases, data were available from two or more sites in the same nation. Data for French and German Swiss are given in Table 2; these two samples differed significantly for all factors except A. English and Northern Irish targets did not differ in N, E, A, or C, but they were dramatically different in O: The English ranked 4th, whereas the Northern Irish ranked 49th. Hong Kong Chinese scored significantly higher than Mainland Chinese on N and lower on O. Where there are linguistic or historical reasons for treating subcultures separately, such treatment appears to be appropriate.

Three sites were sampled in Brazil and four in the United States. There were no significant differences among the Brazilian sites for any of the factors. The American sites, however, differed on N, E, and C, and some of these differences were substantial. In E, for example, the lowest scoring site (San Francisco State University) fell exactly in the middle of the distribution in Table 2, whereas the highest scoring site (University of Iowa) was slightly higher than any of the 51 cultures. Had we relied on data from a single American site, we might have reached a range of conclusions about Americans' level of E.

Ideally, intercultural comparisons would be based on national probability samples; failing that, it would be useful to compare the samples obtained here with larger and perhaps more representative samples in one or more cultures to get a sense of how representative they are. There are no NEO-PI-R norms based on national probability samples, and although the *Manual* (Costa & McCrae, 1992) provides adult Form R norms, it has no college-age norms. However, a sample was recently recruited to provide new normative data (McCrae, Martin, & Costa, in press), and we selected Form R targets aged 18–21 years and 30+ years ($N = 722$) from that sample to provide a basis of comparison. The present American sample scored about one quarter of a standard deviation higher than these new norms on E and O and one quarter of a standard deviation lower on A, but it did not differ significantly on N or C. Thus, it appears that the combined American sample was probably close to being representative of the country as a whole.

Culture-Level Correlates

To examine the validity of aggregate personality scores, we correlated them with culture-level scores from other personality instruments, measures of beliefs and values, and socioeconomic indicators (see Table 3). The most direct comparison was with the factors in self-reports on the NEO-PI-R. Significant, and moderately large, correlations were found for N, E, and O factors, and a trend ($p < .10$) was found for C. None of the discriminant correlations was significant.

Aggregate mean values for the 30 NEO-PI-R facets were reported by McCrae (2002) for self-report data from 36 cultures, of which 26 overlapped with the present sample, and by J. Rossier (personal communication, August 19, 2004) for Burkina Faso and French Switzerland. Culture-level correlations for the facets are given in the last column of Table 1; most (87%) were significant, and the median value was .58. Note that four of the A facets and four of the C facets were significant despite limited agreement on global A and C factor scores. These cross-method data provide strong evidence that a variety of specific traits may be validly assessed at the culture level, just as cross-observer correlations have provided consensual validation of traits at the individual level (McCrae, 1982).

With regard to the EPQ scales, in addition to the links between corresponding N and E scales, it might be hypothesized that A and C would be negatively related to Psychoticism and positively related to Lie (McCrae & Costa, 1985), although these associations were small even in comparisons at the individual level. A significant correlation was found for N using data from Lynn and Martin (1995), but none of the other hypotheses was supported. Thus, this cross-method, cross-instrument comparison provided little evidence of validity for the culture-level scores. Similarly, there was no association with external locus of control, which at the individual level was modestly related to N and low C (Costa, McCrae, & Dye, 1991).

Aggregate personality factor scores were, however, significantly related to a number of culture-level variables that characterize societies' beliefs and values. N was related to uncertainty avoidance, a dimension associated with anxiety (Hofstede, 2001). Cultures whose members were high in E had democratic values, as seen in correlations with Smith et al.'s (1996) Egalitarian Commitment scale and low power distance. E was also related to individualism, an emphasis on self-expression rather than survival, a disbelief in the role of fate, and high subjective well-being. These are generally Western beliefs and values, consistent with research showing that E is highest in Europe and the Americas (McCrae, 2004).

Cultures whose members were high in O were also characterized by low power distance and high individualism. In addition, open cultures valued affective and intellectual autonomy and egalitarian commitment but rejected conservatism; they had a secular–rational approach to life. Open cultures thus appeared to be independent and unconventional. A, another dimension associated with values at the individual level (Roccas, Sagiv, Schwartz, & Knafo, 2002), had a similar set of correlates, except that A was not significantly related to Smith et al.'s (1996) egalitarian commitment. C was unrelated to values and beliefs when zero-order correlations were examined.

The pattern of correlates in Table 3 is meaningful and reasonably consistent with previous findings. As table footnotes show, 17 of the 34 significant correlations (50%) between observer-rated NEO-PI-R factors and other criteria were replicated when aggregated self-report data were used to measure the factors.

The similarity of correlates for O and A suggests that these two theoretically independent factors may have been correlated in this sample. In fact, the correlation was .42 ($p < .01$). A was also positively related to the other three factors (rs = .29–.39, $p < .05$). (None of the other interfactor correlations was significant.) To improve discriminant validity, we orthogonalized the five factors by factoring them and extracting five varimax-rotated factors. Factor scores from this analysis clarified correlations with the

Table 3
Culture-Level Correlates of NEO-PI-R Form R Factors

Criterion	Factor				
	N	E	O	A	C
Personality measures					
NEO-PI-R Form S factors ($N = 28$)					
Neuroticism	.52[c]**	−.23	−.02	.22	.02
Extraversion	−.06	.60[c]***	.30	.37	.01
Openness to Experience	−.17	.09	.50[c]**	.13	.25
Agreeableness	.30	−.06	−.33	.11	.16
Conscientiousness	−.12	−.10	−.30	−.19	.35[c]
EPQ scales ($N = 28$; Lynn & Martin, 1995)[a]					
Neuroticism	.41[b,c]*	.12	.17	.14	.08
Extraversion	−.15	.05	.02	−.13	−.08
Psychoticism	−.05	−.23	−.15	.05	.22
EPQ scales ($N = 27$; van Hemert, van de Vijver, Poortinga, & Georgas, 2002)[a]					
Neuroticism	.19	.12	−.05	−.01	−.05
Extraversion	−.32	.34	.08	−.26	−.04
Psychoticism	−.15	−.25	−.07	−.26	.08
Lie ($N = 25$)	.06	−.55**	−.18	−.51**	.06
Rotter locus of control ($N = 34$; Smith, Trompenaars, & Dugan, 1995)	.25	−.06	−.14	−.07	.02
Beliefs, attitudes, values					
Hofstede (2001) dimensions ($N = 49$)					
Power distance	.20[c]	−.46[b]***	−.41[b]**	−.31*	.11
Uncertainty avoidance	.30[b,c]*	.07	−.03	−.02	.20
Individualism	.05	.51[b,c]**	.33[b]*	.37**	−.14
Masculinity	−.14	.00	.10	.04	.03
Long-term orientation ($N = 30$)	−.09	−.17	−.05	−.18	−.01
Schwartz (1994) values ($N = 22$)					
Conservatism	−.20	−.02	−.70[b,c]**	−.51*	.15
Affective autonomy	.13	.24	.55**	.61[c]**	−.03
Intellectual autonomy	.37	−.15[c]	.51[b]*	.44*	.07
Hierarchy	−.24	−.12	−.32	−.23	−.10
Mastery	−.27	−.31	.10	−.09	−.15
Egalitarian commitment	.25	.20	.55**	.44*	−.09
Harmony	.08	.09	.26	.09	.15
Inglehart & Norris (2003) values ($N = 42$)					
Secular–rational	−.01	.09	.34*	.42**	−.02
Self-expression	−.08	.54[b,c]***	.29	.30	−.08
Social axioms ($N = 29$; Leung & Bond, 2004)					
Social cynicism	−.25	−.24	.03	−.18	.00
Social complexity	−.09	.35	.25	.19	.23
Reward for application	−.30	−.34	−.36	−.23	.21
Religiosity	.05	−.39*	−.36	−.13	.28
Fate control	−.26	−.56[b,c]**	−.19	−.10	−.03
Organizational attitudes ($N = 34$; Smith, Dugan, & Trompenaars, 1996)					
Conservatism vs. egalitarian commitment	−.02	.46[b]**	.34*	.26	−.21
Loyal involvement vs. utilitarian involvement	−.01	.00	−.17	−.31	.03
Subjective well-being ($N = 35$; Diener, Diener, & Diener, 1995)	.01	.63[b,c]***	.35*	.48**	−.02
Economic indicators					
Gross domestic product per capita ($N = 51$)	.04	.44[b]***	.47[b]***	.46***	.02
Gini Index ($N = 40$)	−.02	−.08	−.25	−.26	.11
Human Development Index ($N = 48$)	.02	.54[b,c]***	.34[b]*	.40**	.25[c]

Note. N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness; NEO-PI-R = Revised NEO Personality Inventory; EPQ = Eysenck Personality Questionnaire.
[a] Indian data from Lodhi, Deo, and Belhekar (2002). [b] Replicated, $p < .05$, in culture-level analyses using self-report data. [c] Significant after partialing per capita gross domestic product.
* $p < .05$. ** $p < .01$. *** $p < .001$.

criteria in Table 3. The pattern of significant findings was unchanged for N. Correlations between E and van Hemert et al.'s (2002) E and between C and self-reported NEO-PI-R C attained statistical significance, and O now showed significant negative correlations with Leung and Bond's (2004) reward for application and religiosity axioms. The major changes, however, were with A, which was now significantly related only to van Hemert et al.'s Lie, affective autonomy, secular–rational values, and per capita GDP. Of these, the most easily interpreted is affective autonomy, which implies regard for pleasure and the enjoyment of life—values that antagonistic cultures might disdain. The overall pattern of results with the orthogonalized factors more closely resembled that found in self-report data: Nineteen of 27 significant correlations (70%) were replicated using aggregated self-reports.

Control Analyses

It might have been guessed that GDP would be related to C because C is associated with job performance at the individual level (Barrick, Mount, & Judge, 2001). Instead, aggregate E, O, and A were all related to GDP and to HDI (see Table 3). One interpretation of these findings is that individuals in wealthy nations have moved beyond materialist values and are now more concerned with interpersonal and experiential issues (Inglehart & Oyserman, in press). Alternatively, the associations may simply reflect the fact that Europeans and Americans are both wealthy and high in E and O (Allik & McCrae, 2004).

Because some researchers believe that culture-level correlations should be interpreted net of economic indicators (e.g., Hofstede, 2001; Leung & Bond, 2004), table notes in Table 3 report results of analyses controlling for GDP; only about a third of the significant correlations remained significant. The most pronounced effects of partialing GDP were on the associations of personality with values. By contrast, correlations with NEO-PI-R self-report aggregates were relatively unaffected; indeed, the partial correlation for C was now significant at conventional levels ($r = .43$, $p < .05$). Analyses for facets (see Table 1) controlling for GDP found that all 26 significant correlations remained significant and that A4: Compliance attained significance ($r = .32$, $p < .05$).

NEO-PI-R scales are roughly balanced in keying, but N, E, A, and C domains have a small preponderance of positively keyed items, and all five factors were correlated with acquiescent responding within the 51 cultures (median rs = .25, .22, .15, .03, and .30 for N, E, O, A, and C, respectively). When aggregated across respondents, these small correlations might have affected culture-level means. In fact, however, culture-level acquiescence (see Table 2) was significantly related only to O ($r = –.35$, $p < .05$), and partialing acquiescence out of the correlations reported in Table 3 had little effect. Correlations of O with intellectual autonomy, secular–rational values, Smith et al.'s (1996) egalitarian commitment, subjective well-being, and the HDI became nonsignificant; the remaining 32 significant correlations in Table 3 changed little in magnitude and remained significant. Partialing acquiescence from the correlations between Form S and Form R facets (see Table 1) reduced the correlation for E3: Assertiveness to $r = .27$, *ns.* All other correlations remained significant.

Profile Analyses

It is conceivable that the correlations seen in the last column of Table 1 and in the first five rows of Table 3 are attributable to a subset of cultures—perhaps individualistic societies, in which traits are thought to be more salient (Triandis, 1995). In that case, the data would in fact offer construct validity only within those cultures. Personality profiles provide one way of assessing agreement across methods at the level of each individual culture. McCrae (1993) proposed a coefficient of profile agreement, r_{pa}, that summarizes agreement between two assessments of a target across the five factors. This coefficient was calculated for each of the 28 cultures for which both self-report and observer-rating NEO-PI-R data were available; values ranged from .32 to .42, with a median of .37. This was comparable to the mean r_{pa} of .41 found at the individual level for agreement between self-reports and peer ratings from knowledgeable acquaintances (McCrae, 1993). Most importantly, r_{pa} was unrelated to Hofstede's (2001) individualism (or to acquiescence or the quality index). Agreement across methods thus appears to be the rule for both individualistic and collectivistic cultures.

That interpretation may, however, be misleading because r_{pa} was developed for the analysis of individual-level scores, which have much higher variance than the mean scores analyzed here, and r_{pa} is sensitive to score elevation. Most mean scores from both self-reports and observer ratings were near $T = 50$, so agreement across methods was expectable. A somewhat different approach to profile agreement is given by intraclass correlations calculated by the double-entry method across the 30 facets. This approach reflects similarity in the shape of the profile and the relative elevation of scores, and it has been used to quantify agreement with personality disorder prototypes (Miller, Pilkonis, & Morse, 2004). Aggregate facet data for self-reports were available for 28 cultures that overlapped the present sample. After first being standardized across cultures, intraclass correlations ranged from .04 for Austria to .88 for Burkina Faso. As shown in the last column of Table 2, 22 of these correlations (79%) were significant; the median value was .45. Cultures with the largest profile agreement (rs > .60) were Belgium, Burkina Faso, France, India, Malaysia, Serbia, Turkey, French Switzerland, and the United States—a very diverse group.

Data from Italy and Malaysia illustrate profile agreement in Figure 1 and give some sense of the degree to which profiles from the two studies are—and are not—interchangeable. (Note that this figure plots the unstandardized T scores.) The aggregate self-reports (dashed lines) are more extreme than the aggregate observer ratings (solid lines), but they tend to show similar profile shapes. As is the case with multimethod assessments of individuals (McCrae, 1994), self-reports and ratings appear to have given related but not wholly redundant characterizations.

Multidimensional Scaling

As in Allik and McCrae (2004), we examined geographical patterns in trait scores using MDS. Mean scores for the 30 NEO-PI-R facets were first standardized across the 51 cultures, and the distance between cultures was defined as 1 minus the Pearson correlation

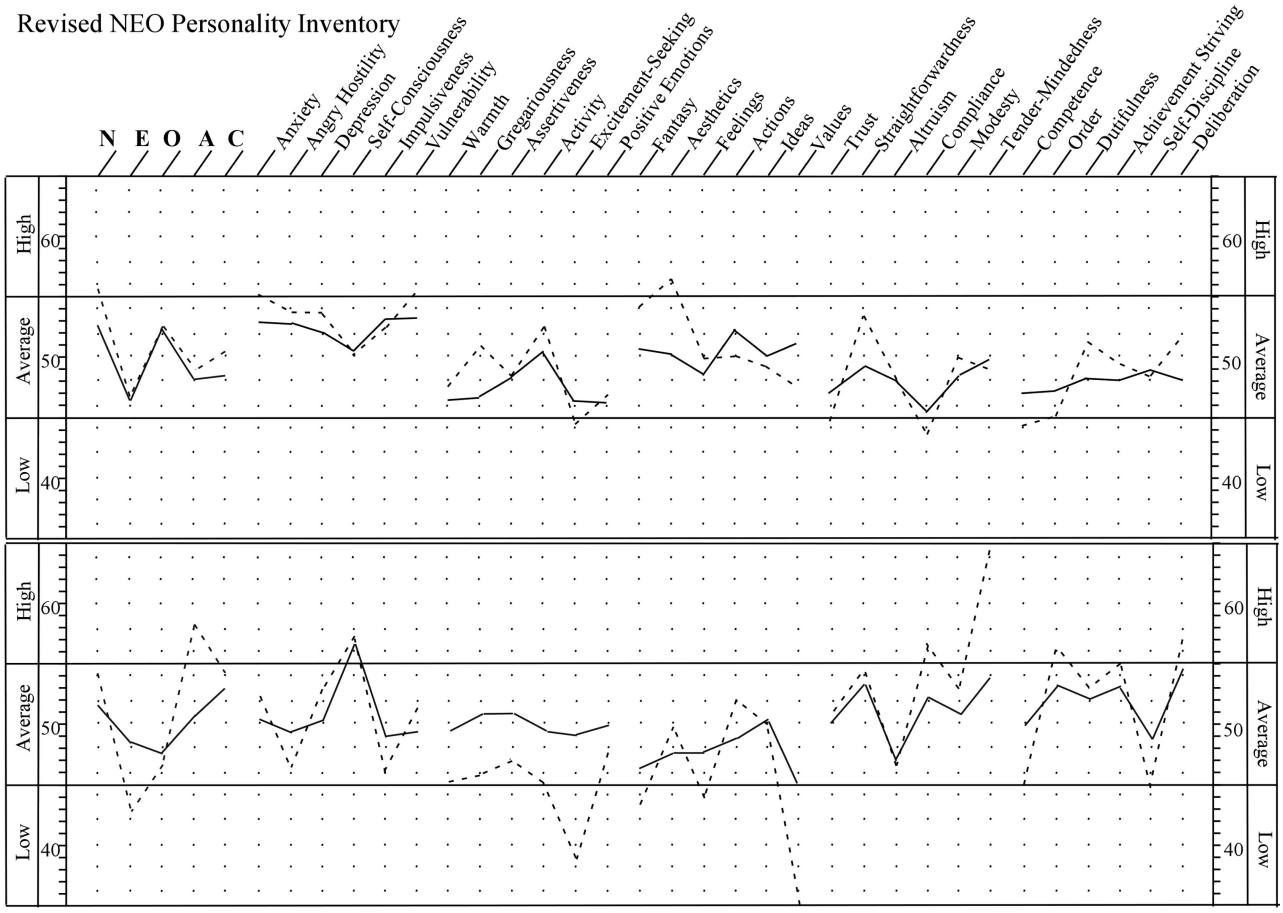

Figure 1. Mean Revised NEO Personality Inventory profiles for Italians (top panel) and Malays (bottom panel) from self-reports (dashed lines) and observer ratings (solid lines). The five factor scores are given on the left; toward the right, the facet scales are grouped by factor. N = Neuroticism; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness. Profile form reproduced by special permission of the Publisher, Psychological Assessment Resources, Inc., 16204 North Florida Avenue, Lutz, Florida 33549, from the Revised NEO Personality Inventory by Paul T. Costa, Jr., and Robert R. McCrae. Copyright, 1978, 1985, 1989, 1992 by PAR, Inc. Further reproduction is prohibited without permission from PAR, Inc.

across the 30 facets.[6] Two dimensions were extracted, and coordinates were correlated with the mean factor scores given in Table 2. As in Allik and McCrae, we rotated the MDS axes to maximize the correlations with N and E, placing cultures higher in N at the top (north) of the plot and those higher in E at the right (east) of the plot.

Figure 2 presents the resulting MDS plot and merits several comments. Cultures near each other in MDS space, and thus similar in personality profiles, tend to be historically and ethnically related. On the far right, the Northern Irish, English, Australians, New Zealanders, Canadians, and Americans are clustered. On the left, East and West African cultures are grouped. Czechs and Slovaks are located at the top of the plot; Germans, Austrians, and German Swiss at the bottom. Lebanese, Turks, and Kuwaitis occupy the center of the plot. The top of the plot includes Southern and Eastern European cultures as well as Brazilians and Argentines; however, Spaniards and most other Latin Americans are found in the bottom half of the plot. More broadly, it can be seen that, except for Russians, all of the cultures on the left side of the plot are Asian or African; all of the cultures on the right are European or American.

In general, these findings resemble those reported by Allik and McCrae (2004). To quantify similarity, we correlated MDS coordinates across the two studies for the 26 cultures common to both. The horizontal coordinates were strongly related ($r = .69$, $p < .001$), but the vertical coordinates were not ($r = .31$, *ns*). In part, this appears to be due to the shift of the three German-speaking cultures from the top of the self-report plot to the bottom of the observer-rating plot. It is not clear why German-speaking people would perceive themselves as higher in N than they perceive their compatriots.

[6] A distance matrix based on (1 – Pearson *r*) is offered by Statistica (SoftStat, Inc., 1995), the program used for MDS analyses. Because the Pearson *r* is insensitive to differences in elevation of the profile, we also calculated a distance matrix based on 1 minus the intraclass correlation, using the double-entry method. Results were virtually identical.

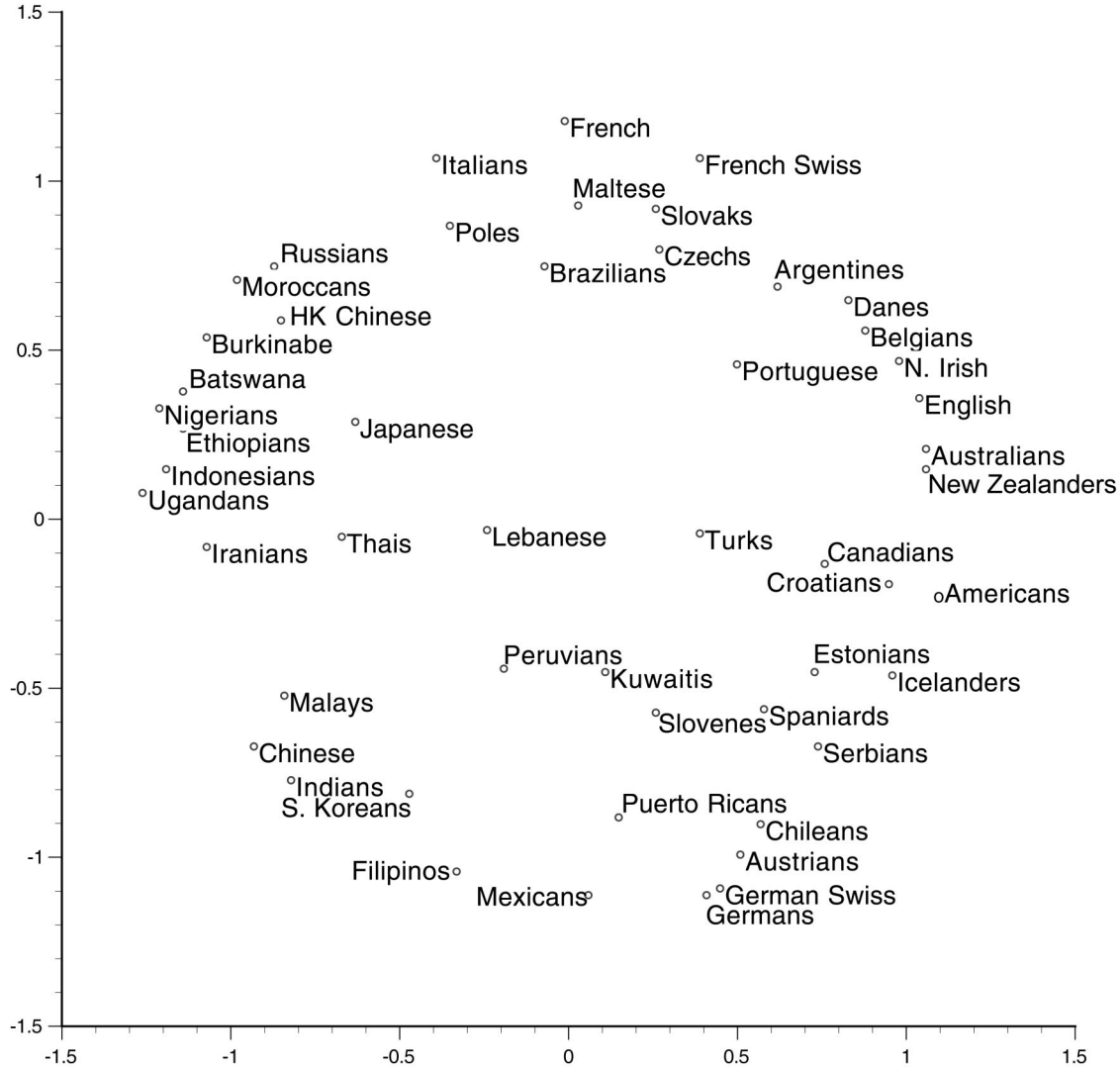

Figure 2. Multidimensional scaling plot of 51 cultures based on a distance matrix of (1 – Pearson *r*) for the 30 Revised NEO Personality Inventory facet scores, standardized across cultures. The vertical axis is maximally aligned with N (Neuroticism), the horizontal axis with E (Extraversion). HK Chinese = Hong Kong Chinese; N. Irish = Northern Irish; S. Koreans = South Koreans.

In one sense, these correlations underestimated the similarity of the two plots because they were based solely on 26 matching cultures. Side-by-side comparison of the plots reveals that the five Black African cultures in Figure 2 have positions close to those of Black South Africans and Zimbabweans in Allik and McCrae (2004). Corresponding positions are also found for Icelanders and Norwegians, English and White South Africans, and Puerto Ricans and Hispanic Americans. Similar profiles in similar cultures add to the construct validity of aggregate personality scores.

Substantively, Figure 2 can be interpreted most directly by correlating MDS coordinates with the factor scores in Table 2. These show that the vertical axis is significantly related to N (r = .63, $p < .001$) versus C (r = –.63, $p < .001$), whereas the horizontal axis is related to E (r = .81, $p < .001$) and O (r = .28, $p < .05$). At the facet level, cultures near the top of the plot are characterized chiefly ($|r|$s > .50) as anxious, hostile, depressed, and vulnerable, whereas those near the bottom are assertive, competent, achievement oriented, self-disciplined, and deliberate. Cultures on the right are impulsive, warm, active, cheerful, imaginative, liberal, trusting, competent, organized, and self-disciplined, whereas those on the left tend to be self-conscious and vulnerable. These are the facets that define the broad E factor in Table 1.

The factor-level correlations replicate those found in Allik and McCrae (2004), except that, in the earlier study, A was negatively related to both axes, whereas in the present study, A was positively related to the horizontal axis (r = .43, $p < .01$).[7] Together, these self-report and observer-rating data suggest that Europeans and Americans are higher in O and especially E than Asians and Africans. There is also some evidence that Southern and Eastern

[7] The orthogonalized A factor described earlier was unrelated to either axis.

Europeans are higher in N and lower in C than Northern Europeans and that people from South and Southeast Asia (e.g., Indians, Filipinos) are low in N and high in C.

Discussion

With a few exceptions, the present analyses replicated findings previously reported for aggregate personality traits measured by self-reports on the NEO-PI-R. Culture-level scores are generalizable across age groups and sex; the culture-level factor structure approximates that found at the individual level; scale variances differ systematically across cultures, with the largest variances found in Western cultures; aggregate scores show meaningful patterns of convergent and discriminant validity with other culture-level variables; and geographically and historically related cultures show similar personality profiles. Such results would be unlikely if personality measures were seriously distorted by cultural differences in language and response biases; the data as a whole thus offer top-down evidence of the rough scalar equivalence of NEO-PI-R factors and facets in some two dozen languages.

If scalar equivalence is maintained when the NEO-PI-R is used in different cultures and if samples are comparable—as the design of this study was intended to make them—then group differences are presumably real: For example, Figure 1, which was plotted against international norms, can be interpreted to mean that Malays are higher in self-consciousness than most other people in the world (Abdullah, 1993), and the analyses of within-nation variability showed that the English are more open to experience than the Northern Irish. (Recall that these statements refer to people on average; there is a wide range of individual differences on all traits in all cultures.) Poortinga and colleagues (2002) are probably not alone in remaining skeptical of such claims of true mean level differences, and researchers who wish to advance them must make systematic efforts to eliminate alternative explanations. Several steps were taken in that direction here.

First, the use of observer ratings eliminated the possibility that results reflected cultural differences in self-presentation. There may, of course, be cultural influences on how raters describe others, but it seems unlikely that they would exactly parallel the cultural effects on self-presentation. In fact, in cultures that promote modesty, self-enhancement should be diminished whereas other enhancement might be increased (but see Bond, Kwan, & Li, 2000, for evidence of separate self- and other enhancement effects). Such effects would tend to reduce culture-level correlations across methods. Second, analyses examining acquiescence showed that it has a very limited effect on the validity of aggregate personality variables, at least when balanced scales such as those of the NEO-PI-R are used. Third and last, we conducted analyses controlling for GDP. Those analyses showed that national wealth and the educational, social, and health variables that attend it may play a role in accounting for observed associations of personality traits with beliefs and attitudes. However, convergence across measures of traits themselves was largely unaffected by partialing out GDP.

This does not mean that we now have definitive values for aggregate trait levels in our sample of cultures. Several findings in the present study suggest caution. Perhaps most puzzling is the failure to find significant correlations with aggregate traits assessed by the EPQ. This is least surprising with respect to Psychoticism because it is only modestly related to low A and C at the individual level and shows a weak structure at the culture level (van Hemert et al., 2002). Weak correlations of N and E with the corresponding EPQ variables may be due to differences in sampling and data analysis. For example, van Hemert et al.'s (2002) meta-analysis did not adjust means for age and sex, and their samples were not based on a uniform sampling strategy. Again, it is possible that differences were due to the particular scales used: EPQ E is not identical in conception or operationalization to NEO-PI-R E. To examine that issue further, we correlated van Hemert et al.'s EPQ E with the six NEO-PI-R E facets and found significant associations for E2: Gregariousness, E3: Assertiveness, and E5: Excitement Seeking (rs = .39–.56, ps < .05). Unfortunately, none of these correlations was replicated when Lynn and Martin's (1995) EPQ E data were used. With Ns of only 27 and 28, it is perhaps not surprising that results were unstable.

Another reason for caution is the finding that different values were obtained from different sites in the American data. Although a similar problem was not found in Brazilian data, this result calls into some question the degree to which one can confidently generalize from data from a single site in any culture. To obtain personality profiles that accurately reflect the culture as a whole, researchers need to obtain more representative samples, and given the rather narrow range of differences between cultures, the samples probably need to be larger than 200. Future designs would also benefit from the inclusion of targets aged 21 to 40 years, a large segment of the population that was deliberately omitted here. An alternative strategy would be to look systematically for subcultures. Perhaps the personality profile of Californians is meaningfully different from that of Midwesterners, and averaging them might obscure important relations.

Despite these cautions, the pattern of evidence so far suggests that aggregating individual personality scores is a useful way to characterize cultures and thus that more rigorous sampling and more extensive data collection would likely be worth the effort. At present, one can be fairly confident about generalizations that characterize large regions of the world: In particular, the evidence that Europeans, on average, are more extraverted than Asians or Africans is quite strong.

The origin of these group similarities and differences is unclear. English and ethnic-majority Australians share a common ancestry, so their similar personality profiles might be due to similar distributions of trait-related alleles. Yet they also share many elements of culture, including language, law, and religion, which could also account for their resemblance. Ancestry and culture are almost completely confounded in the present data, and either or both might be causal. One powerful and underused method for distinguishing them is the study of acculturation effects (McCrae, Yik, Trapnell, Bond, & Paulhus, 1998), but a variety of methods is needed to fully understand personality psychology in the genome era (Anderson & Nickerson, 2005).

Regardless of their origins, aggregate personality traits may have consequences for societies (Hofstede & McCrae, 2004). For example, McCrae, Costa, et al. (2005) found that within cultures at the individual level, HIV stigmatization was negatively related to O, especially O6: Values. This effect seems to be multiplied at the aggregate level: Cultures with very low levels of O6 (see McCrae, 2002) include South Africa and Zimbabwe, where official reluctance to deal with HIV infection has led to devastating epidemics.

The full range of aggregate personality traits might be relevant to a host of social, economic, and health outcomes.

Culture-Level Factor Structure

The major finding from the ecological factor analysis is that a close approximation to the individual-level FFM could be found in these data. Simulations showed that this is not remarkable, but it is testimony to the scalar equivalence of NEO-PI-R scales in different cultures: If the mean level of facet scale scores were seriously distorted across cultures, the intended factor structure could not be replicated.[8] As discussed by Allik and McCrae (2002), the covariation of culture-level traits along the lines of the FFM might be due to (thus far unidentified) cultural mechanisms that affect all facets of a domain similarly. More likely, however, is that the common genetic influences thought to account for structure at the individual level (McCrae, Jang, Livesley, Riemann, & Angleitner, 2001) also operate at the aggregate personality level: The factors emerge because societies differ in the distribution of alleles of genes relevant to each of the factors.

There are, however, two other findings worth noting. The first is the apparent divisibility of observer-rated culture-level N into two factors, one resembling internalizing, the other externalizing disorders. This distinction was not found in the analysis of aggregate self-report data nor in analyses of individual-level data from either method of measurement, so it is not yet clear whether it is a reliable finding or a fluke. The distinction itself, however, is conceptually meaningful, and it is possible that there is a real Level of Analysis × Method of Measurement interaction. For aggregate observer ratings, anger and impulsiveness are different phenomena from depression and self-consciousness, whereas for aggregate self-reports, they are both expressions of negative affect. Why this difference should appear at culture-level analyses but not at individual-level analyses is not clear, but the question is perhaps worth pursuing.

The second is that in the five-factor solution, the E factor is exceptionally broad, including elements of N, O, and C that are not found at the individual level and that have no known genetic association. This appears to be a robust phenomenon, found in both self-report and observer-rating data and in two nonoverlapping samples of cultures. Particularly striking is the pattern of O facets: Cultures high in E were also high in O1: Fantasy and O6: Values but tended to be low in O2: Aesthetics. Introverted cultures (e.g., India; see McCrae, 2002, Figure 1) showed the opposite pattern. Inglehart (1997) reported that imagination and tolerance are among the defining values of the self-expression dimension, which is associated with E ($r = .54$) and with Hofstede's individualism/collectivism ($r = .63$) and which is believed to have increased in the postindustrial world. Thus, one account of the broad E factor would be that, for historical reasons, the extraverted peoples of Europe and the Americas have entered a postmaterialist era in which a number of other traits, including imagination, tolerance, impulsiveness, and sense of competence, are encouraged. If these interpretations are correct, the structure shown in Table 1 can be considered an overlay of historico-cultural influences on the basic structure given in the human genome.

Aggregate Personality, Ethos, and National Character

Do aggregate personality traits resemble the ethos of a culture? If Ruth Benedict had administered the NEO-PI-R to her Pueblo informants, would they have scored low on E and O and high on A and C, as the description *sober, conventional, cooperative,* and *orderly* suggests? There is at present only indirect evidence of such convergence. Hofstede's (2001) dimensions of culture have been related to institutions and customs—for example, high power distance cultures are said to be characterized by centralized political power, an emphasis on agriculture instead of industry, and unquestioning deference to teachers. In the present study, power distance was related to low E, O, and A, suggesting that cultures whose members are introverted, closed to experience, and disagreeable may be deferential, agrarian, and authoritarian. Hofstede and McCrae (2004) have discussed these links at length, including a consideration of the causal directions involved.

Ethos might also be reflected in shared values and beliefs, and the present study provides new information linking aggregate personality traits to culture-level measures provided by Schwartz (1994), Inglehart and Norris (2003), Smith and colleagues (1996), and Leung and Bond (2004). The most predictable associations were with O. Cultures marked by higher levels of O are progressive, humanistic, and free-thinking; those with lower levels of O are conservative, traditional, and religious in orientation. These culture-level associations resemble the individual-level associations (Roccas et al., 2002). A is also strongly associated with values at the individual level, and one might have predicted that cultures high in A would value harmony over mastery, whereas those low in A would be characterized by social cynicism. None of those predictions is confirmed in Table 3, however. Instead, cultures high in A tended to resemble those high in O. Orthogonalizing the factors eliminated many of the unanticipated correlates of A, but the hypothesized correlates were still nonsignificant. Clearly, Table 3 provides more evidence for the construct validity of aggregate O than of aggregate A.

Neither N nor C was strongly related to beliefs and values, but E was associated with an orientation toward self-expression, a repudiation of fatalism, and high subjective well-being. Inglehart and Oyserman (in press) suggested that self-expression arises as industrial societies come to take survival for granted and become postmaterialist in outlook. The strong link between self-expression and E and the fact that much of the world is rapidly becoming postindustrial suggests the hypothesis that E should increase in the coming decades—a conclusion consistent with cohort differences documented by Twenge (2001).

Do the data in Table 2 reflect perceptions of national character? Americans tend to think of East Asians as being prototypically hardworking, but in the present data, Japan and Hong Kong were merely average in C. Instead, the highest scoring countries were Kuwait, Puerto Rico, Malaysia, German-speaking Switzerland, and the Philippines. These might seem surprising, but most Americans are not very knowledgeable about Kuwaitis or Filipinos, so

[8] Some forms of scalar inequivalence are consistent with a replicable structure: In particular, if the effect of the distortion is uniform across all facets in a domain, it will not affect factor structure. Factor replication at the culture level is thus not proof of scalar equivalence, but it does provide one piece of evidence for it.

their perceptions here may not be trustworthy. Although it would be ideal to have information on the perception of each culture's character by itself and all other cultures, such data are not yet available. The Personality Profiles of Cultures Project provides data for most of the 51 cultures studied here that can be used to examine correspondences between aggregate personality and national character—as perceived by members of the culture itself—at both the factor and facet levels (Terracciano et al., in press).

References

Abdullah, M. M. (1993). Konsep malu dan segan orang Melayu [Self-consciousness and the shame concept in Malay people]. In A. H. Othman (Ed.), *Psikologi Melayu* (pp. 244–294). Kuala Lumpur, Malaysia: Dewan Bahasa dan Pustaka.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232.

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.

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *American Psychologist, 57,* 1060–1073.

Anderson, N. B., & Nickerson, K. J. (Eds.). (2005). Genes, race, and psychology in the genome era [Special issue]. *American Psychologist, 60*(1).

Au, K. Y. (2000). Intra-cultural variation as another construct of international management: A study based on secondary data of 42 countries. *Journal of International Management, 6,* 217–238.

Barnouw, V. (1985). *Culture and personality* (4th ed.) Belmont, CA: Wadsworth.

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9,* 9–30.

Benedict, R. (1934). *Patterns of culture.* Boston: Houghton Mifflin.

Bond, M. H., Kwan, V. S. Y., & Li, C. (2000). Decomposing a sense of superiority: The differential social impact of self-regard and regard-for-others. *Journal of Research in Personality, 34,* 537–553.

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

Costa, P. T., Jr., & McCrae, R. R. (1992). *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. (in press). Trait and factor theories. In J. C. Thomas & D. L. Segal (Eds.), *Comprehensive handbook of personality and psychopathology* (Vol. I). New York: Wiley.

Costa, P. T., Jr., McCrae, R. R., & Dye, D. A. (1991). Facet scales for Agreeableness and Conscientiousness: A revision of the NEO Personality Inventory. *Personality and Individual Differences, 12,* 887–898.

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.

Cota, A. A., Longman, R., Stewart, R., Holden, R. R., & Fekken, G. C. (1993). Comparing different methods for implementing parallel analysis: A practical index of accuracy. *Educational and Psychological Measurement, 53,* 865–876.

Diaz-Loving, R., & Draguns, J. G. (1999). Culture, meaning, and personality in Mexico and in the United States. In Y.-T. Lee, C. R. McCauley, & J. G. Draguns (Eds.), *Personality and person perception across cultures* (pp. 103–126). Mahwah, NJ: Erlbaum.

Diener, E., Diener, M., & Diener, C. (1995). Factors predicting the subjective well-being of nations. *Journal of Personality and Social Psychology, 69,* 851–864.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41,* 417–440.

Du Bois, C. (1944). *The people of Alor.* Minneapolis: University of Minnesota Press.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire.* San Diego, CA: EdITS.

Georgas, J., & Berry, J. W. (1995). An ecocultural taxonomy for cross-cultural psychology. *Cross-Cultural Research, 29,* 121–157.

Gülgöz, S. (2002). Five-factor model and the NEO-PI-R in Turkey. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 167–188). New York: Kluwer Academic/Plenum Publishers.

Haven, S., & ten Berge, J. M. F. (1977). *Tucker's coefficient of congruence as a measure of factorial invariance: An empirical study* (Heymans Bulletin No. 290 EX). Groningen, the Netherlands: University of Groningen.

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.

Huang, C. D., Church, A. T., & Katigbak, M. S. (1997). Identifying cultural differences in items and traits: Differential item functioning in the NEO Personality Inventory. *Journal of Cross-Cultural Psychology, 28,* 192–218.

Inglehart, R. (1997). *Modernization and postmodernization: Cultural, economic, and political change in 43 societies.* Princeton, NJ: Princeton University Press.

Inglehart, R., & Norris, P. (2003). *Rising tide: Gender equality and cultural change around the world.* New York: Cambridge University Press.

Inglehart, R., & Oyserman, D. (in press). Individualism, autonomy, self-expression and human development. In H. Vinken, J. Soeters, & P. Ester (Eds.), *Comparing cultures: Dimensions of culture in a comparative perspective.* Leiden, the Netherlands: Brill.

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

Kroeber, A. L. (1917). The superorganic. *American Anthropologist, 19,* 163–213.

Lee, Y. T., & Ottati, V. (1993). Determinants of in-group and out-group perceptions of heterogeneity: An investigation of Sino-American stereotypes. *Journal of Cross-Cultural Psychology, 24,* 298–318.

Leung, K., & Bond, M. H. (2004). Social axioms: A model for social beliefs in multi-cultural perspective. *Advances in Experimental Social Psychology, 36,* 119–197.

Lodhi, P. H., Deo, S., & Belhekar, V. M. (2002). The five-factor model of personality: Measurement and correlates in the Indian context. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 227–248). New York: Kluwer Academic/Plenum Publishers.

Lynn, R., & Martin, T. (1995). National differences for thirty-seven nations in extraversion, neuroticism, psychoticism and economic, demographic and other correlates. *Personality and Individual Differences, 19,* 403–406.

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

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

McCrae, R. R. (1993). Agreement of personality profiles across observers. *Multivariate Behavioral Research, 28,* 13–28.

McCrae, R. R. (1994). The counterpoint of personality assessment: Self-reports and observer ratings. *Assessment, 1,* 159–172.

McCrae, R. R. (2001). Trait psychology and culture: Exploring intercultural comparisons. *Journal of Personality, 69,* 819–846.

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. (2004). Human nature and culture: A trait perspective. *Journal of Research in Personality, 38,* 3–14.

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

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

McCrae, R. R., & Costa, P. T., Jr. (1985). Comparison of EPI and Psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences, 6,* 587–597.

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., Martin, T. A., Oryol, V. E., Senin, I. G., & O'Cleirigh, C. (2005). *Personality correlates of HIV stigmatization in Russia and the United States.* Manuscript submitted for publication.

McCrae, R. R., Herbst, J. H., & Costa, P. T., Jr. (2001). Effects of acquiescence on personality factor structures. In R. Riemann, F. Ostendorf, & F. Spinath (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp. 217–231). Berlin, Germany: Pabst Science.

McCrae, R. R., Jang, K. L., Livesley, W. J., Riemann, R., & Angleitner, A. (2001). Sources of structure: Genetic, environmental, and artifactual influences on the covariation of personality traits. *Journal of Personality, 69,* 511–535.

McCrae, R. R., Martin, T. A., & Costa, P. T., Jr. (in press). Age trends and age norms for the NEO Personality Inventory-3 in adolescents and adults. *Assessment.*

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., Terracciano, A., & 78 Members of the Personality Profiles of Cultures Project. (2005). 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., Yik, M. S. M., Trapnell, P. D., Bond, M. H., & Paulhus, D. L. (1998). Interpreting personality profiles across cultures: Bilingual, acculturation, and peer rating studies of Chinese undergraduates. *Journal of Personality and Social Psychology, 74,* 1041–1055.

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.

Miller, J. D., Pilkonis, P. A., & Morse, J. Q. (2004). Five-factor model prototypes for personality disorders: The utility of self-reports and observer ratings. *Assessment, 11,* 127–138.

Morey, L. C., Gunderson, J., Quigley, B. D., Shea, M. T., Skodol, A. E., McGlashan, T. H., et al. (2002). The representation of Borderline, Avoidant, Obsessive-Compulsive, and Schizotypal personality disorders by the five-factor model of personality. *Journal of Personality Disorders, 16,* 215–234.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108,* 291–310.

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.

Paunonen, S. V., & Ashton, M. C. (1998). The structured assessment of personality across cultures. *Journal of Cross-Cultural Psychology, 29,* 150–170.

Peabody, D. (1985). *National characteristics.* New York: Cambridge University Press.

Pinker, S. (2002). *The blank slate: The modern denial of human nature.* New York: Penguin Books.

Poortinga, Y. H., van de Vijver, F., & van Hemert, D. A. (2002). Cross-cultural equivalence of the Big Five: A tentative interpretation of the evidence. In R. R. McCrae & J. Allik (Eds.), *The five-factor model of personality across cultures* (pp. 273–294). New York: Kluwer Academic/Plenum Publishers.

Raju, N. S., van der Linden, W., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Applied Psychological Measurement, 19,* 353–368.

Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The Big Five personality factors and personal values. *Personality and Social Psychology Bulletin, 28,* 789–801.

Rossier, J., Dahourou, D., & McCrae, R. R. (2005). Structural and mean level analyses of the five-factor model and Locus of Control: Further evidence from Africa. *Journal of Cross-Cultural Psychology, 36,* 227–246.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80*(1).

Schwartz, S. H. (1994). Beyond individualism/collectivism: New cultural dimensions of values. In U. Kim, H. C. Triandis, C. Kagitcibasi, S.-C. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications* (pp. 85–119). Thousand Oaks, CA: Sage.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

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

Smith, P. B., Dugan, S., & Trompenaars, F. (1996). National culture and values of organizational employees. *Journal of Cross-Cultural Psychology, 27,* 231–264.

Smith, P. B., Trompenaars, F., & Dugan, S. (1995). The Rotter Locus of Control scale in 43 countries: A test of cultural relativity. *International Journal of Psychology, 30,* 377–400.

SoftStat, Inc. (1995). Statistica (Vol. 3): Statistics II [Computer software and manual]. Tulsa, OK: Author.

Steel, P., & Ones, D. S. (2002). Personality and happiness: A national-level analysis. *Journal of Personality and Social Psychology, 83,* 767–781.

Terracciano, A., Abdel-Khalek, A. M., Ádám, N. Adamovová, L., Ahn,

C.-K., Ahn, H.-N., et al. (in press). National character does not reflect mean personality trait levels in 49 cultures. *Science.*

Triandis, H. C. (1995). *Individualism and collectivism.* Boulder, CO: Westview Press.

Twenge, J. M. (2001). Birth cohort changes in extraversion: A cross-temporal meta-analysis, 1966–1993. *Personality and Individual Differences, 30,* 735–748.

van de Vijver, F. J. R., & Leung, K. (1997). Methods and data analysis of comparative research. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology: Vol. 1. Theory and method* (pp. 257–300). Boston: Allyn & Bacon.

van Hemert, D. A., van de Vijver, F. J. R., Poortinga, Y. H., & Georgas, J. (2002). Structural and functional equivalence of the Eysenck Personality Questionnaire within and between countries. *Personality and Individual Differences, 33,* 1229–1249.

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin, 96,* 465–490.

Appendix

Contributing Members of the Personality Profiles of Cultures Project

The 79 coauthors of this article, listed alphabetically by country, are Nora B. Leibovich and V. Schmidt, University of Buenos Aires and National Research Council, Buenos Aires, Argentina; Jane Shakespeare-Finch, Queensland University of Technology, Brisbane, Australia (now at the University of Tasmania); Aljoscha Neubauer, Karl-Franzens-University Graz, Graz, Austria; Filip De Fruyt, Ghent University, Ghent, Belgium; Margaret Munyae, University of Botswana, Gaborone, Botswana; Carmen Flores-Mendoza, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; Donatien Dahourou, University of Ouagadougou, Ouagadougou, Burkina Faso; Lindsay E. Ayearst and Krista K. Trobst, York University, Toronto, Ontario, Canada; Lidia Alcalay and Franco Simonetti, Pontificia Universidad Católica de Chile, Santiago, Chile; Denis Bratko, University of Zagreb, and Iris Marušić, Center for Educational Research and Development, Zagreb, Croatia; Martina Hřebíčková, Academy of Sciences of the Czech Republic, Brno, Czech Republic; Erik Lykke Mortensen and Hans Henrik Jensen, University of Copenhagen, Copenhagen, Denmark; Jüri Allik and Anu Realo, University of Tartu, Tartu, Estonia; Tilahun Sineshaw (Ethiopia), Ramapo College of New Jersey, Mahwah, New Jersey, U.S.A.; Jean-Pierre Rolland, Jean-Michel Petot, and Nathalie Camart, Paris 10-Nanterre University, Paris, France; Alois Angleitner and Fritz Ostendorf, Universität Bielefeld, Bielefeld, Germany; Michelle Yik, Hong Kong University of Science and Technology, Hong Kong, People's Republic of China; Friðrik H. Jónsson, University of Iceland, Reykjavík, Iceland; V. S. Pramila, Andhra University, Visakhapatnam, India; Magdalena S. Halim, Atma Jaya Indonesia Catholic University, Jakarta, Indonesia; Shahriar Shahidi, Shahid Beheshti University, Tehran, Islamic Republic of Iran; Claudio Barbaranelli, University of Rome "La Sapienza," Rome, Italy; Yoshiko Shimonaka, Bunkyo Gakuin University, Irumagun, and Katsuharu Nakazato, Iwate Prefectural University, Takizawa, Japan; Ahmed M. Abdel-Khalek and Badr M. Alansari, Kuwait University, Safat, Kuwait; Brigitte Khoury, American University of Beirut Medical Center, Beirut, Lebanon; Khairul A. Mastor, Universiti Kebangsaan Malaysia, Selangor Darul Ehsan, Malaysia; Ruth Falzon, Mary Ann Lauri, Mary Ann Borg Cunen, and Sandra Scicluna Calleja, University of Malta, Msida, Malta; Rolando Diaz-Loving, Universidad Nacional Autónoma de México, Mexico City, Mexico; M. Arif Ghayur, Al Akhawayn University, Ifrane, Morocco (now at San Diego State University, San Diego, California, U.S.A.); Ronald Fischer, Victoria University of Wellington, Wellington, New Zealand; A. Oluyinka Ojedokun, University of Ibadan, Ibadan, Nigeria; Garry Prentice and Margaret McRorie, Queen's University, Belfast, Northern Ireland; Lei Wang, Peking University, Beijing, People's Republic of China; Norma Reátegui and Marina Brunner-Sciarra, Universidad Peruana Cayetano Heredia, Lima, Peru; Gregorio E. H. del Pilar, University of the Philippines, Quezon City, the Philippines; Andrzej Sekowski and Waldemar Klinkosz, Catholic University of Lublin, Lublin, Poland; Margarida Pedroso de Lima, Universidade de Coimbra, Coimbra, Portugal; José Porrata, University of Puerto Rico, San Juan, Puerto Rico; Thomas A. Martin (Russia), Susquehanna University, Selinsgrove, Pennsylvania, U.S.A; 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; Hyun-nie Ahn and Chang-kyu Ahn, Pusan National University, Pusan, South Korea; Maria D. Avia and Maria L. Sánchez-Bernardos, Universidad Complutense de Madrid, Madrid, Spain; Willibald Ruch, University of Zürich, Zürich, Switzerland; Jérôme Rossier, University of Lausanne, Lausanne, Switzerland; Niyada Chittcharat, Srinakharinwirot University, Bangkok, Thailand; Sami Gülgöz, Koç University, Istanbul, Turkey; Florence Nansubuga, Makerere University, Kampala, Uganda; Peter B. Smith, University of Sussex, Brighton, United Kingdom; David Matsumoto, San Francisco State University, San Francisco, California, U.S.A.; Ed Diener, University of Illinois at Urbana–Champaign, Urbana-Champaign, Illinois, U.S.A.; Andrew Beer and John Humrichouse, University of Iowa, Iowa City, Iowa, U.S.A.; Louisa Budzinski and Shigehiro Oishi, University of Minnesota, Minneapolis, Minnesota, U.S.A.; and Goran Knežević, Department of Psychology, and Dragana Djuric Jocic, Institute for Psychiatry, Belgrade, Yugoslavia.

Received October 27, 2004
Revision received March 21, 2005
Accepted April 13, 2005 ■