Trait Psychology and Culture: Exploring Intercultural Comparisons

Robert R. McCrae

National Institute on Aging, NIH

ABSTRACT   Personality traits, studied for decades by Western personality psychologists, have recently been reconceptualized as endogenous basic tendencies that, within a cultural context, give rise to habits, attitudes, skills, beliefs, and other characteristic adaptations. This conceptualization provides a new framework for studying personality and culture at three levels. *Transcultural* research focuses on identifying human universals, such as trait structure and development; *intracultural* studies examine the unique expression of traits in specific cultures; and *intercultural* research characterizes cultures and their subgroups in terms of mean levels of personality traits and seeks associations between cultural variables and aggregate personality traits. As an example of the problems and possibilities of intercultural analyses, data on mean levels of

Revised NEO Personality Inventory scales from college age and adult samples ($N = 23,031$) of men and women from 26 cultures are examined. Results showed that age and gender differences resembled those found in American samples; different subsamples from each culture showed similar levels of personality traits; intercultural factor analysis yielded a close approximation to the Five-Factor Model; and factor scores were meaningfully related to other culture-level variables. However, mean trait levels were not apparent to expert raters, casting doubt on the accuracy of national stereotypes. Trait psychology can serve as a useful complement to cultural perspectives on human nature and personality.

In 1950, after dismissing the psychological characterizations of cultures by anthropologists as "oversimplified, distorted descriptions, stillborn from an infantile methodology," (p. 391) Cattell admitted that "no quantitative, scientific results exist" (p. 440) linking personality traits to culture. Translations of objective personality measures were scarce, and opportunities for cross-cultural research rare. But by 1973, Cattell's own 16PF had been translated into 19 languages, and today most major personality questionnaires have foreign language versions (e.g., Butcher, 1996). Methodologists (van de Vijver & Leung, 1997) have formalized principles of comparative research, and over the past quarter century, cross-cultural psychologists have routinely compared different cultures on various personality traits.

The accumulated results of these studies, together with advances in trait psychology itself, have now made it possible to define a discipline of personality traits and culture (McCrae, 2000; see also Church, 2000). The goal of anthropologists in the first half of the 20th century was to show how patterns of culture shaped personality; the goal of the new discipline is to determine how personality traits interact with culture in shaping people's lives.

The new conceptualization of traits as independent rather than dependent variables is based on several lines of research (McCrae et al., 2000). Longitudinal studies have shown that traits are stable over long periods despite intervening life events; behavior genetic studies have shown that genes, not child-rearing practices, are major determinants of adult traits; cross-cultural studies have suggested that both the five-factor structure of personality traits and their developmental course in adulthood are universal. Cultures may yet have some detectable influence on personality traits—a question to which most of the present article is addressed—but in a very profound way traits transcend culture.

What cultures incontestably do influence are the acquired skills, habits, attitudes, interests, roles, and relationships that McCrae and Costa (1999) have called *characteristic adaptations.* Traits also influence characteristic adaptations, and through them, the behavior that expresses traits. A garrulous Frenchman and a talkative Korean share the same extraverted tendencies, but they express them in culture-specific form and language.

Three Levels of Analysis

As detailed elsewhere (McCrae, 2000), relations between traits and culture can be addressed at three levels. The *transcultural level* concerns questions about universality. Can the same traits be found in different cultures? Are associations among traits similarly structured? Are there universal age and gender differences? Evidence so far suggests that the answer to these questions is "yes." In particular, imported measures of the Five-Factor Model of personality show very similar structures in all the cultures in which they have so far been examined (McCrae, Costa, del Pilar, Rolland, & Parker, 1998; Paunonen, Jackson, Trzebinski, & Forsterling, 1992).[1] It is in fact strong evidence of universality that makes the study of traits across cultures possible: We could not ask whether Indonesians were more agreeable than Mexicans if Agreeableness were not a factor of personality in both cultures.

Studies at the *intracultural level* are concerned with the culture-specific expression or operation of a trait. Benet-Martínez and John (2000), for example, identified indigenous trait adjectives in Castilian Spanish associated with the five universal personality factors. Openness to Experience was expressed by such terms as *extravagante, excéntrica, bohemia* (extravagant, eccentric, Bohemian) that emphasize an unconventional lifestyle. Elsewhere, Openness might be seen chiefly in intellectual interests or in aesthetic appreciation. Because such emic versions of traits must be induced from a study of the indigenous language and culture, ethnographic methods might be particularly useful in intracultural studies of traits.

---

1. This does not preclude the possibility that there may be additional, emic, dimensions of personality unique to particular cultures. If so, the full scope of personality traits would not be universal, although there would still be a universal core.

Finally, analyses at the *intercultural level* ask about cultural differences in traits, and seek associations between traits and features of culture. Most often, mean levels of traits will be the focus of interest; national character could be described in terms of a mean personality profile. But one could also ask about cultural differences in standard deviations of trait scores, or cultural differences in the alignment of factors (Kallasmaa, Allik, Realo, & McCrae, 2000; McCrae, Costa et al., 1998).

Perhaps the most ambitious attempt to date to make intercultural comparisons in personality examined the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975) scales. Factor analyses of translated versions had been used to eliminate items that did not work well in translation, and mean scores were then prorated to equal item lengths. Barrett and S. Eysenck (1984) reported prorated means for men and women in 25 countries, and Lynn and Martin (1995) extended their list to 37. Personality profiles appeared to be similar for men and women, but no other evidence was offered that the scores were representative of the cultures or comparable across cultures.

Intercultural analyses are generally multicultural analyses. When only two cultures are compared, it is difficult to identify the features of culture that may be causally associated with differences in personality, but if a range of cultures is considered, associations can be tested. There are countless features of culture that differentiate Americans and Chinese, any one of which might account for differences in Extraversion (McCrae, Yik, Trapnell, Bond, & Paulhus, 1998). But if individualistic and collectivistic cultures in general differ in Extraversion, then an association between social structure and trait level can plausibly be asserted.

Intercultural comparisons can lead to objective assessments of national character, which, if accurate, could be of considerable use to travelers, diplomats, and businesspersons. To the extent that these assessments are inaccurate, however, or indiscriminately applied to all members of the culture, they may contribute to harmful stereotypes. There are ethical as well as scientific reasons to be cautious in interpreting cultural differences in personality scores.

Intercultural Comparisons: The Ideal
and the Feasible

Basic intercultural comparisons require data on the mean levels of personality traits in three or more cultures, together with some understanding of the distinctive features of the cultures. With such data it is possible to search for associations between personality traits and patterns of culture. Over the past quarter century, cross-cultural methodologists have made it clear that this is not as simple a task as it might appear. Personality measures have routinely been translated and used in other cultures, but whether they yield accurate assessments of mean levels of traits is subject to considerable question.

A first concern has to do with translation. Even if we assume that the translator is competent and has accurately conveyed the sense of each item (generally checked by examination of an independent back-translation), it remains possible that subtle shifts in phrasing will lead to greater or lessor endorsement and thus to higher or lower means. Item Response Theory is sometimes used to test the scalar equivalence of items; bilingual studies, in which the same subjects complete original and translated versions, can be used to assess comparability of items or scales.

But even if bilinguals score identically on the two versions, that may not guarantee that the test is truly equivalent across cultures. Members of a particular culture may be more prone to endorse any item, or to use extreme categories (Cheung & Rensvold, 2000)—artifacts that would affect responses in both languages. Again, respondents in different cultures may have different self-presentational styles. Judgments by observers, perhaps from outside the culture, may be necessary to detect self-report biases (cf. McCrae, Yik et al., 1998). In principle, all these distortions can and should be corrected before assaying cross-cultural comparisons.

Sampling poses another set of problems for intercultural comparisons. Most psychologists use the term *culture* informally and interchangeably with *country*; they tend to assume that all Peruvians share one culture, distinct from that of all Canadians. Bock (2000) has called this fallacy the *uniformity assumption*, and contemporary anthropologists are careful to characterize their samples in terms of age, gender, social class, and other relevant groupings. Researchers who wish to make intercultural trait comparisons should probably use stratified samples, allowing

empirical test of the degree of uniformity of personality traits across subsamples.

Sampling is critical in intercultural comparisons because the data collected are to be interpreted as representative of an entire culture. Convenient groups like undergraduates may or may not be accurate proxies for national probability samples. It would be clearly inappropriate to use deviant groups like psychiatric patients, unless the same groups are sampled in each culture; even here, it would not be clear that cultural differences could be generalized beyond the group in question.

Marsella, Dubanoski, Hamada, and Morse (2000) argued that ethnocultural identity ought to be assessed as a first step in any cross-cultural comparison in order to screen the sample. If we are interested in the effects of culture, we need to study individuals who have actually adopted the beliefs, values, and practices of the culture, not simply those whose citizenship or ethnic ancestry gives them nominal membership. Undergraduates around the world tend to be Westernized in many respects; they thus make a relatively poor choice for intercultural comparisons. Note, however, that the effect of using Westernized respondents is presumably to minimize intercultural differences; any effects that emerge in analyses of student samples would likely be even stronger in ethnoculturally screened samples.

In the ideal design, then, personality measures of demonstrated scalar equivalence would be administered to stratified representative samples of individuals with the appropriate ethnocultural identity, across a wide range of different cultures. It is doubtful whether anyone has yet collected such data; the data to be analyzed here certainly do not meet these ideal requirements. But it remains to be seen how much departures from the ideal actually affect results. It is possible that personality assessments are more robust than methodologists imagine. The first goal of the present exploration is simply to evaluate the feasibility of intercultural comparisons using the suboptimal data typically available. Do any meaningful patterns emerge from the data at all? Can we have any confidence in the results? Both internal and external checks can be used to answer these questions.

Intercultural Differences on the NEO-PI-R

Gathering data to meet the specifications of an ideal intercultural design would be an enormous task. In what might be seen as a pilot study for

such a project, the present article analyzes data already collected by investigators in 26 cultures. The instrument used is a comprehensive measure of personality traits, and there is sufficient experience with its cross-cultural application to suggest that meaningful results may emerge despite methodological limitations.

Measure

The Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) assesses 30 specific traits, or facets, that define the five basic factors of personality: Neuroticism (N), Extraversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C). These factors are approximated by domain scores, which are the simple sum of six facets, and, more precisely, by factor scores that use weighted combinations of all 30 facets. The instrument has been translated into over 30 languages, with back-translations into English reviewed by the original test authors. In general, these translations have shown adequate reliability (see McCrae et al., 1999). Some are well validated; others have not yet been studied (see Costa, McCrae, & Jónsson, in press, for a review of European translations).

Several articles have addressed the factor structure of the NEO-PI-R across cultures (e.g., Mastor, Jin, & Cooper, 2000; McCrae, Costa et al., 1998; Kallasmaa et al., 2000), and several other unpublished studies have examined the structure; indeed, this is a routine step in evaluating a new translation. In each of the 26 cultures examined here, the five-factor structure has been replicated well beyond chance (McCrae, Zonderman, Costa, Bond, & Paunonen, 1996), although with varying degrees of fidelity to the original. In general, translations into Western languages show closer convergence with the American original than do those into non-Western languages, with the clearest replications for N and C, the least clear for O.

Only a few translations have been evaluated for scalar equivalence with the original American instrument, but results are instructive. In four studies, bilingual volunteers completed both English and translated versions, usually a week apart. Correlations between the two versions for the five domain scores ranged from .64 to .95 (median = .86), suggesting that in each case the translation assessed the same construct. Furthermore, mean level differences were minimal. In Hong Kong, there was a significant difference only in E, with scores approximately one-quarter

SD lower in the Chinese version (McCrae, Yik et al., 1998). In the Hispanic American sample, women scored about .2 *SD* lower on N when completing the Spanish version, but there were no differences for men (Psychological Assessment Resources, 1994). A Korean sample showed no consistent mean level differences (Piedmont & Chae, 1997). Finally, a study in Zimbabwe showed a significant difference only for N, with slightly lower scores in the Shona version than the English (Piedmont, Bain, McCrae, & Costa, in press). To the extent that these studies are representative, they suggest that translation itself is unlikely to introduce large biases in mean level comparisons—probably because the errors introduced in the translation of individual items average out when summed into scales.

But bilingual studies do not address such issues as response styles or self-presentational tendencies. Three studies are relevant here. Huang, Church, and Katigbak (1997) used differential item functioning (DIF) analyses to compare American and Filipino responses to the original NEO-PI administered in English to both samples. They found that almost 40% of the items showed DIF, and many significant cultural differences disappeared when those items were removed. Certainly this argues for caution in interpreting cross-cultural comparisons, especially at the item level. However, Grimm and Church (1999) compared response sets in U.S. and Philippine college students and found that controlling for response styles had little effect on conclusions about cultural differences. McCrae, Yik et al. (1998) supplemented self-reports with peer ratings of Chinese undergraduates and found that Hong Kong-born raters had higher standards for assessing Competence and Achievement Striving and lower thresholds for attributing Depression and Vulnerability than did Canadian-born raters. However, no significant effects were found for the remaining 26 NEO-PI-R facets. Thus, response style effects at the scale level appear to be relatively circumscribed, at least in this instance. If an accurate personality profile for a specific culture is desired, then detailed studies of biases in self-reports are essential. But if such biases are small and scattered, then they are likely to contribute only random error to intercultural studies with a relatively large number of cultures.

Samples

Investigators around the world have collected NEO-PI-R data on samples relevant to their own research interests. Mean values have been

published by several researchers (Cassaretto, 1999; Chen, 1996;[2] Costa & McCrae, 1992; Lee, 1995; Hoekstra, Ormel, & De Fruyt, 1996; Marušić, Bratko, & Eterović, 1997; Mastor, Jin, & Cooper, 2000; McCrae, Yik et al., 1998; Psychological Assessment Resources, 1994; Shimonaka, Nakazato, Gondo, & Takayama, 1999); unpublished data, or data from unpublished analyses of subgroups, were made available by others (for descriptions of related or overlapping data see Avia, Sanz, & Sánchez-Bernardos, 1999; De Fruyt & Mervielde, 1999; Heuchert, Parker, Stumpf, & Myburgh, 2000; Hoekstra & De Fruyt, in press; Kallasmaa et al., 2000; Katigbak, Church, & Akamine, 1996; Knežević, Radović, & Opačić, 1997; McCrae et al., 1999; Piedmont et al., in press; Rolland, 1998; Rolland, Parker, & Stumpf, 1998). Participants in all these studies were volunteers; clinical and occupational selection samples were excluded. Only two samples—Flemish and Portuguese—approximated representative samples, and no assessments of ethnocultural identity were made. The range of cultures was fairly broad, with representatives of five continents and Indo-European, Malayo-Polynesian, Sino-Tibetan, Uralic, and Bantu language families. Some characteristics of the samples are reported in Table 1.

Previous research (including data from the Croatian, Estonian, German, Italian, Japanese, Korean, Portuguese, and Russian samples studied here) has shown age differences within cultures for all five factors (Costa et al., 2000; McCrae et al., 1999). Samples were therefore divided into college age (generally age 18–21, but varying somewhat across cultures) and adult (age 22+). Samples were also stratified by gender, giving a total of 84 subsamples. When raw scores from the 32 adult subsamples are compared to the 52 college-aged subsamples, the expected differences are seen: Adults are lower in N, E, and O, and higher in A and C across the 26 cultures (all $p < .01$). Further, gender differences found across these cultures closely resemble those noted in American samples (Costa, Terracciano, & McCrae, 2001). These results provide a first hint that analyses of NEO-PI-R scales across multiple cultures can yield meaningful results. To control for these age and gender differences and to make scores more easily interpretable, all mean values were converted to *T*-scores (which have a mean of 50 and standard deviation of 10 in the normative group), using American norms for adult and college-age men

---

2. Chen (1996) contains a typographical error. The mean O5: Ideas scores for men should be 17.17.

Table 1
Characteristics of the Samples

Country	Language	Sample Size (Men/Women)		Source/Reference
		College Age	Adult	
Hong Kong	Chinese	60/62		McCrae, Yik et al., 1998
Taiwan	Chinese	173/371		Chen, 1996
Croatia	Croatian	233/233	123/133	Marušić et al., 1997
The Netherlands	Dutch	615/690		Hoekstra et al., 1996
Belgium	Dutch/Flemish	34/68	527/490	F. DeFruyt
United States	English	148/241	500/500	Costa & McCrae, 1992
South Africa (Blacks)	English	19/46		W. Parker
South Africa (Whites)	English	41/168		W. Parker
Estonia	Estonian	119/398	189/331	J. Allik
The Philippines	Filipino	134/375		G. del Pilar
	English	152/236		A. T. Church
France	French	54/338	279/395	J.P. Rolland; Rolland, 1998
Germany	German	290/454	1185/1801	F. Ostendorf
Indonesia	Indonesian	34/138		L. Halim
Italy	Italian	26/41	315/308	G. V. Caprara
Japan	Japanese	176/177	164/164	Shimonaka et al., 1999
South Korea	Korean (1)	1257/1096		Lee, 1995
	Korean (2)		278/315	R. L. Piedmont
Malaysia	Malay	124/327		Mastor et al., 2000
India	Marathi	107/107		S. Deo
Norway	Norwegian (1)	74/18	397/295	H. Nordvik
	Norwegian (2)		148/210	Ø. Martinsen
Portugal	Portuguese	205/253	606/816	M. P. de Lima
Zimbabwe	Shona	36/35	135/106	R. L. Piedmont
United States	Spanish	24/49		PAR, 1994
Peru	Spanish	274/165		Cassaretto, 1999
Spain	Spanish		89/107	M. Avia
Yugoslavia	Serbian	72/547	256/245	G. Knežević
Russia	Russian	26/91	201/192	T. Martin

and women (Costa & McCrae, 1992). These normed scores are used in all further analyses.[3]

3. Although some correction for age and gender differences seems to be required, there is no reason to think that American norms necessarily provide the best correction. A somewhat less parochial alternative would be to create residual scores (net of age and gender) within this sample of 84 cultures. Analyses of these residual scores showed a similar but not identical pattern of results. Correlations of the aggregated factor scores

Figure 1 provides illustrations of personality profiles for two cultures. The five factor scores are given on the left; scores for the 30 facets, grouped by domain, are given toward the right. The top panel plots means for two different subsamples of adult Norwegian women; the bottom panel, means for two subsamples of college-age Filipino men. One notable feature of Figure 1 is the range of scores. For both these cultures, the great majority of means lie within the *Average* band (relative to American norms). Across all 84 subsamples, this is true of three-quarters of the means. These data demonstrate that there is more variation in personality within cultures than across cultures—a fact that led to the demise of the classic hypotheses of basic or modal personality structures (Singer, 1961). One implication is that personality differences between cultures are likely to be relatively subtle. National stereotypes (Peabody, 1999) may or may not be accurate reflections of such subtle differences.

The two cultures plotted in Figure 1 were chosen because they provide a check on the generalizability of results across different versions of the NEO-PI-R. The two Norwegian subsamples completed two independently made Norwegian translations of the NEO-PI-R; the two Filipino subsamples were tested in different languages (Filipino or English). Yet in both cases, very similar profiles are seen within culture, and clearly different profiles appear across the two cultures. A similar consistency of profiles was seen across languages and raters in a study of Chinese undergraduates (McCrae, Yik et al., 1998).

A more formal way to assess generalizability across subsamples considers correlations across gender and age group. For the gender analyses, means for the five domains for men were matched with means from women of the same age group and culture. The correlations across these 42 pairs of subsamples were .78, .85, .83, .85, and .80, all $p < .001$, for N, E, O, A, and C, respectively. Thus, cultural effects for men closely paralleled those seen in women. For the age group analyses, means for the five domains for college-aged participants were matched with means from adults of the same gender and culture. Because some samples included only college aged participants or adults, only 28 pairs were available. Correlations were .74, .58, .47, .58, and .55, all $p < .05$, for N, E, O, A, and C, respectively. Thus, cultural effects show considerable generalizability across age groups. These findings demonstrate internal

in Table 3 with aggregated residual domain scores across the 26 cultures ranged from .80 for N to .95 for O.
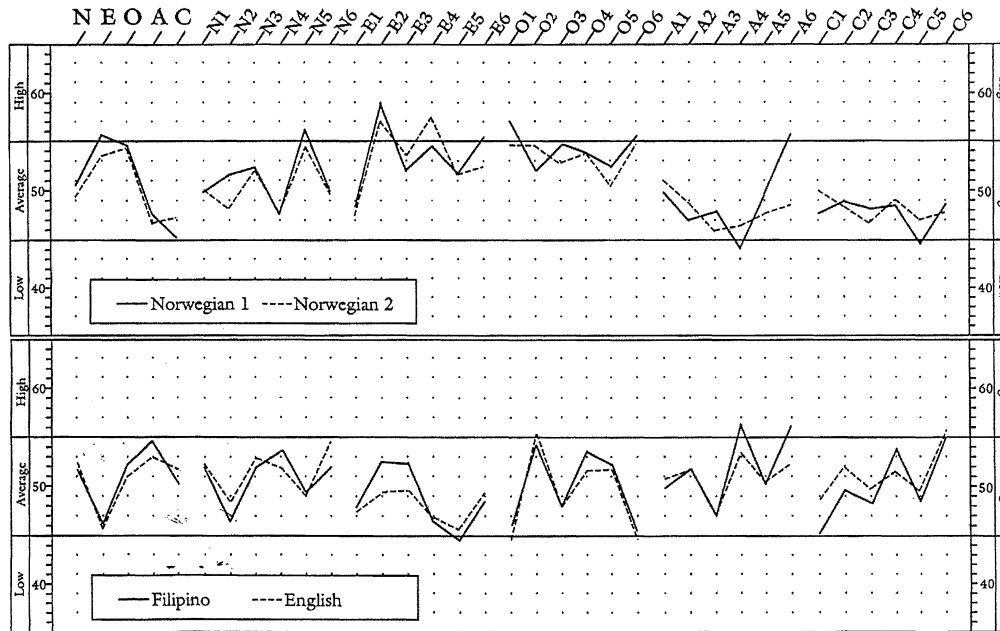
Figure 1

Top panel: Mean NEO-PI-R profiles for adult Norwegian women responding to two independent translations. Bottom panel: Mean NEO-PI-R profiles for college-age Filipino men responding to Filipino vs. English versions. Full labels for facet scales are given in Table 2. Profile form adapted by special permission of the Publisher, from the Revised NEO Personality Inventory, by Paul T. Costa, Jr., and Robert R. McCrae. Copyright 1978, 1985, 1989, 1992 by PAR, Inc. Further reproduction is prohibited without permission of PAR, Inc.

consistency in the pattern of mean values and make it reasonable to combine data from different subsamples to estimate values for the culture as a whole.

Factor Structure

Although ultimately a full understanding of culture and personality will need to examine personality traits at the facet level, for exploratory analyses, it seems preferable to focus on the five broad factors. These factors are approximated by the domain scores, but more precise—and more nearly orthogonal—estimates are given by factor scores. The factor scoring weight matrix is given in the Manual (Costa & McCrae, 1992, Table 2). These weights are based on a factor analysis of the American normative adult sample; here they are applied to the mean T-scores of the 84 subsamples.

This approach is based on the assumption that the five factors are meaningful trait dimensions at the level of cultures. There is to date no evidence in support of that assumption (although Schwartz, 1994, has reported similar structures of values in individuals and cultures). Traits like trust, altruism, and modesty covary in individuals, probably because of shared genetic influences (McCrae, Jang, Livesley, Riemann, & Angleitner, 2001). But there is no a priori reason why they must covary in the same way at the level of cultures. Some cultures might encourage modesty, but discourage trust; some might promote openness to ideas, but inhibit openness to feelings. (Note that such effects would have no influence on the factor structure of individuals within the culture, because individuals would all be equally affected by the cultural influences.) The culture-level factor structure need bear no resemblance to the individual-level factor structure (Bond, in press), but if it does not, then it makes little sense to compare cultures in terms of N, E, O, A, and C.

To resolve this issue, the 30 facets of the NEO-PI-R were factored across the 84 subsamples, with each subsample constituting a single case. (Cattell, 1973, called this design a *combined intercultural* factor analysis—although, at the time, the method had "never been fed enough data to be used," p. 341.) Normally, factor analyses require samples of at least 150 to yield stable results (Guadagnoli & Velicer, 1988); however, the increased reliability of aggregated scores may compensate for the small number of observations (Bond, in press). Principal components analysis showed seven factors with eigenvalues greater than 1.0, but

parallel analysis (Cota, Longman, Holden, Fekken, & Xinaris, 1993) clearly indicated that only five factors should be extracted.

Table 2 reports both Varimax and Procrustes rotations of the five factors (McCrae et al., 1996), which show an unmistakable resemblance to the individual-level structure. In the Varimax solution, the N and E factors are inflated and the O and A factors weak; these divergences from the individual-level structure are substantially corrected in the Procrustes rotation. In the latter solution, 27 of the 30 facets have their highest loading on the expected factor, and secondary loadings resemble those seen in individual-level analyses. Factor congruence coefficients greater than .90 are usually regarded as evidence of replication, although Haven and ten Berge (1977) have argued that .85 is a more reasonable criterion. By that more liberal standard, all factors except O are replicated in the Procrustes solution, and O barely misses the criterion. Further, 22 of the 30 facets show greater-than-chance variable congruence coefficients. Perhaps the most anomalous finding is the loading of .50 for A5: Modesty on the E factor. In individual-level analyses, Modesty usually has a small negative loading on E.

Factor scores calculated from this analysis at the culture level can be correlated with the factor scores based on normative weights from the individual level. Those correlations are factor comparabilities (Everett, 1983), and are presented in the last row of Table 2. These very large values clearly show that the ordering among cultures determined by an analysis of mean values from 26 cultures is virtually identical to what one obtains from an application of the American normative weights from individuals. It appears that data for the present analysis are meaningfully structured and that factor scores can legitimately be used to compare cultures.

Scores for Cultures

Given evidence of the internal consistency and factor structure of scores across cultures, it is possible to create a composite five-factor personality profile for each, averaging across gender and age group. I have treated Black and White South African samples and Hispanic and general American samples as separate cultures. Table 3 lists the mean values for peoples from 26 cultures for the five factor scores. (Note that factor scores for any new group can be added to this list simply by applying American norms and factor scoring weights.) For most factors, the range is about 1 *SD*.

Table 2
Culture-Level Factor Structure of NEO-PI-R Facet Scales: Varimax and Procrustes Rotations

NEO-PI-R Facet Scale	Varimax Factor					Procrustes Factor[a]					VC[b]
	N	E	O	A	C	N	E	O	A	C	
N1: Anxiety	**.84**	.05	.13	.23	−.19	**.89**	.00	−.03	.01	−.11	.99**
N2: Angry Hostility	**.81**	.23	.09	−.35	−.15	**.73**	−.06	−.02	**−.57**	−.07	.99**
N3: Depression	**.83**	−.12	−.11	.13	−.36	**.82**	−.15	−.22	−.03	−.35	.96**
N4: Self-Consciousness	**.71**	**−.40**	−.18	.11	−.10	**.63**	**−.40**	−.34	.10	−.16	.91*
N5: Impulsiveness	.14	**.65**	.08	.19	**−.60**	.32	**.68**	.22	−.14	**−.47**	.87*
N6: Vulnerability	**.74**	−.28	−.09	−.09	**−.42**	**.68**	−.37	−.16	−.14	**−.43**	.95**
E1: Warmth	−.33	**.48**	.28	**.43**	−.01	−.10	**.63**	.34	.27	.11	.97**
E2: Gregariousness	−.11	**.75**	.08	.05	−.11	.00	**.71**	.18	−.24	.00	.86*
E3: Assertiveness	−.26	**.71**	−.06	.02	.00	−.19	**.69**	.06	−.22	.06	.84
E4: Activity	.11	**.84**	.21	.01	.12	.20	**.71**	.20	−.35	.28	.95**
E5: Excitement Seeking	.12	**.84**	−.25	−.07	−.07	.11	**.74**	−.18	**−.45**	−.02	.93*
E6: Positive Emotions	**−.54**	**.63**	.02	.25	−.09	−.38	**.76**	.19	.07	−.02	.91*
O1: Fantasy	.06	**.57**	**.48**	−.11	**−.51**	.23	**.46**	**.61**	−.34	−.31	.94**
O2: Aesthetics	.07	−.05	**.82**	−.20	.04	.18	−.22	**.76**	−.12	.23	.88*
O3: Feelings	−.04	**.57**	**.61**	.21	−.27	.22	**.57**	**.67**	−.02	−.04	.94**
O4: Actions	**−.56**	−.05	.21	.22	−.29	−.39	.14	.37	.33	−.25	.76
O5: Ideas	**−.51**	.02	**.62**	.05	.10	−.34	.05	**.67**	.19	.24	.90*
O6: Values	−.23	**.43**	**.43**	.17	**−.43**	.00	**.48**	**.57**	.03	−.26	.83

Table 2 (cont.)

NEO-PI-R Facet Scale	Varimax Factor					Procrustes Factor[a]					VC[b]
	N	**E**	**O**	**A**	**C**	**N**	**E**	**O**	**A**	**C**	
A1: Trust	**–.61**	–.20	–.02	.39	.00	**–.49**	.07	.08	**.56**	–.04	.95**
A2: Straightforwardness	–.26	**–.47**	.13	**.53**	.13	–.13	–.19	.08	**.73**	.10	.94**
A3: Altruism	–.19	**.50**	.34	**.47**	.35	.01	**.60**	.28	.28	**.48**	.83
A4: Compliance	**–.40**	**–.68**	–.22	.37	.10	–.39	–.39	–.21	**.68**	–.05	.87*
A5: Modesty	.13	.24	–.15	**.74**	–.06	.30	**.50**	–.19	**.52**	–.04	.65
A6: Tender-Mindedness	–.08	–.06	–.07	**.65**	.35	.03	.19	–.18	**.63**	.32	.81
C1: Competence	**–.44**	**.44**	.20	.10	**.55**	–.36	**.44**	.18	.03	**.61**	.94**
C2: Order	–.21	–.17	–.19	.09	**.68**	–.30	–.12	–.31	.20	**.58**	.85*
C3: Dutifulness	–.06	.05	.23	.23	**.77**	–.03	.05	.04	.23	**.80**	.96**
C4: Achievement Striving	–.13	.01	–.01	–.02	**.83**	–.22	–.05	–.18	.02	**.79**	.83
C5: Self-Discipline	**–.45**	.06	.19	.09	**.66**	**–.42**	.08	.12	.18	**.67**	.95**
C6: Deliberation	.11	–.36	–.24	.10	**.72**	–.03	–.34	**–.46**	.22	**.59**	.83
Factor Congruence[c]	.87	.83	.81	.81	.90	.93**	.88**	.84**	.90**	.94**	.90**
Factor Comparability	.94	.93	.83	.78	.97	.97	.96	.94	.97	.97	

Note: These are principal components from 84 subsamples. Loadings greater than |.40| are given in boldface. [a]Targeted to American normative factor structure. [b]Variable congruence coefficients; total congruence coefficient in the last row. [c]Congruence with American normative factor structure. *Congruence higher than that of 95% of rotations from random data. **Congruence higher than that of 99% of rotations from random data.

Table 3
Composite Factor *T*-Scores for Peoples From 26 Cultures

Culture	Factor				
	N	**E**	**O**	**A**	**C**
Hong Kong Chinese	53.3	37.6	49.2	54.6	49.2
Taiwan Chinese	51.5	42.0	50.2	54.5	48.1
Croatians	52.8	45.1	49.0	47.5	53.2
Dutch	48.6	43.9	55.7	54.6	48.6
Belgians	53.0	47.7	51.8	50.0	46.6
Americans	50.0	50.0	50.0	50.0	50.0
South African Blacks	49.1	41.4	47.7	50.4	47.9
South African Whites	51.9	47.2	54.4	52.2	47.9
Estonians	49.7	49.9	52.6	50.8	49.6
Filipinos	50.8	43.8	51.8	52.9	51.5
French	55.4	47.3	54.1	52.1	47.4
Germans	52.8	47.3	56.7	49.1	46.7
Indonesians	48.6	43.3	49.9	51.9	50.3
Italians	55.6	46.6	52.6	48.9	50.4
Japanese	55.3	41.7	51.7	47.7	42.6
South Koreans	53.6	40.0	51.4	52.3	48.8
Malaysians	54.2	42.5	46.6	58.5	54.2
Indians (Marathi)	49.1	40.7	51.4	56.7	55.7
Norwegians	47.4	53.6	51.5	49.9	45.7
Portuguese	55.5	46.3	49.2	51.2	50.3
Zimbabweans	50.9	42.3	47.0	51.0	51.8
Hispanic Americans	49.5	47.5	51.2	47.1	51.6
Peruvians	50.8	45.5	50.0	48.6	49.0
Spaniards	57.1	48.3	48.0	49.4	48.3
Yugoslavians	51.1	47.6	56.0	48.4	51.7
Russians	53.6	45.2	49.1	46.7	46.5

Is it possible to validate these factor scores? In the case of N and E factors, comparisons can be made to the mean scores reported by Barrett and S. Eysenck (1984) and Lynn and Martin (1995) on the EPQ. Matches were found with 15 of the cultures in the present study. The correlation of EPQ N with the NEO-PI-R N factor was $r = .74$, $p < .01$, but EPQ E was not significantly related to NEO-PI-R E ($r = .21$, *n.s.*). Further, the NEO-PI-R C factor was positively (and unexpectedly) related to both EPQ E and EPQ P. Scatterplots revealed that the Indian EPQ data

appeared to be anomalous; when that sample was excluded, convergent correlations for both N and E were significant (rs = .80 and .61, respectively, $N = 14$, $p < .05$), whereas none of the other EPQ/NEO-PI-R correlations reached significance.

Other culture-level criteria can also be examined. Diener, Diener, and Diener (1995) ranked 55 cultures with respect to mean subjective well-being, a variable inversely related to N and directly related to E in individual analyses of Western samples (Costa & McCrae, 1980). Twenty of the present samples could be matched with those in the Diener et al. (1995) list (e.g., Russia with the U.S.S.R., Hong Kong with China), and the correlations of N and E with subjective well-being were -.35 and .64, respectively. Both correlations are in the hypothesized direction, although in this small sample only the latter attains statistical significance ($p < .01$). (Analyzed at the level of subsamples, these correlations are -.32 and .56, $N = 68$, $p < .01$.) O, A, and C were unrelated to subjective well-being.

Probably the best-known dimensions of culture are those identified by Hofstede (1991): Power Distance, Uncertainty Avoidance, Individualism, and Masculinity. If Black South Africa and Zimbabwe are considered as East African cultures,[4] then Hofstede's ratings are available for 23 of the 26 cultures (Hofstede, 1994; Peabody, 1999). Table 4 reports correlations between Hofstede's ratings and the factor scores for the 26 cultures. Considering that Hofstede's dimensions originated in ratings of work-related values among employees of a multinational corporation, there is remarkable convergence between the two sets of variables. Power Distance, which reflects an acceptance of status differences, is highest in cultures whose members are conscientious and introverted—obedient and docile workers who prefer the order and formality that hierarchical social structures provide. Uncertainty Avoidance describes cultures that seek to minimize the threat of ambiguous situations, and one might guess that it would be inversely related to O. But the data suggest that it is neurotic and antagonistic cultures that are most sensitive to threat. Hofstede himself (1991; cited in Peabody, 1999, p. 77) described nations

---

4. A reviewer questioned whether it was appropriate to identify Russia with the U.S.S.R., Hong Kong with China, and South Africa with East Africa. Analyses of SWB and Hofstede dimensions were therefore repeated, omitting these questionable matches. Numerical differences were small, and there were no changes in the pattern of significant results.

Table 4
Correlations Between Hofstede (1991) Dimensions
and Personality Factors

			Factor		
Hofstede Dimension	**N**	**E**	**O**	**A**	**C**
Power Distance	.11	−.43*	−.35	.38	.63**
Uncertainty Avoidance	.55**	.27	.19	−.61**	−.34
Individualism	.02	.61**	.48*	−.16	−.31
Masculinity	.39	−.25	.09	−.14	−.13
GDP[a]	.16	.27	.20	−.12	−.42*

Note. N = 23 cultures for Hofstede dimensions, 26 for GDP. [a]GDP = per capita Gross Domestic Product, 1998 estimate (Hong Kong estimate separate from China). Source: www.infoplease.com/countries.html.
*$p < .05$. **$p < .01$.

high in Uncertainty Avoidance as "anxious cultures" and characterizes their members as "fidgety, emotional, aggressive"—traits associated with N and low A. The association of N with Uncertainty Avoidance replicates a finding reported by Lynn and Martin (1995), who analyzed the Eysenck N scale. Individualism is associated with both E and O: Self-assertive and free-thinking people are found in cultures that value the individual over the group. Masculinity is unrelated to any of the personality factors in the 23 cultures, which is perhaps not surprising because, on the individual level, masculinity is not a clear definer of any of the five factors (Saucier & Goldberg, 1998).[5]

Power Distance in this sample, as in most, is inversely related to Individualism (r = -.61), and both these variables are related to per capita Gross Domestic Product (GDP; rs = -.63, .66). The last row of Table 4 shows that GDP itself is inversely correlated with C, but is unrelated to the other factors. When GDP is partialled from the correlations between Hofstede dimensions and personality factors, the correlations of Power Distance with E and C (rs = -.31 and .41, respectively) are nonsignificant,

---

5. Only 13 of the present list of cultures overlapped with those rated for values by Schwartz (1994). Trends in the data suggested that E and O were associated with Egalitarian Commitment and Harmony (including *helpful, freedom, protecting environment,* and *world of beauty* values), whereas A and C were associated with Conservatism (including *politeness, forgiving, self-discipline,* and *obedient*).

but there are no other changes in the patterns of significant findings. It appears that most of these personality/culture associations are not simply a matter of economics.

Finally, we may ask if the scores in Table 3 square with perceptions of national character. It is certainly possible to find post hoc explanations for many of the results. The moody and melancholic temperament of Spaniards matches their high N scores. Germany, the land of *Dichter und Denker*, is home to high O scores. Malaysians are reputed to be polite and nonconfrontational (Mastor et al., 2000), as their high A scores suggest. However, it is surprising that the industrious Japanese score so low in C, and that Hispanic Americans, for whom *simpatía* is a core value (Triandis, Marín, Lisansky, & Betancourt, 1984), score so low in A.

It would be possible to evaluate all the scores in Table 3 by reviewing the published literature. For example, the relatively high A scores of Filipinos are consistent with the emphasis on smooth interpersonal relations in Filipino culture, and their low E scores are consistent with low needs for exhibitionism and affiliation (see Church, 1987). But to gain an overall evaluation of the extent to which results in Table 3 are consistent with perceptions of national character, I turned to a panel of experts (see Author Note). On five separate sheets of paper presented in random order, I listed the seven lowest and seven highest scoring cultures for each factor. I asked the experts to identify the personality factor that had been used to rank the cultures on each page. Which personality factor, for example, is lowest among Hong Kong Chinese and South Koreans, but highest among Norwegians and Americans? Judges were instructed not to consult the literature, but simply to rely on their own impressions.

The judges were in near perfect agreement that this task was exceedingly difficult. Several declined even to offer guesses, and of the 25 judgments that were made, only 4 were correct—about what one would expect by chance. In view of the internal consistency and external correlates of the factors, these findings are puzzling. It is possible that the scores, and thus the rankings, of the cultures are seriously distorted by response biases, differing standards of self-presentation, or quirks of sampling.[6] It is possible that the judges were not sufficiently familiar with

6. Peter Smith suggested that scores might be distorted by cultural differences in social desirability response bias, and he recommended a reanalysis. The mean of the desirable poles of the five factors (i.e., with N reverse scored as Emotional Stability) was taken as an index of bias and used to adjust factor scores so that all cultures were equated on

the full range of cultures to make informed judgments. It is also possible that human judges, even experts, are relatively insensitive to differences in average levels of personality traits in large groups and that mean trait differences are not reflected in obvious ways in social institutions and customs. One implication of the last interpretation is that intercultural analyses may uncover many new and hitherto unsuspected findings because, like microscopes, they operate at a level of analysis inaccessible to the "naked eye." Another is that national stereotypes—already regarded with suspicion by many social scientists—are unlikely to be good estimates of mean levels of personality traits.

Some Tentative Interpretations

The present analyses were frankly exploratory; it was not clear beforehand that any meaningful results would be found at all. Samples might have been completely unrepresentative of their culture; translations might have rendered scores incommensurable across cultures; personality factors found within cultures might have disappeared when examined across cultures; scores might have no interpretable pattern of external correlates. But as it turned out, results were consistent and relatively clear. Age and gender differences resembled those found in American samples; different subsamples from each culture showed similar levels of personality traits; intercultural factor analysis yielded a close approximation to the Five-Factor Model (FFM); and factor scores were meaningfully related to other culture-level variables. These findings encourage intercultural analyses even when ideal conditions for data interpretation are not met.

However, the resulting scores apparently bore little relation to perceptions of national character. It was not obvious to expert judges that Indians and Malaysians were, on average, more conscientious than Japanese and Norwegians, nor that Norwegians and Americans were more extraverted than Hong Kong Chinese and South Koreans. Perhaps this simply implies that these relatively subtle group differences in personality cannot be detected by informal observation. Alternatively, this may be reason to doubt the veridicality of the factor scores.

overall desirability. Adjustments were small, suggesting that desirability played a limited role in determining culture scores. Adjusted scores correlated from .86 to .93 with the raw scores, and had little effect on the ranking of cultures.

Cultural differences in response styles have been proposed as a major source of bias in cross-cultural comparisons (Cheung & Rensvold, 2000). Acquiescence is not relevant here, because all versions of the NEO-PI-R have balanced keying, and extreme responding would affect *SD*s rather than means. If social desirability were the major determinant of responses, we would expect a single, favorable versus unfavorable factor instead of the five substantive factors that were found.

However, other biases are possible. For example, Hong Kong Chinese have higher standards of competence than Americans, and judged by these standards, rate themselves lower (McCrae, Yik et al., 1998). Clearly, alternative ways of assessing personality—such as ratings by observers from other cultures—would be helpful in assessing the veridicality of these self-report-based means.

Interpreting Cultural Effects and Associations

Why should two cultures differ in the mean level of a personality trait? Anthropologists would likely assume that cultural institutions or practices encouraged or inhibited the trait. Behavior geneticists, however, might suggest that differences in gene pools could equally account for differences in trait phenotypes. Acculturation studies (McCrae, Yik et al., 1998) are particularly useful in sorting out these possibilities because individuals from the same gene pool are studied under different cultural conditions.

Associations between personality traits and features of culture are thus subject to two basic causal interpretations: Culture may shape personality, or personality may shape culture. These possibilities can be illustrated by considering the correlations of personality factors with Hofstede's dimensions reported in Table 4. Power Distance, for example, might exemplify the first causal route. In societies with high Power Distance there is a heavy institutional emphasis on status and authority: Some people give orders, some take them. As in military settings (cf. Ottati, Triandis, & Hui, 1999), the outcome may well be order and discipline, leading to a population high in C.

The association of N with Uncertainty Avoidance might exemplify the opposite causal pattern. If many members of a group are by temperament particularly prone to anxiety and anger, the group may have evolved social structures to minimize stress, just as psychiatric hospitals seek to provide a safe and predictable environment for their patients. This

interpretation seems more plausible than the converse argument that rules and routines somehow lead to elevations in N.

How should the association of E with Individualism be interpreted? Extraverts choose to interact intensively with others, and it might seem that they would be found in collectivistic cultures. But in fact, collectivism is associated with close interaction only with one's ingroup; it is "people in individualist cultures [who] are very good at meeting outsiders, forming new ingroups, and getting along with new people" (Triandis, Bontempo, Villareal, Asai, & Lucca, 1988, p. 324). This suggests the hypothesis that collectivism is an institutional compensation for low levels of E. Social interaction is essential for the survival of every culture; if people are not temperamentally inclined to interact, then perhaps collectivistic institutions must arise to guide them in social directions.

It must be recalled that Hofstede's dimensions, although interpreted (and validated) in terms of social structural variables, were derived from analyses across cultures of mean individual responses within cultures, and thus represent cultural patterns inferred from the values of individuals. Perhaps more interesting would be correlations of mean personality levels with dimensions of ethos, derived from analyses of the beliefs, institutions, art, and myths of different cultures. At a more sociological level, economic, political, and social variables might be related to national personality profiles (Bond, in press). Although there is at present no generally accepted taxonomy of culture-level variables (but see Georgas & Berry, 1995, for a proposal), there are many individual features of culture whose relation to mean personality levels could profitably be investigated.

CONCLUSION

Nations have been characterized throughout history in the language of personality traits, and anthropologists and psychoanalysts offered qualitative analyses of personality and culture in the first half of the last century. But quantitative data on personality traits across cultures have only become available in the past 50 years, most in the past decade. We have only begun to be able to address scientifically some of our oldest questions about humankind and society.

We know already that personality traits are in many respects universal, showing the same structure and adult development in widely different cultures. The present article addressed a different question: How and why

do cultures differ in mean levels of personality traits? The exploratory analyses offered here surely raised more questions than they answered. The data appeared to be consistent and meaningful by several criteria, but the resulting profiles were by no means intuitively obvious. Are the factor scores examined here veridical or biased? Would aggregation of judgments of national character across many judges (Peabody, 1999) converge with personality factor scores? With what features of culture, if any, are mean personality scores associated? Do national differences in the levels of personality traits have genetic or environmental explanations? Why is it that almost identical five-factor structures emerge in culture-level and individual-level analyses of personality traits?

Researchers who wish to characterize cultures in terms of personality profiles have both an ethical and a scientific obligation to be as accurate as they can, and the recommendations of methodologists (e.g., van de Vijver & Leung, 1997) should be followed as closely as possible. But lapses from an ideal design should not inhibit thoughtful exploration of available data. The more we learn, the sooner we will discover our errors.

One of the errors we ought to avoid is the disparaging attitude toward anthropologists and their methods seen in the opening quotation of this article (Cattell, 1950). Ethnographic methods are probably not the ideal way to assess mean levels of personality traits, but they can yield rich insights into the operation and significance of personality traits in a particular cultural context, confirming or calling into question the results of standardized questionnaires. Personality and culture is an interdisciplinary field; without the analyses of culture provided by anthropologists and other social scientists, there would be nothing to which to relate trait levels.

Conversely, it is time for all the social sciences to acknowledge the role of personality traits in human affairs. N, E, O, A, and C are not inventions of Western psychologists; they are part of human nature— dimensions of enduring dispositions that somehow find expression in every culture. How group differences in the levels of these traits are related to group processes and outcomes is as yet largely unknown, but it is a topic that should be of great interest to all social scientists.

REFERENCES

Avia, M. D., Sanz, J., & Sánchez-Bernardos, M. L. (1999). *Version Espanola del Inventario de Personalidad NEO Revisado (NEO-PI-R) de Paul T. Costa y Robert R. McCrae (1992). Edicion Experimental.* Unpublished research report, University Complutense de Madrid.

Barrett, P., & Eysenck, S. (1984). The assessment of personality factors across 25 countries. *Personality and Individual Differences,* **5,** 615–632.

Benet-Martínez, V., & John, O. P. (2000). Measuring *Los Cinco Grandes* in Spain with indigenous Castilian markers. *American Behavioral Scientist,* **44,** 141–157.

Bock, P. K. (2000). Culture and personality revisited. *American Behavioral Scientist,* **44,** 32–40.

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

Butcher, J. N. (Ed.). (1996). *International adaptations of the MMPI-2: A handbook of research and clinical applications.* Minneapolis: University of Minnesota Press.

Cassaretto, M. (1999). *Adaptacion del Inventario de personalidad NEO Revisado (NEO-PI-R) Forma S en un grupo de estudiantes universitarios [Adaptation of the Revised NEO Personality Inventory in a sample of university students].* Unpublished thesis, Pontificia Universidad Catolica del Peru, Lima, Peru.

Cattell, R. B. (1950). *Personality: A systematic theoretical and factual study.* New York: McGraw-Hill.

Cattell, R. B. (1973). *Personality and mood by questionnaire.* San Francisco: Jossey-Bass.

Chen, M. C. (1996). *Psychosocial correlates of prosocial behavior among college students in Taiwan.* Unpublished doctoral dissertation, Loyola College in Maryland.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology,* **31,** 187–212.

Church, A. T. (1987). Personality research in a non-Western culture: The Philippines. *Psychological Bulletin,* **102**, 272–292.

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

Costa, P. T., Jr., & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology,* **38,** 668–678.

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., & Jónsson, F. H. (in press). Validity and utility of the Revised NEO Personality Inventory: Examples from Europe. In B. DeRaad & M. Perugini (Eds.), *Big Five assessment.* Göttingen, Germany: Hogrefe & Huber.

Costa, P. T., Jr., McCrae, R. R., Martin, T. A., Oryol, V. E., Senin, I. G., Rukavishnikov, A. A., Shimonaka, Y., Nakazato, K., Gondo, Y., Takayama, M., Allik, J., Kallasmaa, T., & Realo, A. (2000). Personality development from adolescence through adulthood:

Further cross-cultural comparisons of age differences. In V. J. Molfese & D. Molfese (Eds.), *Temperament and personality development across the life span* (pp. 235–252). Hillsdale, NJ: Erlbaum.

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. S., Holden, R. R., Fekken, G. C., & Xinaris, S. (1993). Interpolating 95th percentile eigenvalues from random data: An empirical example. *Educational and Psychological Measurement,* **53,** 585–596.

De Fruyt, F., & Mervielde, I. (1999). RIASEC types and Big Five traits as predictors of employment status and nature of employment. *Personnel Psychology,* **52,** 701–727.

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

Everett, J. E. (1983). Factor comparability as a means of determining the number of factors and their rotation. *Multivariate Behavioral Research,* **18,** 197–218.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire.* London: Hodder and Stoughton.

Georgas, J., & Berry, J. W. (1995). An ecocultural taxonomy for cross-cultural psychology. *Cross-Cultural Research: The Journal of Comparative Social Science,* **29,** 121–157.

Grimm, S. D., & Church, A. T. (1999). A cross-cultural study of response biases in personality measures. *Journal of Research in Personality,* **33**, 415–441.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin,* **103,** 265–275.

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

Heuchert, J. W. P., Parker, W. D., Stumpf, H., & Myburgh, C. P. H. (2000). The Five-Factor Model of personality in South African college students. *American Behavioral Scientist,* **44,** 112–125.

Hoekstra, H. A., & De Fruyt, F. (in press). *Een aanvullende normering van de NEO-PI-R op een bevolkingssteekproef [NEO-PI-R normative population study].* Lisse: Swets & Zeitlinger.

Hoekstra, H. A., Ormel, J., & De Fruyt, F. (1996). *Handleiding NEO Persoonlijkheidsvragenlijsten NEO-PI-R en NEO-FFI [Manual for NEO Personality Inventories NEO-PI-R and NEO-FFI].* Lisse, The Netherlands: Swets & Zeitlinger.

Hofstede, G. (1991). *Cultures and organizations: Software of the mind.* London: McGraw-Hill.

Hofstede, G. (1994). Images of Europe. *Netherlands Journal of Social Sciences,* **30,** 63–82.

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.

Kallasmaa, T., Allik, J., Realo, A., & McCrae, R. R. (2000). The Estonian version of the NEO-PI-R: An examination of universal and culture-specific aspects of the Five-Factor Model. *European Journal of Personality,* **14,** 265–278.

Katigbak, M. S., Church, A. T., Akamine, T. X. (1996). Cross-cultural generalizability of personality dimensions: Relating indigenous and imported dimensions in two cultures. *Journal of Personality and Social Psychology,* **70,** 99–114.

Knežević, G., Radović, B., & Opačić, G. (1997). An evaluation of the "Big Five Model" of personality through an analysis of the NEO-PI-R Personality Inventory. *Psihologija,* **1-2,** 3–24.

Lee, K.-I. (1995). *Factor structure and maladaptive group profiles of the Revised NEO Personality Inventory for Koreans.* Unpublished doctoral dissertation, Pusan National University.

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

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

Marušić, I., Bratko, D., & Eterović, H. (1997). A contribution to the cross-cultural replicability of the five-factor personality model. *Review of Psychology,* **3,** 23–35.

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

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

McCrae, R. R., & Costa, P. T., Jr. (1999). A Five-Factor Theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 139–153). New York: Guilford.

McCrae, R. R., Costa, P. T., Jr., del Pilar, G. H., Rolland, J. P., & Parker, W. D. (1998). Cross-cultural assessment of the Five-Factor Model: The Revised NEO Personality Inventory. *Journal of Cross-Cultural Psychology,* **29,** 171–188.

McCrae, R. R., Costa, P. T., Jr., Lima, M. P. d., Simões, A., Ostendorf, F., Angleitner, A., Marušić, I., Bratko, D., Caprara, G. V., Barbaranelli, C., Chae, J.-H., & Piedmont, R. L. (1999). Age differences in personality across the adult life span: Parallels in five cultures. *Developmental Psychology,* **35,** 466–477.

McCrae, R. R., Costa, P. T., Jr., Ostendorf, F., Angleitner, A., Hřebríčková, M., Avia, M. D., Sanz, J., Sánchez-Bernardos, M. L., Kusdil, M. E., Woodfield, R., Saunders, P. R., & Smith, P. B. (2000). Nature over nurture: Temperament, personality, and lifespan development. *Journal of Personality and Social Psychology,* **78,** 173–186.

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

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

Ottati, V., Triandis, H. C., & Hui, C. H. (1999). Subjective culture and the workplace: Hispanic and mainstream naval recruits. In Y.-T. Lee, C. R. McCauley, & J. G. Draguns (Eds.), *Personality and person perception across cultures* (pp. 235–253). Mahwah, NJ: Erlbaum.

Paunonen, S. V., Jackson, D. N., Trzebinski, J., & Forsterling, F. (1992). Personality structure across cultures: A multi-method evaluation. *Journal of Personality and Social Psychology,* **62,** 457–473.

Peabody, D. (1999). National characteristics: Dimensions for comparison. In Y.-T. Lee, C. R. McCauley, & J. G. Draguns (Eds.), *Personality and person perception across cultures* (pp. 65–100). Mahwah, NJ: Erlbaum.

Piedmont, R. L., Bain, E., McCrae, R. R., & Costa, P. T., Jr. (in press). The applicability of the Five-Factor Model in a sub-Saharan culture: The NEO-PI-R in Shona. In R. R. McCrae & J. Allik (Eds.), *The Five-Factor Model across cultures*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Piedmont, R. L., & Chae, J. H. (1997). Cross-cultural generalizability of the five-factor model of personality: Development and validation of the NEO-PI-R for Koreans. *Journal of Cross-Cultural Psychology,* **28,** 131–155.

Psychological Assessment Resources. (1994). *The Revised NEO Personality Inventory: Manual Supplement for the Spanish Edition.* Odessa, FL: Author.

Rolland, J. P. (1998). *NEO-PI-R Inventaire de Personnalité-Révisé: Manuel [Revised NEO Personality Inventory: Manual].* Paris: Les Editions du Centre de Psychologie Appliquée.

Rolland, J. P., Parker, W. D., & Stumpf, H. (1998). A psychometric examination of the French translations of the NEO-PI-R and NEO-FFI. *Journal of Personality Assessment,* **71,** 269–291.

Saucier, G., & Goldberg, L. R. (1998). What is beyond the Big Five? *Journal of Personality,* **66,** 495–524.

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

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

Singer, M. (1961). A survey of culture and personality theory and research. In B. Kaplan (Ed.), *Studying personality cross-culturally* (pp. 9–90). Evanston, IL: Row, Peterson.

Triandis, H. C., Bontempo, R., Villareal, M. J., Asai, M., & Lucca, N. (1988). Individualism and collectivism: Cross-cultural perspectives on self-ingroup relationships. *Journal of Personality and Social Psychology,* **54,** 323–338.

Triandis, H. C., Marín, G., Lisansky, J., & Betancourt, H. (1984). *Simpatía* as a cultural script of Hispanics. *Journal of Personality and Social Psychology,* **47,** 1363–1375.

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.