This is the author-manuscript version of a paper that was published in:

Journal of Personality and Social Psychology, Vol 89(3), Sep 2005. pp. 407-425.

Personality Profiles of Cultures: Aggregate Personality Traits

Robert R. McCrae

and

Antonio Terracciano

and

79 Members of the Personality Profiles of Cultures Project

Keywords: Personality, Five-Factor Model, Cross-cultural, Culture-level analyses

Corresponding Author:

Robert R. McCrae, Box #03

Gerontology Research Center

5600 Nathan Shock Drive

Baltimore, MD 21224-6825

email: mccraej@grc.nia.nih.gov

Abstract

The 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 = 12,122). Aggregate scores on Revised NEO Personality Inventory (NEO-PI-R) scales generalized across age and gender groups, yielded a close approximation to the individual-level Five-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 were replicated, but appeared to be artifactual. 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.

Personality Profiles of Cultures, I: Aggregate Personality Traits

There is enormous appeal in the idea that cultures have distinctive personalities. Ruth Benedict's (1934) classic description of Pueblo culture as *Apollonian*—sober, conventional, cooperative, and orderly—seems apt and insightful. But 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 are 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. But 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 (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 (McCrae, 2004). But 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, we know that anxiety, hostility, and depression covary among individuals to define a Neuroticism factor (Watson & Clark, 1984), but are anxious cultures also usually hostile and depressed cultures? If not, the concept of Neuroticism 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 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 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 Filipino traits, but these judgments did not match observed mean profiles. The Personality Profiles of Cultures Project was designed to help resolve these issues.

Conceptualizing Personality in Cultures

There are at least three ways in which the personality of a culture might be conceptualized, which we will call *ethos, national character,* and *aggregate personality*. *Ethos*, at a superorganic level (Kroeber, 1917), refers to trait-like 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 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 is high in

Norway compared to other cultures. The whole culture is represented by the mean of its parts—the culture members—in this formulation, 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 computed. In one sense, this is precisely like comparing other groups, such as patients with different personality disorders (Morey et al., 2002). But cross-cultural psychologists have long noted that cross-cultural comparisons pose special challenges (McCrae, 2001; van de Vijver & Leung, 1997). Cross-cultural comparisons 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 employs 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 universal (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 (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. A second problem with IRT analyses is that samples from two cultures might have identical distributions of item scores, and thus no differential item 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 showed 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.

But 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 took a test.

Cross-cultural methodologists have focused on these bottom-up approaches because most cross-cultural studies are 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. But with the recent availability of data from large numbers of cultures, a completely different, top-down approach is now possible 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. Multi-method 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 Neuroticism scores with the mean Eysenck Personality Questionnaire Neuroticism 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 expectable result. When two variables covary, groups that happen for any reason to be high on one will tend also to be high on the other; when group-level data are analyzed, these two variables will 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, due to 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 without affecting 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.

As a basis for interpreting the ecological factor analyses reported here, we will conduct simulations of these conditions and evaluate the resulting factor congruences with the normative individual-level structure. A first simulation will randomly reassign subjects to "cultures," to show that such groupings retain the individual-level structure. A second simulation will add 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 will model systematic variation between "cultures" by contrasting hypothetical Thinking and Feeling cultures.

Aggregate Personality Profiles in 51 Cultures

The present study builds upon 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 gender groups; (b) culture-level factor analysis replicated the individual-level factor structure, though with a broader Extraversion 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. All of these findings argued 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, despite the widespread perception that the Japanese are an industrious people. Poortinga, van de Vijver, and van Hemert (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 responses 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 are due to shared response biases, an alternative method of measurement—observer ratings—was used 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 et al., in press) showed that the basic structure of personality traits was universal, and that age and sex differences seen in self-report studies (Costa, Terracciano, & McCrae, 2001; McCrae et al., 1999) were generally replicated 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 were 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 is included, along with a measure of data quality, to examine associations of these artifacts with variations in scale variances.

We also assess the generalizability of aggregate personality scores across men and women and college-age and adult subsamples and the interrater reliability of the aggregate scores; examine the culture-level factor structure of the NEO-PI-R; and correlate 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) makes it possible to examine the culture-level convergence for specific traits in the present study. To characterize cultures as a whole, we analyze 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 consider the effects of national wealth, aggregate acquiescence, and within-culture sampling on these cross-cultural comparisons.

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 are reported in McCrae et al. (in press). For the present paper, data from Iran (*N*s = 35 male, 38 female raters; 137 targets) became available. Domain reliabilities in the Iranian sample were .92, .88, .84, .93, and .95 for Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C), respectively. After targeted rotation, factor congruence coefficients comparing the Iranian structure to 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 is given in McCrae et al. (in press). 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-aged men, adult (over 40) 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 don't 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 24 to 305, with a total of N = 12,122 valid ratings.

Instrument

The NEO-PI-R is a 240-item measure of the FFM. It contains 30 8-item facet scales, six for each of the five basic personality factors, N, E, O, A, and C. Responses are made on a five-point Likert 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 *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 can be 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 the first authors of this article 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 naysaying, proportion of missing data, the first language of the respondent, the 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 reliability and factor replicability (McCrae et al., in press).

The Quality Index was based on ranking within the group of 50 cultures. To estimate quality in the Iranian sample, a multiple regression was used to predict the total Quality Index from its components in the original 50 cultures. Four predictors were significant: The percent of the unscreened sample with valid protocols (VALID); the judgment that respondents had problems with the questionnaire (PROBLEM; 0 = no, 1 = yes); the percent of the unscreened sample which exceeded the cut-offs 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 non-native, 0 = somewhat fluent non-native 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, due to 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, in press); the Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1975) as reported 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 towards 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 vs. Secular-Rational values and Survival vs. 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 vs. Egalitarian Commitment and Loyal Involvement vs. 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 Gross Domestic Product (GDP; 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 is required in matching cultures across these studies, because cultures were 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, N. Ireland was matched with the U. K. or 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 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., McCrae, 2002; Leung & Bond, 2004; Steel & Ones, 2002) have reported correlations between aggregate-level NEO-PI-R self-report data and other culture-level variables. For the present study, these correlations were recalculated using all available cultures

and the matching rules noted above, to assess replicability of culture-level associations across methods. Note that these are not strict replications, 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, the mean raw domain scores for college-aged subsamples were correlated with mean domain scores for adult subsamples matched on culture and gender (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 .67, .46, .52, .62, and .33, respectively (all ps < .001), suggesting that culture-level scores generalize at least minimally across these age groups. To assess generalizability across gender, mean raw domain scores for female subsamples were correlated 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 .54, .78, .76, .64, and .84, respectively (all ps < .001), suggesting generalizability across genders.

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). Intraclass correlations usually apply to ratings given by a set of judges of the same target. Here, the targets are the different individuals, but all are representatives of the same culture. These values were

.88, .91, .92, .91, and .89 for N, E, O, A, and C, respectively. As shown in the eighth column of Table 1, ICCs for the 30 facets ranged from .80 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 targets.

Age and gender differences at the group level were examined by paired *t*-tests. Older subsamples scored lower on N, E, and O, and higher on A and C than younger subsamples (all *p*s < .001); female groups scored higher than male groups on all five factors (all *p*s < .01). To adjust for these differences, the 30 NEO-PI-R facet scores were standardized as *T*-scores within age and gender groups across all 51 cultures, and all subsequent analyses used these facet 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 et al., in press).

Ecological Factor Analysis Simulations

To test for the effects of cultural influences on ecological factor analyses, all cases were randomly reassigned to 201 "cultures" to parallel the 201 subsamples. A culture-level principal components analysis was conducted on the means of the 30 facet scales in these randomly-constituted "cultures," five factors were extracted, and the factors were rotated to maximal fit 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 .95 to .98.

To simulate the effect of random cultural contributions to the factor structure, 30 random variables were created with an expected mean of 0 and standard deviation of 4 *T*-score points. These perturbations were added to the facet scores of the 201 "cultures;" the mean absolute

change in facet scores was 3.2 *T*-score points. However, these relatively modest random changes had a pronounced effect on the factor structure: Factor congruence coefficients ranged from .24 for O to .62 for E and A; the total congruence coefficient was .49. 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. In this analysis, factor congruence coefficients were .86, .86, .48, .82, and .88 for N, E, O, A, and C, respectively, with a total congruence coefficient of .79. It thus appears that even small deviations from scalar equivalence can degrade the factor structure.

Finally, to simulate the effect of systematic cultural contributions to ecological factor structures, we divided the 201 "cultures" 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 were .98, .90, .61, .95, and .97 for N, E, O, A, and C, respectively; five of the O facets had positive loadings on the O factor, whereas O3: Feelings loaded –.58. Systematic cultural contributions of this magnitude are thus clearly noticeable in ecological factor analyses.

Ecological Factor Analysis

A culture-level principal components analysis was conducted on the means of the 30 facet scales in 201 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 8.18, 4.23, 2.99, 2.32, 1.79, 1.58, and .98, and parallel analysis (Cota, Longman, Stewart, Holden, & Fekken, 1993) indicated that six factors should be 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 to .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, with O and C facets distributed across three factors. Thus, the six-factor solution is 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. But 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 similarity was beyond chance for all five factors, 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 four of the six E facets and by O3: Feeling and A3: Altruism, which have secondary loadings on the E factor in individual-level analyses. But 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.

Table 1 about here

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 .83, .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 (2002) 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. The anomalies with the E factor thus replicate 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 (2002) noted was related to cultural differences in individualism-collectivism.

On the other hand, 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, are 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, overall mean factor and facet scores were calculated. 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, 11.3, 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.

Table 2 about here

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 gender; correlations ranged from .15 to .73, of which 28 were significant ($p < .05$). Similar analyses showed generalizability across gender, rs = .28 to .76, 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 201 subsamples showed a single large factor accounting for 39% 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 *SD*s for each culture are reported in column 7 of Table 2, and the Table entries have been sorted in ascending magnitude of this value. As in McCrae (2002), this arrangement highlights the geographical organization of results: Asian and African cultures show lower variability, whereas European and American cultures show higher. These values are significantly

correlated (r = .61, N = 26, p < .001) with mean *SD*s in self-reports (McCrae, 2002), but also with Acquiescence (r = –.28, N = 51, p < .05) and especially the Quality Index (r = .66, N = 51, p < .001). Acquiescent responding, when applied to a balanced scale, reduces variance, as does random error. These correlations suggest that apparent differences in facet scale variance across cultures may be due largely or entirely to artifacts of response style.

Within-Nation Variability

In four 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. Data for English and Northern Irish are also in Table 2. These two parts of the United Kingdom do not differ in N, E, A, or C, but they are dramatically different in O: the English rank 4th, whereas the Northern Irish rank 49th. Where there are linguistic or historical reasons for treating subcultures separately, that 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 higher than any of the 51 cultures. Had we relied on data from a single American site, we might have reached a wide range of conclusions about Americans' level of E.

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 socio-economic indicators (see Table 3). The most direct comparison is with the factors in self-reports

on the NEO-PI-R. Significant, and moderately large, correlations are found for N, E, and O factors, and a trend ($p < .10$) is found for C. Observer-rated A is related to self-reported E rather than A, but there are no other failures of discriminant validity.

Table 3 about here

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 are small even in comparisons at the individual level. A significant correlation is found for N using data from Lynn and Martin (1995), but none of the other hypotheses is supported. Thus, this cross-method, cross-instrument comparison provides little evidence of validity for the culture-level scores. Similarly, there is no association with external Locus of Control, which at the individual level is modestly related to N and low C (Costa, McCrae, & Dye, 1991).

Aggregate personality factor scores are, however, significantly related to a number of culture-level variables that characterize societies' beliefs and values. N is related to Uncertainty Avoidance, a dimension associated with anxiety (Hofstede, 2001). Cultures whose members are high in E have democratic values, as seen in correlations with Smith et al.'s (1996) Egalitarian Commitment scale and low Power Distance. E is 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 are high in O also are characterized by low Power Distance and

high Individualism. In addition, Open cultures value Affective and Intellectual Autonomy and Egalitarian Commitment, but reject Conservatism. They have a secular-rational approach to life, and limited belief in religion. Open cultures thus appear to be independent and unconventional. Agreeableness, another dimension associated with values at the individual level (Roccas, Sagiv, Schwartz, & Knafo, 2002), has a similar set of correlates, except that high A cultures do not reject religion, and they score higher on subjective well-being (cf. McCrae & Costa, 1991). C is unrelated to values and beliefs when zero-order correlations are examined.

The pattern of correlates in Table 3 is meaningful and generally consistent with previous findings. As Table footnotes show, 17 of the 31 significant correlations between observer-rated NEO-PI-R factors and other criteria are replicated when aggregated self-report data are used to measure the factors.

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 overlap 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 (80%) are significant, and the median value is .58. Note that four of the A facets and four of the C facets are significant, despite limited agreement on A and C factor scores. These data provide evidence that a variety of specific traits may be validly assessed at the culture level.

Control Analyses

Aggregate E, O, and A are all related to GDP and to HDI (see Table 3), and some researchers believe that culture-level correlations should be interpreted net of economic indicators (e.g., Hofstede, 2001; Leung & Bond, 2004). As indicated by Table 3 footnotes, only about a third of the significant correlations in Table 3 remain significant after controlling for

GDP. The most pronounced effects of partialling GDP are on the associations of personality with values. By contrast, the correlations with NEO-PI-R self-report aggregates are relatively unaffected; indeed, the partial correlation for C is now significant at conventional levels (r = .41, p < .05). Controlling for GDP also improves discriminant validity: The unexpected correlation of observer-rated A with self-reported E is reduced to nonsignificance. Analyses for facets (see Table 1) controlling for GDP found that 23 of the 24 significant correlations remained significant (E1: Warmth was the exception).

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 are 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 affect culture-level means. In fact, however, culture-level Acquiescence (see Table 2) was significantly related only to O (r = –.37, p < .01), and partialling it out of the correlations reported in Table 3 had little effect. Correlations of O with Intellectual Autonomy, Religiosity, Smith et al.'s (1996) Egalitarian Commitment, and the HDI became non-significant; the remaining 32 significant correlations in Table 3 changed little in magnitude and remained significant. Partialling Acquiescence from the correlations between Form S and Form R facets (Table 1) reduced the correlation for N2: Angry Hostility to r = .38, p < .10. 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 mean of .38. This is comparable to the mean r_{pa}, .41, found at the individual level for agreement between self-reports and peer ratings from knowledgeable acquaintances (McCrae, 1993). Most importantly, it is similar for all 28 cultures, suggesting that aggregate assessments are valid across a wide range of 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. Most mean scores from both self-reports and observer ratings are near $T = 50$, so agreement across methods is expectable. As an alternative, the aggregate scores were standardized across the 28 cultures, and r_{pa} was calculated on these standardized scores. The resulting values ranged from –.26 for Denmark to .83 for Malaysia, with a mean of .40. These standardized r_{pa}s correlated .71 with the unstandardized r_{pa}s, and neither coefficient was related 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.

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 rather than the 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 (McCrae, 2002; J. Rossier, personal communication, August 19, 2004)

are available for 28 cultures that overlap the present sample. After first standardizing across cultures, intraclass correlations ranged from .04 for Austria to .88 for Burkina Faso. Eighteen of these correlations were significant, with three more showing a trend ($p < .10$). Cultures with the largest profile agreement ($rs > .60$) were Belgium, Burkina Faso, France, India, Malaysia, Serbia, Turkey, French Switzerland, and the U. S. The median value (.45) was found for Italy and Croatia.

Data from Italy, a typical case, and Malaysia, a case of good agreement, were chosen to illustrate profile agreement in Figure 1. (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 give related but not wholly redundant characterizations.

Figure 1 about here

Discussion

With few exceptions, the present analyses replicate findings previously reported for aggregate personality traits measured by 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 (a fact probably attributable to artifacts rather than substantive differences in the homogeneity of trait levels); and aggregate scores show meaningful patterns of

convergent and discriminant validity with other culture-level variables. 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: Malaysians are indeed higher in self-consciousness than most other people in the world (see Figure 1), and the English are more open to experience than the Northern Irish.[5] Poortinga and colleagues (2002) are probably not alone in remaining skeptical of such claims, 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 reflect 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 finally, 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. But convergence across measures of traits themselves was

largely unaffected by partialling out GDP.

This does not mean that we now have definitive values for aggregate trait levels in our sample of cultures. Assessments using the NEO-PI-R did not square well with assessments using the EPQ, and as Figure 1 shows, there are clear discrepancies for some facets in some cultures even when different forms of the NEO-PI-R are used. Analyses of within-country variation in the U. S. showed that different sites could yield somewhat different personality profiles.

But the pattern of evidence so far suggests that aggregating individual personality scores is a useful way to characterize cultures. To obtain personality profiles that accurately reflect the culture as a whole, researchers will 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, a large segment of the population that was deliberately omitted here. A most interesting design would include self-reports and observer ratings of the same individuals, to understand better method-of-measurement effects.

Culture-Level Factor Structure

The major finding from the ecological factor analysis was 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 equivalance of NEO-PI-R scales in different cultures. 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 interaction of level-of-analysis by method-of-measurement. 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 but not 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 non-overlapping samples of cultures. Particularly puzzling is the pattern of O facets: Cultures high in E are also high in O1: Fantasy, and O6: Values, but tend to be low in O2: Aesthetics. Introverted cultures (e.g. India; see McCrae, 2002, Figure1) show the opposite pattern. Inglehart (1997) reports that *imagination* and *tolerance* are among the defining values of the self-expression dimension, which is strongly associated with E. Perhaps the culture-level E is generated by the post-materialist values of the post-industrial world.

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 respondents, 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 this. 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, Inglehart and Norris, Smith and colleagues, and Leung and Bond. The most predictable associations were with Openness to Experience. 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). Agreeableness 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.

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) suggest that self-expression arises as industrial societies come to take survival for granted and become post-materialist in outlook. The strong link between self-

expression and Extraversion and the fact that much of the world is rapidly becoming post-industrial 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 hard-working, but in the present data, Japan and Hong Kong are merely average in C. Instead, the highest scoring countries are 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 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 will provide 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.

References

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.

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

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

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: 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): 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, J. D., & 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.

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.

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 Publishers.

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

McCrae, R. R., & Costa, P. T., Jr. (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. (1991). Adding *Liebe und Arbeit*: The full Five-Factor Model and well-being. *Personality and Social Psychology Bulletin, 17*, 227-232.

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., 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: Pabst Science Publishers.

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., Terracciano, A., & 78 Members of the Personality Profiles of Cultures Project. (in press). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*.

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

McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology, 70*, 552-566.

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 perrsonality. *Journal of Personality Disorders, 16*, 215-234.

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.

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. (in press). Structural and mean level analyses of the Five-Factor Model and Locus of Control: Further evidence from Africa. *Journal of Cross-Cultural Psychology*.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 80*, No. 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.

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.

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

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 and 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.

Author Notes

Robert R. McCrae and Antonio Terracciano, National Institute on Aging, National Institutes of Health, Department of Health and Human Services; and 79 members of the Personality Profiles of Cultures Project. A complete list of the 79 coauthors, listed alphabetically by country, can be found at the end of this article.

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 Butkovi□, 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 Zba□nik, 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.

German, Russian, and Czech data were taken from earlier studies (McCrae, Costa, Martin et al., 2004; Ostendorf & Angleitner, 2004), and portions of the Thai, Brazilian, and Lebanese data are also reported in chapters (Costa & McCrae, in press; McCrae, in press; McCrae, Terracciano, & Khoury, in press). Portions of these data were presented at the 2nd World Congress on Women's Mental Health, March, 2004, Washington, DC. Czech participation was supported by Grant 406/01/1507 from the Grant Agency of the Czech Republic and is related to research plan AV 0Z7025918 of the Institute of Psychology, Academy of Sciences of the Czech Republic. S. Gulgöz's participation was supported by the Turkish Academy of Sciences.

Burkinabè and French Swiss participation was supported by a grant from the Swiss National Science Foundation to J. Rossier. The data collection in Hong Kong was supported by RGC Direct Allocation Grants (DAG02/03.HSS14 and DAG03/04.HSS14) awarded to M. Yik. Data collection in Malaysia was supported by UKM Fundamental Research Grant 11JD/015/2003.

Robert R. McCrae receives royalties from the Revised NEO Personality Inventory.

Correspondence concerning this article may be sent to Robert R. McCrae, Box #03, Gerontology Research Center, 5600 Nathan Shock Drive, Baltimore, Maryland, 21224-6825. Email: mccraej@grc.nia.nih.gov

Footnotes

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

[2]The self-report correlations are available from the first author.

[3]There were no Canadian data for adult males, and no Iranian data for college-age targets, so the total number of subsamples was 201.

[4]Previous research had used U. S. age and gender 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. For comparison with previous work, data in the present study were also standardized using the U. S. data collected in the present study, with very similar results. The international norms used in the present study are available from the first author.

[5]These statements refer to people on average. Recall that there is a wide range of individual differences on all traits in all cultures.

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	**.78**	.09	–.14	.07	.17	.93[d]	.90	.69***
N2: Angry Hostility	**.66**	–.07	–.18	**–.43**	–.09	.97[e]	.86	.39*
N3: Depression	**.53**	–.22	–.23	.17	**–.42**	.84	.89	.53**
N4: Self-Consciousness	.33	**–.41**	–.18	.35	–.14	.70	.91	.61***
N5: Impulsiveness	**.51**	**.52**	.17	–.19	–.27	.96[e]	.87	.63***
N6: Vulnerability	**.62**	–.38	–.16	–.07	–.35	.94[e]	.88	.57***
E1: Warmth	–.02	**.67**	.19	**.45**	.19	.99[e]	.94	.43*
E2: Gregariousness	–.37	**.63**	–.11	.17	–.18	.92[d]	.88	.34
E3: Assertiveness	**–.49**	.30	.00	–.28	.31	.91[d]	.80	.23
E4: Activity	.06	**.44**	.33	.10	.35	.82	.87	.61***
E5: Excitement Seeking	–.21	.39	–.23	–.24	**–.48**	.62	.96	.47*
E6: Positive Emotions	–.26	**.72**	.18	.24	.12	.95[e]	.91	.52**
O1: Fantasy	.18	**.57**	**.58**	.02	–.18	.86[d]	.92	.59***
O2: Aesthetics	–.10	–.25	**.69**	.21	.21	.88[d]	.90	.50**
O3: Feelings	.04	**.49**	**.59**	.26	.21	.84	.95	.63***
O4: Actions	–.20	–.14	**.72**	.04	–.20	.84	.89	.45*
O5: Ideas	–.35	–.01	**.62**	.10	.22	.92[d]	.85	.65***
O6: Values	.12	**.53**	**.53**	.21	.05	.62	.97	.74***
A1: Trust	–.28	**.41**	.21	**.60**	.05	.97[e]	.91	.40*
A2: Straightforwardness	.15	**.45**	.07	**.57**	.22	.62	.94	.26
A3: Altruism	.08	**.65**	.17	**.43**	**.42**	.92[d]	.96	.72***
A4: Compliance	–.34	–.24	.14	**.73**	–.07	.94[e]	.88	.32
A5: Modesty	.34	.34	–.09	**.60**	.04	.76	.92	.63***
A6: Tender-Mindedness	.11	.16	.13	**.69**	.25	.93[d]	.93	.60***
C1: Competence	–.25	**.46**	.24	.12	**.70**	.92[d]	.94	.65***
C2: Order	–.15	–.29	–.16	.23	**.65**	.84	.84	.68***
C3: Dutifulness	.02	.22	.08	.35	**.85**	.92[d]	.94	–.02
C4: Achievement Striving	–.22	.10	–.06	.05	**.74**	.91[d]	.91	.68***
C5: Self-Discipline	–.17	.18	.04	.13	**.83**	.97[e]	.85	.08
C6: Deliberation	–.36	**–.44**	–.17	.23	**.56**	.97[e]	.85	.70***
Factor Congruence[c]	.87[e]	.80[e]	.89[e]	.90[e]	.94[e]	.88[e]		
Factor Comparability	.84	.93	.94	.90	.96			

Note: These are principal components from 201 subsamples targeted to the American normative

factor structure. Loadings greater than .40 in absolute magnitude are given in boldface. [a]Variable congruence coefficient; 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. $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$.

Table 2.

Aggregate Factor T-*scores, Mean Facet Standard Deviation, Quality Index, and Acquiescence in 51 Cultures.*

	Factor						Quality	Acquiescence
Culture	**N**	**E**	**O**	**A**	**C**	*SD*[a]	Index	Score
Moroccans	50.5	44.9	48.4	46.1	45.5	7.6	5.4	516.7
Ethiopians	48.8	47.1	48.5	47.3	47.2	7.9	10.8	522.5
Malays	51.8	48.3	47.5	51.7	53.0	7.9	13.3	521.3
Nigerians	47.8	44.4	49.1	46.6	45.8	8.0	13.1	507.3
Ugandans	49.4	46.5	49.5	48.3	48.2	8.3	5.9	518.7
Indians	50.1	48.5	48.7	51.7	52.3	8.4	15.8	555.1
Burkinabé	53.1	48.8	49.2	51.3	49.7	8.5	21.3	534.7
Kuwaitis	51.9	52.9	47.5	51.0	52.6	8.7	18.9	542.2
Russians	51.4	45.8	49.7	50.3	49.1	8.7	16.3	527.7
Peruvians	48.5	50.1	48.9	48.5	48.7	8.8	19.2	501.2
Botswana	48.9	46.8	47.7	48.0	46.8	8.9	13.4	514.8
P. R. C. Chinese	46.6	46.6	50.0	48.6	48.0	8.9	16.1	517.4
Germans	48.1	49.6	54.9	52.1	52.3	9.1	37.6	516.8
Slovaks	49.2	49.7	48.1	50.5	48.6	9.2	30.1	508.5
Lebanese	50.0	51.2	48.1	46.3	50.5	9.2	9.9	519.4
H. K. Chinese	50.5	46.2	47.3	46.9	49.6	9.3	25.9	522.9
Indonesians	50.0	45.5	48.9	49.0	49.6	9.3	22.7	515.3
Filipinos	48.3	48.9	50.8	47.3	53.5	9.4	18.1	504.2
Mexicans	46.2	47.9	50.2	47.4	50.8	9.5	15.8	493.9
Canadians	49.6	53.1	47.8	49.9	49.1	9.6	27.6	523.3
S. Koreans	48.4	50.7	50.9	50.3	48.3	9.6	26.3	494.0
Belgians	49.6	52.3	50.4	49.8	47.4	9.7	32.9	502.7
Icelanders	48.6	51.5	51.2	52.0	49.4	9.7	29.6	508.8
Portuguese	51.6	51.4	51.3	51.0	50.7	9.7	32.6	512.9
Puerto Ricans	49.9	51.7	49.6	48.8	52.9	9.7	13.6	521.6

Japanese	50.7	49.5	51.2	48.8	49.5	9.9	26.6	486.8
German Swiss	47.6	48.3	58.4	54.0	53.5	10.0	34.3	496.0
Americans	48.1	52.3	50.3	49.1	48.9	10.0	25.3	504.6
Italians	52.6	46.5	52.3	48.1	48.4	10.0	25.5	488.6
Croatians	49.3	51.0	49.1	48.4	50.3	10.1	17.0	514.0
Australians	48.6	53.9	50.7	50.0	47.5	10.1	27.2	522.2
Poles	50.7	49.2	48.6	48.5	49.4	10.2	31.3	515.7
Slovenians	50.7	49.5	48.8	49.0	52.4	10.2	13.7	515.0
Thais	48.9	49.7	48.4	49.6	48.9	10.2	24.7	521.0
Argentineans	51.3	52.3	46.1	50.6	50.1	10.3	22.6	497.1
English	50.1	53.7	53.5	50.2	48.2	10.3	28.4	512.5
Brazilians	53.7	52.2	49.0	50.3	51.5	10.3	26.0	517.7
French	52.7	48.0	51.4	51.3	48.4	10.3	35.3	496.9
New Zealanders	47.9	52.5	50.1	50.1	47.8	10.4	32.9	514.1
Turks	51.4	53.1	48.1	51.0	51.5	10.4	31.9	517.8
N. Irish	50.1	55.7	47.4	52.4	47.4	10.4	30.2	516.4
Estonians	47.0	50.2	47.9	48.4	48.8	10.5	30.3	506.6
Maltese	53.1	50.6	48.4	49.4	51.6	10.6	31.3	518.8
Serbians	49.4	49.4	51.6	48.4	51.7	10.7	31.3	528.3
Chileans	50.0	51.7	51.8	50.8	52.3	10.8	33.1	496.9
French Swiss	53.6	51.1	51.6	53.0	49.8	10.8	36.7	501.8
Austrians	48.3	50.8	50.5	50.6	52.4	10.8	28.8	512.8
Danes	50.3	51.9	55.2	53.1	48.5	10.8	35.6	499.0
Spaniards	49.7	50.5	48.8	51.4	51.3	11.0	37.4	500.2
Iranians	48.4	48.2	50.1	48.6	47.0	11.1	10.2	528.9
Czechs	51.5	48.2	50.3	54.2	51.5	11.1	30.7	536.1

Note. Quality Index is taken from McCrae et al. (in press); Acquiescence is the mean sum of all NEO-PI-R raw score responses before reflecting. [a]Mean *T*-score-standardized standard deviation across 30 NEO-PI-R facet scales.

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

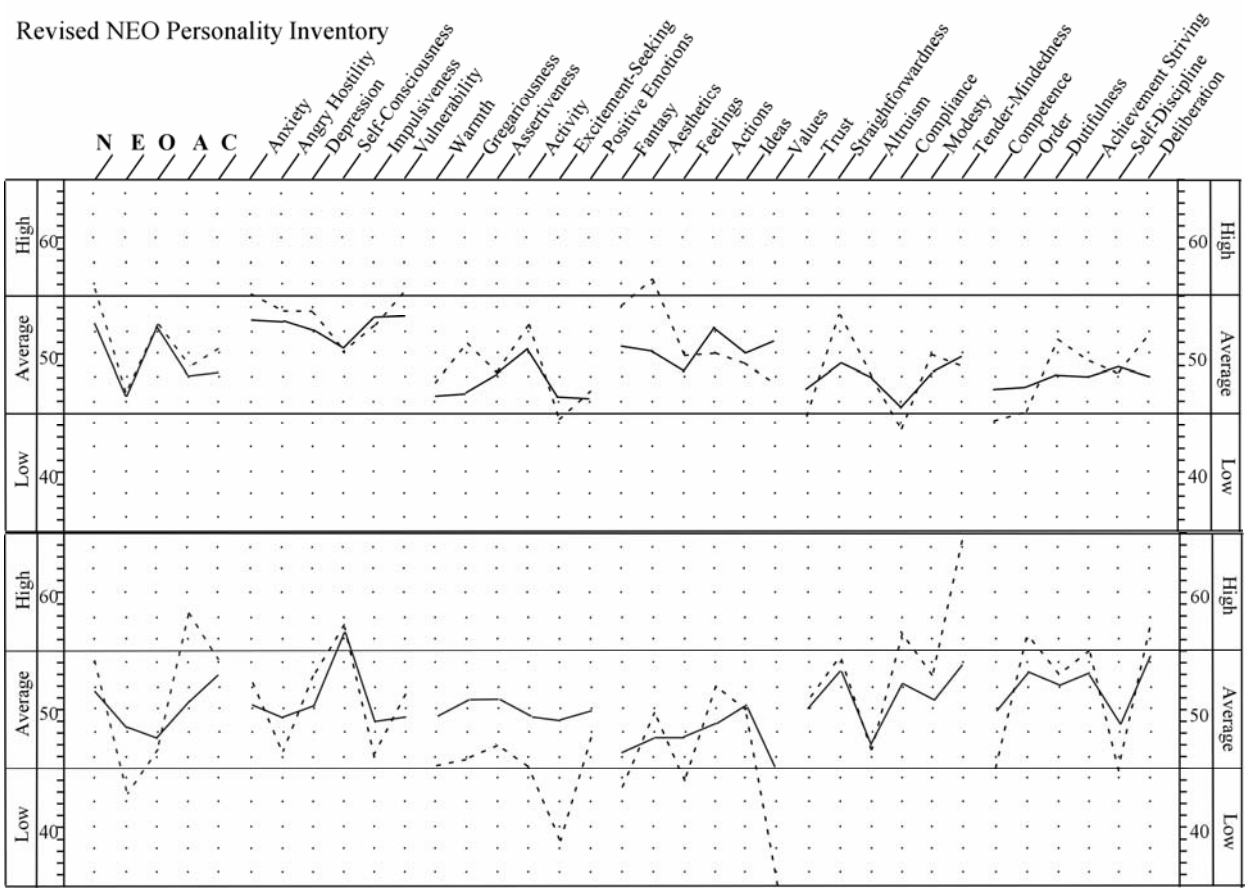
	Factor				
Criterion	**N**	**E**	**O**	**A**	**C**
		Personality Measures			
NEO–PI–R Form S Factors (N = 28)					
Neuroticism	.52**[c]	–.20	–.03	.21	.06
Extraversion	–.07	.58***[c]	.30	.39*	–.03
Openness to Experience	–.17	.08	.51**[c]	.14	.24
Agreeableness	.29	–.06	–.34	.11	.15
Conscientiousness	–.12	–.10	–.30	–.19	.34[c]
EPQ Scales (N = 28; Lynn & Martin, 1995)[a]					
Neuroticism	.41*[b,c]	.11	.19	.14	.09
Extraversion	–.15	.05	.02	–.14	–.08
Psychoticism	–.05	–.23	–.13	.05	.23
EPQ Scales (N = 27; van Hemert et al., 2002)[a]					
Neuroticism	.19	.13	–.06	–.01	–.06
Extraversion	–.31	.35	.07	–.26	–.05
Psychoticism	–.15	–.26	–.05	–.26	.09
Lie (N = 25)	.06	–.55**	–.16	–.51**	.08
Rotter Locus of Control (N = 34; Smith et al., 1995)	.25	–.06	–.14	–.07	.02
		Beliefs, Attitudes, Values			
Hofstede (2001) Dimensions (N = 49)					
Power Distance	.21[c]	–.45**[b]	–.42**[b]	–.32*	.13
Uncertaintly Avoidance	.29*[b,c]	.07	–.03	–.03	.21
Individualism	.05	.51***[b,c]	.33*[b]	.38**	–.15
Masculinity	–.13	.02	.09	.03	.05
Long–Term Orientation (N = 30)	–.09	–.18	–.04	–.18	.00

Schwartz (1994) Values (N = 22)					
Conservatism	–.22	–.07	–.69***[b,c]	–.50*	.11
Affective Autonomy	.15	.28	.54**	.60**[c]	.01
Intellectual Autonomy	.39[c]	–.09[c]	.49*[b]	.43*	.13
Hierarchy	–.22	–.09	–.34	–.24	–.08
Mastery	–.23	–.24	.06	–.13	–.08
Egalitarian Commitment	.24	.21	.56**[c]	.44*	–.08
Harmony	.06	.05	.29	.11	.11
Inglehart & Norris (2003) Values (N = 42)					
Secular–Rational	–.02	.06	.36*	.43**	–.04
Self–expression	–.06	.58***[b,c]	.26	.29	–.07
Social Axioms (N = 29; Leung & Bond, 2004)					
Social Cynicism	–.27	–.28	.07	–.17	–.01
Social Complexity	–.10	.35	.26	.20	.20
Reward for Application	–.30	–.35	–.36	–.23	.20
Religiosity	.07	–.37	–.38*[b]	–.14	.30[c]
Fate Control	–.26	–.58***[b,c]	–.17	–.09	–.04
Organizational Attitudes (N = 34; Smith et al., 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 et al., 1995)	.01	.64***[b,c]	.33	.47**	–.03
Economic Indicators					
Gross Domestic Product per capita (N = 51)	.04	.45***[b]	.46***[b]	.46***	–.02
Gini Index (N = 40)	–.03	–.08	–.25	–.26	.11
Human Development Index (N = 48)	.02	.55***[b,c]	.34*[b]	.40**	.24[c]

[a]Indian data from Lodhi, Deo, & Belhekar (2002). [b]Replicated ($p < .05$, one-tailed) in culture-level analyses using self-report data. [c]Significant after partialling GDP.

Figure Caption

Figure 1. Mean NEO-PI-R 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; towards the right the facet scales are grouped by factor. Profile form reproduced by special permission of the Publisher, from the Revised NEO Personality Inventory, by Paul T. Costa, Jr., and Robert R. McCrae. Copyright 1978, 1989, 1992 by PAR, Inc. Further reproduction is prohibited without permission of PAR, Inc.



The 79 coauthors of this article 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; 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, 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 H⊆ebí□ková, 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, Ramapo College of New Jersey, Mahwah, New Jersey, USA; 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 S. M. Yik, Hong Kong University of Science and Technology, Hong Kong, People's Republic of China; Friξrik 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, 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, CA); Ronald Fischer,Victoria University of Wellington,Wellington, New Zealand; A. Oluyinka Ojedokun, University Of Ibadan, Ibadan, Nigeria; 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; Jose Porrata, University of Puerto Rico, San Juan, Puerto Rico; Thomas A Martin**,** Susquehanna University, Selinsgrove, Pennsylvania, USA; 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, KoH University, Istanbul, Turkey; Florence Nansubuga, Makerere University, Kampala, Uganda; Peter B. Smith, University of Sussex, Brighton, United Kingdom; Garry Prentice and Margaret McRorie, Queen's University, Belfast, Northern Ireland; Louisa Budzinski and Shigehiro Oishi, University of Minnesota, Minneapolis, Minnesota, USA; Ed Diener, University

of Illinois at Urbana-Champaign, Urbana-Champaign, Illinois, USA; Andrew Beer and John Humrichouse, University of Iowa, Iowa City, Iowa, USA; David Matsumoto, San Francisco State University, San Francisco, CA, USA; and Goran Kne⟩evi□, Department of Psychology, and Dragana Djuric Jocic, Institute for Psychiatry, Belgrade, Yugoslavia.