The Validity and Structure of Culture-Level Personality Scores: Data From Ratings of Young Adolescents

Robert R. McCrae,[1] Antonio Terracciano,[1] Filip De Fruyt,[2] Marleen De Bolle,[2] Michele J. Gelfand,[3] Paul T. Costa, Jr.,[1] and 42 Collaborators of the Adolescent Personality Profiles of Cultures Project

[1]National Institute on Aging
[2]Ghent University
[3]University of Maryland

ABSTRACT   We examined properties of culture-level personality traits in ratings of targets ($N = 5,109$) ages 12 to 17 in 24 cultures.

Aggregate scores were generalizable across gender, age, and relationship groups and showed convergence with culture-level scores from previous studies of self-reports and observer ratings of adults, but they were unrelated to national character stereotypes. Trait profiles also showed cross-study agreement within most cultures, 8 of which had not previously been studied. Multidimensional scaling showed that Western and non-Western cultures clustered along a dimension related to Extraversion. A culture-level factor analysis replicated earlier findings of a broad Extraversion factor but generally resembled the factor structure found in individuals. Continued analysis of aggregate personality scores is warranted.

The idea that the citizens of different nations have distinctive personalities can be traced to antiquity, and it was a central tenet of early 20th century culture and personality studies (LeVine, 2001). For a number of reasons, including the declining influence of psychoanalysis and ethical concerns about ethnocentrism (see Church, 2001), the topic fell out of favor, and interest has only recently been revived, this time from the perspective of trait psychology (Lynn & Martin, 1995; McCrae, Terracciano, & 79 Members of the Personality Profiles of Cultures Project, 2005; Schmitt et al., 2007). In this new approach, personality profiles of cultures can be obtained by averaging traits assessed in a sample of culture members, yielding a

Jean-Pierre Rolland, Université Paris Ouest Nanterre La Défense, France; Vanina Schmidt, University of Buenos Aires, Argentina; Andrzej Sekowski, The John Paul II Catholic University of Lublin, Poland; Jane Shakespeare-Finch, Queensland University of Technology, Australia; Yoshiko Shimonaka, Bunkyo Gakuin University, Japan; Franco Simonetti, Pontificia Universidad Catolica De Chile, Chile; Jerzy Siuta, Jagiellonian University, Poland; Barbara Szmigielska, Jagiellonian University, Poland; Vitanya Vanno, Srinakharinwirot University, Thailand; Lei Wang, Peking University, People's Republic of China; Michelle Yik, The Hong Kong University of Science and Technology, Hong Kong.

Correspondence concerning this article should be addressed to Robert R. McCrae, 809 Evesham Avenue, Baltimore, MD 21212. E-mail: RRMcCrae@gmail.com.

set of aggregate personality traits. This is an etic approach, in which the same set of traits (usually identified in one culture) are studied across a range of cultures.

The validity of these culture-level scores must be established, and there are at least two reasons to be skeptical about their accuracy. The first is that the personality trait scales that are aggregated may not themselves be commensurable across cultures: They may assess different constructs in different cultural contexts, or they may lack scalar equivalence (Nye, Roberts, Saucier, & Zhou, 2008; van de Vijver & Leung, 1997) due to problems in translation, in the relevance of particular items, or to cultural differences in response styles. These are theoretical threats to the validity of all cross-cultural measures.

The second reason to doubt the validity of aggregate personality scores is that research to date suggests that they do not correspond to national character stereotypes (Perugini & Richetin, 2007). It is widely believed, for example, that the English are reserved—yet their aggregate personality scores suggest that they are in fact quite extraverted (McCrae, Terracciano, & 79 Members, 2005). This finding is not a fluke; analyses of data from 49 cultures suggested that national stereotypes are almost completely unrelated to aggregate personality traits (Terracciano et al., 2005). Many stereotypes have at least a kernel of truth (Madon et al., 1998), so the failure to find any association of national character stereotypes with aggregate personality scores is a legitimate source of concern.

National character data from the Personality Profiles of Cultures (PPOC) project used by Terracciano and colleagues (2005)—and reanalyzed in the present article—were obtained by asking raters in each culture to describe the typical member of their own culture. Such judgments are sometimes called *autostereotypes*, in contrast to the *heterostereotypes* held by members of one culture about members of another. Several studies, however, have shown general agreement between these two kinds of stereotypes (Boster & Maltseva, 2006; Peabody, 1985). People around the world think that Americans are assertive and arrogant, and so do Americans (Terracciano & McCrae, 2007). Thus, the apparent inaccuracy of national character stereotypes is unlikely to be the result of ethnocentric or ethnophobic biases or of the way national character stereotypes were assessed.

It is logically possible that both stereotypes and aggregated scores are invalid, but if forced to choose between them, researchers must rely on patterns of supporting evidence. Heine, Buchtel, and

Norenzayan (2008), for example, showed that per capita Gross Domestic Product (GDP) is better predicted by stereotypes of Conscientiousness than by aggregate Conscientiousness scores. But this evidence is ambiguous, because in stereotypic thinking, industriousness is generally (mis)attributed to the wealthy (Fiske, Cuddy, Glick, & Xu, 2002), by a kind of variant of the fundamental attribution error. The weight of evidence to date favors the view that aggregate scores are accurate and national stereotypes are not (McCrae, Terracciano, Realo, & Allik, 2007b), largely because national stereotypes do not make psychological sense as indicators of national trait levels. For example, climate is one of the strongest correlates of national stereotypes of interpersonal warmth (McCrae, Terracciano, Realo, & Allik, 2007a), though few personality psychologists today believe that ambient temperature is a powerful influence on personality development. Stereotypes also fail to obey simple mathematical laws: The stereotype of Italians is not the mean of the stereotype of Northern and Southern Italians, but is almost identical with the latter (et al., 2007a).

A number of cross-cultural methodologists (see Nye et al., 2008) have argued that the scalar equivalence of test items across cultures must be established before mean level comparisons are made—a strategy McCrae, Terracciano, and 79 Members (2005) labeled *bottom-up*. In contrast, McCrae and colleagues advocated a *top-down* strategy in which the construct validity of aggregate scores is examined directly. There is some support for the convergent validity of aggregate personality scores (e.g., Oishi & Roth, 2009), but it is still limited. Rentfrow, Gosling, and Potter (2008) provided validity data on aggregate personality scores for U.S. states, although those data do not address the difficulties posed by translation and cultural variations in response styles. McCrae, Terracciano, and 79 Members (2005) correlated culture-level scores from studies of self-reported personality traits with scores from observer-rated traits across 28 cultures. They found significant agreement for three (Neuroticism, Extraversion, and Openness) of the five factors and 26 of 30 facets of the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992). Analyzed as profile agreement across the 30 facets within each culture, significant agreement was found for 22 of the 28 cultures. Aggregate personality scores also showed evidence of construct validity in their prediction of Hofstede's (2001) dimensions of culture (Hofstede & McCrae, 2004) and in their geographical

patterns (Allik & McCrae, 2004; McCrae, Terracciano, & 79 Members, 2005), in which Western cultures tended to cluster together in contrast to non-Western cultures. Using a different measure of personality, Schmitt and colleagues (2007) reported significant convergent validity between NEO-PI-R factor scores and Big Five Inventory (BFI) scales (John, Donahue, & Kentle, 1991) for three of the factors (Neuroticism, Extraversion, and Conscientiousness) across 27 cultures. (Discriminant validity was more problematic.)

Persuasive evidence of the validity of culture-level aggregate personality scores would have important consequences for cross-cultural psychology. First, it would provide researchers with relatively accurate accounts of the prevailing personality traits in a variety of cultures, scores that might be used to predict a variety of nation-level outcomes of interest (McCrae & Terracciano, 2008). Second, it would reinforce the conclusion that national character stereotypes are almost completely unfounded—an observation with consequences both for the psychology of stereotypes and for the practice of international relations. Third, it would imply that the many theoretical concerns—potential threats to scalar equivalence—that have been raised about cross-cultural comparisons may have limited applicability in real-world data, and thus these concerns may have had an unwarranted chilling effect on mean comparisons in cross-cultural research. Certainly, every cross-cultural researcher must continue to be vigilant against artifactual explanations of apparent cultural differences, but the validity of aggregate personality traits would serve as an encouragement to study such differences.

With so much at stake, further evidence on the validity of aggregate personality traits is surely needed. The present article reports new data from the Adolescent Personality Profiles of Cultures (APPOC) Project, in which aggregate personality traits are scored from observer ratings of adolescents aged 12 to 17 in a sample of 24 cultures. Although this is a relatively small number, it includes 8 cultures (Argentina, Australia, Chile, Islamic Republic of Iran, Puerto Rico, Slovakia, Thailand, and Uganda) not previously included in culture-level studies of the validity of personality profiles.

In studies of personality at the individual level, factor replication is an aspect of construct validity: If scales retain their validity in translation (and if the structure of personality is universal), then the same factor structure should emerge within each culture—as, for the most part, it does in analyses of the NEO-PI-R (McCrae,

Terracciano, & 78 Members, 2005) and in world regional analyses of the BFI (Schmitt et al., 2007). However, replication of the individual-level factor structure at the culture level is not necessarily required, because the structure of personality may vary across levels of analysis. Previous research on the culture-level structure of the NEO-PI-R (McCrae, 2002; McCrae, Terracciano, & 79 Members, 2005) has suggested that the individual-level Five-Factor Model (FFM) is approximately replicated, but that the Extraversion factor is expanded to include aspects of other factors, including Impulsiveness, Openness to Fantasy and Values, and Competence—characteristics that appear to be higher in wealthier and more extraverted cultures. The present study provides an opportunity to replicate this culture-level finding.

As a general rule, the analysis of aggregate scores ought to reproduce the individual level structure, unless there are specific effects on structure due to culture (J. Allik, personal communication, August 10, 2004; McCrae & Terracciano, 2008). The present study uses data on college students' perceptions of adolescents ages 12 to 17, and previous analyses of these data at the individual level (De Fruyt et al., 2009) suggest one deviation from the universal adult factor structure: Openness to Ideas shows a substantial loading on Conscientiousness, perhaps because both diligence and an interest in ideas are attributed to adolescents who are known to be good students. It might therefore be hypothesized that a culture-level factor analysis of these adolescent data will show that aggregate Openness to Ideas loads on the Conscientiousness factor as well as the Openness factor.

METHOD

Procedure

As detailed elsewhere (De Fruyt et al., 2009), collaborators from 27 sites representing 18 different languages from 24 cultures provided data. Ratings from multiple sites were available for the United States (three collaborating sites) and Poland (two collaborating sites). Collaborators were asked to collect anonymous observer ratings from college students who were randomly assigned one of four targets: a boy or girl ages 12 to 14 or 15 to 17 years. College student ratings were used instead of self-reports from adolescents for several reasons (convenience, data quality, comparability to PPOC data), but American studies (Costa, McCrae, & Martin,

2008; McCrae, Costa, & Martin, 2005) suggest that self-reports from adolescents would likely yield similar data. Collaborators were asked to provide data on 50 targets in each category.

Participants received the following general instructions (cf. McCrae, Terracciano, & 78 Members, 2005): "This is a study of personality across cultures. We are interested in how people view others and rate their personality traits, and we will be comparing your responses to those of college students in other countries. Please think of a *boy* [*girl*] *aged 12–14* [*15–17*] whom you know well. He [She] should be someone who is a native-born citizen of your country. He [She] can be a relative or a friend or neighbor—someone you like or someone you don't like." Valid ratings were obtained for 5,109 targets.

Measures

The NEO-PI-R (Costa & McCrae, 1992) is among the most frequently used inventories to assess the FFM and its dimensions of Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. The inventory has 30 facets, organized under the five domains, and includes 240 items (8 items per facet), presented with a 5-point Likert response scale. (For a discussion of the adequacy of this selection of facets to represent the five factors, see McCrae & Costa, 2008.) For the present study, participants were administered a questionnaire consisting of the 240 items of the NEO-PI-R and 37 additional items developed for the NEO-PI-3, a more readable version of the instrument (McCrae, Costa, et al., 2005). Previous analyses (De Fruyt et al., 2009) demonstrated that the psychometric properties of the NEO-PI-3 are maintained in the translations used in this study, and that the instrument is essentially equivalent to the NEO-PI-R in both structure and mean levels. It is therefore appropriate to compare NEO-PI-3 scores in the present sample with NEO-PI-R scores obtained in previous studies. NEO-PI-3 facet scales were standardized as T scores within the full sample (i.e., using individual level data, $N = 5,109$, as international adolescent Form R NEO-PI-3 norms); factor scores were computed using the factor scoring weights for observer ratings presented in the manual (Costa & McCrae, 1992, Table 2, bottom panel). Aggregate scores were the mean T scores in each sample or subgroup.

An index of data quality was also computed for each sample, based on four indicators: Number of protocols with more than 40 missing items, percentage of missing responses in valid protocols, number of protocols with evidence of acquiescence or naysaying, and responses in the unscreened sample to a single-item validity check asking respondents if they had answered honestly and accurately. Internal consistency of this quality index was .67.

Criteria

Validity of aggregate APPOC scores was examined by comparing scores to those previously reported in other samples. These include aggregate self-report NEO-PI-R data from a collection of available data sets (McCrae 2002; McCrae & Terracciano, 2008), observer rating NEO-PI-R data from the adult PPOC (McCrae, Terracciano, & 79 Members, 2005), and self-report BFI data (Schmitt et al., 2007). In addition, APPOC scores are also compared to national character stereotype (NCS) data (McCrae et al., 2007a), in which the ''typical'' member of a culture was rated by culture members on 30 scales corresponding to the facets of the NEO-PI-R. For example, the N1: Anxiety facet was assessed by asking if the typical culture member was ''anxious, nervous, worrying vs. at ease, calm, relaxed.'' When factored across nations, the structure of these stereotype ratings roughly replicated the structure of the NEO-PI-R (Terracciano et al., 2005). If stereotypes are, in fact, groundless, then NCS data provide information on the discriminant validity of aggregate trait scores.

RESULTS AND DISCUSSION

Preliminary Analyses

We compared personality profiles in the three sites in the United States and the two sites in Poland. Using the SPSS Reliability program, treating sites as items and NEO-PI-3 facets as cases, we calculated average measure intraclass correlations under the absolute agreement definition. These values were .77 for the United States and .82 for Poland ($ps < .001$). Data from these cultures were therefore collapsed (as the unweighted means of the different sites) for further analyses.

In previous research (McCrae, Terracciano, & 79 Members, 2005), the variance of facet scores was related to geography, with larger standard deviations across the full range of facet scores for modern, Western cultures. The same pattern was found in the present study, with the lowest mean SDs in Malaysia, Peru, and Uganda, and the highest mean SDs in France, Australia, and Estonia. The correlation of mean SD in the present study with mean SD in the PPOC sample was $r = .73$, $N = 24$, $p < .001$. These geographical variations might be due to real differences in the homogeneity of traits in different cultures, to different response styles (e.g., acquiescence), or to differences in data quality, which also tends to be lower in non-Western countries (see McCrae, Terracciano, & 79 Members, 2005).

Also in previous research (Costa, Terracciano, & McCrae, 2001; Schmitt, Realo, Voracek, & Allik, 2008), the magnitude of gender differences was geographically ordered, with the most marked differences found in modern cultures. As in PPOC (McCrae, Terracciano, & 78 Members, 2005), we calculated gender difference indexes for each of the five factors, based on the facets on which adult women scored higher than men in self-reports (Costa, Terracciano, & McCrae, 2001). For example, because women scored higher than men on Openness to Aesthetics, Feeling, and Actions and lower on Openness to Ideas, a Female Openness/Closedness index was defined as (O2: Aesthetics+O3: Feelings+O4 Actions−O5: Ideas)/4. Girls were rated significantly higher than boys in 74 of the 120 comparisons on the five indexes in 24 cultures. As in previous studies, the five indexes were positively intercorrelated and were summed to represent a general gender differentiation score ($\alpha = .78$). As expected, the smallest differentiation was seen in Puerto Rico, Peru, and Uganda and the largest in Hong Kong, Slovakia, and Estonia. However, there were also some anomalous findings: Gender differentiation was low in Australia but relatively high in Malaysia. The correlation of gender differentiation in the present study with gender differentiation in the PPOC sample was only marginally significant ($r = .37$, $N = 23$, $p < .05$, one-tailed). In adult samples, lack of gender differentiation in traditional cultures has been attributed to the tendency of traditional men and women to compare themselves only to others of their own sex, in effect norming away gender differences in observed scores (Guimond et al., 2007). If so, then true gender differences are likely to be similar in all cultures.

In any culture-level analysis it is necessary to recall that variation within cultures is usually far larger that variation across cultures. A components-of-variance analysis conducted on PPOC data (McCrae & Terracciano, 2008) showed that culture accounted for about 4% of the total variance, age (college vs. adult) for 3%, and sex for about 1%. Table 1 provides parallel information for APPOC. Here the effect of age is far smaller because the age groups differ very little. The effects of culture and sex, however, are similar to those seen in adult targets, although in adolescent targets, the effects of culture are most pronounced for Extraversion and least for Agreeableness.

The top panel of Table 2 presents evidence on the generalizability of aggregate personality scores across gender and age groups. For these analyses, culture means for factor scores were derived for boys

Table 1

Percentage of Variance in Observer-Rated NEO-PI-3 Factor Scores
Attributable to Culture, Sex, and Age

	Factor					
Source	N	E	O	A	C	Mean
Culture	3.6*	5.0*	2.9*	1.5*	4.3*	3.46
Sex	2.8*	0.1*	1.2*	0.8*	2.2*	1.42
Age	0.2*	0.0	0.0	0.2*	0.2*	0.12
Culture × Sex	0.8*	0.6	0.9*	0.5	0.5	0.66
Culture × Age	0.8*	0.7	1.0*	0.5	0.7	0.74
Sex × Age	0.1*	0.0	0.0	0.0	0.0	0.02

Note. $N = 5,109$. Age groups: 12 to 14 versus 15 to 17 years. Values are partial η^2 from a multivariate ANOVA. Three-way interactions were not significant.
*$p < .05$.

and girls (or younger and older targets) separately and correlated across the 24 cultures. All correlations are significant, suggesting that similar estimates of culture-level means would be obtained regardless of the age or gender of the targets.

We asked about the relationship of raters to targets and found that it varied somewhat across cultures. For example, 30% of the targets in Thailand were relatives of the raters, whereas 87% were relatives in Iran. De Fruyt and colleagues (2009) created a familiarity index based on questions about how well the raters knew the target, how often they saw them, and in how many different contexts. On a 0 to 4 scale, familiarity values ranged from 1.88 in Japan to 3.35 in Australia. Raters reported that they had known targets for from 0 to 17 years, with a mean of 9.2 years, but none of the raters had known their targets for over 10 years in Croatia or Portugal. Because of these differences across samples, we conducted analyses of variance on the five factors with culture and each of the dichotomized relationship categories as classifying variables. Most of the effects, even when significant in this large sample, were trivial in magnitude, and none of the main effects for relationship category or interaction effects accounted for more than 1% of the variance. The largest main effect showed that, unsurprisingly, well-known targets were rated higher in Extraversion ($M = 50.7$) than less well-known targets ($M = 48.5$). We also examined the generalizability of aggregate

Table 2
Generalizability and Convergent Correlations of Culture-Level Factor Scores

	APPOC Factor				
	N	E	O	A	C
Generalizability					
Across gender	.68***	.82***	.56**	.54**	.83***
Across age	.61***	.79***	.50**	.49**	.72***
Across relationships					
Type	.84***	.80***	.59**	.49*	.73***
Length[a]	.79***	.78***	.56**	.33	.63**
Familiarity	.82***	.65***	.65***	.43*	.76***
Convergent correlation					
Form R	.50**	.55**	.37*	− .02	.09
Form S	.44*	.74***	− .14	.35	.36
BFI	.44*	.45*	− .27	− .05	.17

Note. Type = friend or acquaintance ($N = 2{,}456$) versus relative ($N = 2{,}588$). Length = known for less than ($N = 2{,}528$) versus more than ($N = 2{,}300$) 10 years. Familiarity = lower ($N = 2{,}327$) versus higher ($N = 2{,}629$). Form R = observer rating NEO-PI-R data, $N = 24$, from McCrae, Terracciano, & 79 Members of the Personality Profiles of Cultures Project (2005); Form S = self-report NEO-PI-R data, $N = 16$, from McCrae (2002) and McCrae and Terracciano (2008); BFI = self-report Big Five Inventory data, $N = 18$, from Schmitt et al. (2007).
[a]Across 22 cultures.
*$p < .05$, **$p < .01$, ***$p < .001$, one-tailed.

scores across relationship categories. The top panel of Table 2 shows that, in general, there is strong replicability. Within this pool of generally well-acquainted raters, the details of the relationship do not seem to have major effects, so sample differences in these details are unlikely to affect results.

Convergent and Discriminant Validity of Aggregate Scores

Validity of Scales Across Cultures

The bottom panel of Table 2 shows correlations with aggregate observer ratings (Form R) and self-reports (Form S) on the NEO-PI-R from previous studies. It also presents correlations with aggregated BFI self-reports. There is strong evidence of convergent validity for

the Neuroticism and Extraversion factors, only weak evidence for Openness, and no evidence in these data for the validity of aggregate Agreeableness and Conscientiousness scores. Nonsignificant correlations for the Agreeableness factor across studies were also reported by McCrae, Terracciano, and 79 Members (2005) and Schmitt and colleagues (2007).

Table 3 provides convergent validity information at the level of the facet scales. The intraclass correlation (first data column; $ICC(1, k) = [BMS–WMS]/BMS$) reflects agreement among raters on targets from each of the 24 cultures and estimates the reliability of the aggregate scores. These values are very slightly smaller than those found in analyses of adult targets (*Mdn ICC* = .91; McCrae, Terracciano, & 79 Members, 2005).

The second and third data columns in Table 3 show convergent correlations with observer rating and self-report data on the NEO-PI-R. For Form R, 23 (76.7%) of the facets show significant cross-study agreement; for Form S, 20 (66.7%) are significant. E2: Gregariousness, O4: Actions, O5: Ideas, C3: Dutifulness, and C5: Self-Discipline failed to reach significance in either comparison; Dutifulness and Self-discipline also failed to show cross-study agreement in the PPOC study (McCrae, Terracciano, & 79 Members, 2005). However, the present data relate aggregate traits in ratings of adolescents using the NEO-PI-3 to aggregate traits in ratings and self-reports of adults using the original NEO-PI-R; from this perspective the overall degree of convergence is striking.

A comparison of Tables 3 and 2 highlights a puzzling finding: Why are the traits that define the Agreeableness and Conscientiousness factors generally related across studies, whereas the factors themselves are not? In both PPOC (McCrae, Terracciano, & 79 Members, 2005) and APPOC (reported below in Table 5), culture-level analyses clearly show Agreeableness and Conscientiousness factors because the facets covary as expected. But the cross-facet, cross-study correlations are not consistently positive. For example, the correlation between aggregate A4: Compliance in adolescents and aggregate A5: Modesty in adults is − .53, $p < .01$. Such anomalies may be due to the small sample size ($N = 24$), but they may also imply that there is more agreement on facet-specific variance than on common variance at the culture level.

The last column of Table 3 reports correlations between APPOC aggregate traits and NCS scores across 22 cultures. Five correlations are significant, but three of them are negative. The positive associ-

Table 3
Intraclass Reliability and Cross-Instrument Correlations for NEO-PI-3
Facet Scales

NEO-PI-3 Facet Scale	$ICC(1,k)$	r^a		
		Form R	Form S	NCS
N1: Anxiety	.90	.65***	.79***	.05
N2: Angry Hostility	.79	.52**	.03	.18
N3: Depression	.86	.55**	.46*	.17
N4: Self-Consciousness	.77	.40*	.43*	− .10
N5: Impulsiveness	.87	.51**	.60**	.05
N6: Vulnerability	.90	.61***	.72***	.54**
E1: Warmth	.90	.60***	.33	− .40(*)
E2: Gregariousness	.84	− .18	.27	.27
E3: Assertiveness	.76	.37*	.67**	.00
E4: Activity	.89	.39*	.51*	− .26
E5: Excitement Seeking	.91	.49**	.82***	.35
E6: Positive Emotions	.81	.43*	.35	− .41(*)
O1: Fantasy	.91	.54**	.40	− .10
O2: Aesthetics	.90	.58**	.12	− .21
O3: Feelings	.90	.78***	.56*	− .14
O4: Actions	.88	.34	− .04	− .29
O5: Ideas	.84	.28	.08	.07
O6: Values	.92	.61***	.75***	− .04
A1: Trust	.90	.48**	.48*	− .20
A2: Straightforwardness	.82	.24	.65**	.26
A3: Altruism	.90	.74***	.72***	− .04
A4: Compliance	.91	.60***	.44*	.36*
A5: Modesty	.80	.63***	.70**	.08
A6: Tender-Mindedness	.89	.32	.47*	− .02
C1: Competence	.81	.52**	.63**	− .37(*)
C2: Order	.88	.47*	.48*	.12
C3: Dutifulness	.86	− .10	.42	− .10
C4: Achievement Striving	.90	.44*	.52*	− .33
C5: Self-Discipline	.84	.24	.18	.31
C6: Deliberation	.92	.58**	.68**	.16
Mdn	.89	.50	.48	− .01

[a]Correlations with aggregate NEO-PI-R facet scores and NCS scales: Form R (observer rating data, $N = 24$) from McCrae, Terracciano, and 79 Members (2005); Form S (self-report data, $N = 16$) from McCrae (2002) and McCrae and Terracciano (2008); NCS data ($N = 22$) from McCrae et al. (2007a).
*$p < .05$, **$p < .01$, ***$p < .001$, one-tailed. (*)Significant as one-tailed test in the wrong direction.

ations of assessed Vulnerability and Compliance with corresponding national stereotypes and the negative correlation of Warmth with its stereotype replicate findings in observer rating data on adults but not in self-report data (Terracciano et al., 2005). Otherwise, these data are consistent with the findings of Terracciano and colleagues, who reported no association of assessed personality with national stereotypes.

Validity of Profiles Within Cultures

Table 4 provides data on comparisons of the 30-facet profiles within each culture. As in previous research, means for each facet were first standardized across the set of cultures used in each analysis; intraclass correlations were then calculated across the 30 facets by the double-entry method (see Griffin & Gonzalez, 1995). Comparing APPOC data to adult Form R data (first data column), significant profile agreement was found for 18 cultures (75.0%), including 6 of 8 cultures not included in the earlier PPOC comparison (McCrae, Terracciano, & 79 Members, 2005). Comparing APPOC data to adult Form S data (third data column), agreement was found for 9 of 16 cultures (56.3%). The magnitude of cross-study agreement was not related to data quality or n of targets in APPOC.

The fifth data column of Table 4 reports *ICC* values for profile agreement with national character stereotypes for 22 cultures. Significant positive correlations were found for Argentina and Turkey, whereas significant negative correlations—contradicting the hypothesis of veridical stereotypes—were found for Australia, the Czech Republic, France, Hong Kong, and Peru. None of these correlations replicated findings reported by Terracciano and colleagues (2005), and the median intraclass correlation was $-.01$. These analyses confirm that national character stereotypes in general do not reflect mean personality trait levels.

The second, fourth, and sixth data columns of Table 4 report a second measure of profile agreement, r_c (Cohen, 1969). Intraclass correlations are sensitive to the shape and relative elevation of profiles, but they do not take into account the direction of scoring. A profile that included measures of Introversion would look quite different from one that included measures of its polar opposite, Extraversion, and would generally yield different *ICC* values, but it would contain the same information. Cohen's r_c is invariant over the

Table 4
Agreement of Adolescents' NEO Personality Inventory-3 Profiles With Adults' Revised NEO Personality Inventory Profiles and National Character Survey Scales

| Culture | Adult NEO-PI-R | | | | | |
| | Form R | | Form S | | NCS | |
	ICC	r_c	ICC	r_c	ICC	r_c
Argentina[a]	.43**	.43**	—	—	.39*	.40*
Australia[a]	.45**	.47**	—	—	− .34(*)	− .32(*)
Chile[a]	.24	.51**	—	—	− .10	− .11
Croatia	.59***	.63***	.25	.26	.04	.06
Czech Republic	− .13	− .13	.53**	.57***	− .33(*)	− .30
Estonia	.58***	.59***	.84***	.85***	− .18	− .18
France	.65***	.65***	.54**	.56***	− .37(*)	− .39(*)
Hong Kong	.47**	.58***	.65***	.70***	− .40(*)	− .34(*)
Islamic Republic of Iran[a]	.04	.05	—	—	—	—
Japan	.77***	.78***	.47**	.48**	.24	.25
Malaysia	.72***	.72***	.65***	.66***	.18	.19
People's Republic of China	.48**	.58***	.03	.06	.23	.30
Peru	.15	.15	.23	.24	− .54(**)	− .52(**)
Poland	.35*	.38*	.33*	.37*	.27	.29
Portugal	.20	.56***	.28	.42**	− .14	.07
Puerto Rico[a]	.41*	.43**	—	—	− .26	− .21
Russia	− .34(*)	− .27	.12	.20	− .06	− .03
South Korea	.52**	.52**	.51**	.52**	.18	.24
Serbia	.51**	.51**	.27	.30	.12	.12
Slovakia[a]	.56**	.64***	—	—	− .17	− .13
Thailand[a]	.42*	.46**	—	—	—	—
Turkey	.58***	.64***	.16	.21	.42*	.43**
Uganda[a]	.58***	.61***	—	—	.07	.11
United States	.67***	.69***	.51**	.56***	.17	.18
Mdn	.48	.54	.40	.45	− .01	.07

Note. $N = 30$ facets. ICC = intraclass correlations (double-entry method). r_c = Cohen's r. Form R (observer rating) data from McCrae, Terracciano, & 79 Members (2005). Form S (self-report) data from McCrae (2002) and McCrae and Terracciano (2008). NCS = National Character Survey; NCS data from McCrae et al. (2007a).
[a]Not included in previous studies of culture-level convergent validity.
*$p<.05$, **$p<.01$, ***$p<.001$, one-tailed. (*),(**)Significant as one-tailed test in the wrong direction.

direction of scale scoring because each scale's reflection around the mean (in this case, $T = 50$) is also included in the profile. It is sensitive to both the shape and the absolute elevation of the two profiles. Reanalysis of data on profile agreement across observers (McCrae, 2008) showed that r_c is as effective as ICC in identifying matched versus mismatched data. Table 4 reports r_c values and provides further support for the view that aggregate adult personality scores, but not national character stereotypes, are related to aggregate adolescent scores. Adolescent profiles for Chile and Portugal are significantly related to adult profiles when r_c is used as the measure of profile agreement.

Geographical Patterns

Associations among aggregate personality profiles were examined using nonmetric Multidimensional Scaling (MDS) to see if profile similarity was associated with geographical patterns. Analysis followed the methods used in previous research (Allik & McCrae, 2004; McCrae, Terracciano, & 79 Members, 2005): Aggregate scores for the 24 cultures were standardized across cultures, a distance matrix was calculated based on (1–Pearson r) across the 30 NEO-PI-3 facets, coordinates for two MDS dimensions were derived (StatSoft, 1995), and these coordinates were correlated with factor scores and rotated to maximize the correlations of the vertical axis with Neuroticism ($r = .75$) and the horizontal axis with Extraversion ($r = .83$). The standardized stress value for the two-dimensional solution was .21, which suggests the need for additional dimensions (five dimensions showed a stress value of .06), but because our intent was to compare these results to previous MDS results, we report the two-dimensional solution.

Figure 1 displays results. As in previous studies, Western cultures are found on the right (extraverted) side of the plot, non-Western cultures on the left. French, Czechs, Argentines, and Hong Kong Chinese are again found at the top of the figure and Estonians and Mainland Chinese at the bottom. There is one notable difference: Russian adolescents are located in the bottom right of the figure and thus appear to be more adjusted and extraverted than older Russians (McCrae, Terracciano, & 79 Members, 2005). Resemblance to the MDS analysis of PPOC data can be quantified by correlating the coordinates across the two studies. Agreement was strong for the horizontal axis, $r = .71$, $N = 24$, $p < .001$; for the vertical axis,
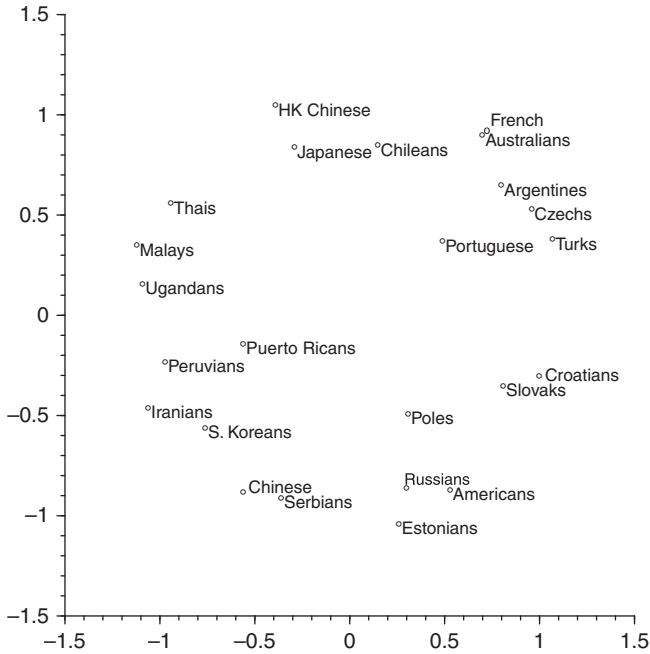
Figure 1
Multidimensional scaling plot of 24 cultures based on a distance matrix of (1−Pearson r) for the 30 NEO Personality Inventory-3 facet scores, standardized across cultures. The vertical axis is maximally aligned with Neuroticism and the horizontal axis with Extraversion. HK Chinese = Hong Kong Chinese. S. Koreans = South Koreans.

however, it was $r = .34$, *ns*. Omitting the Russians, the correlation for the vertical axis increased to $r = .51$, $N = 23$, $p < .05$.

Culture-Level Factor Structure

As in previous studies, principal component analyses at the culture level were undertaken using mean values from subsamples in order to obtain a reasonably large number of cases. For the present study, 108 subsamples were used, representing older and younger adolescent boys and girls from each of the 27 sites. Results after Procrustes rotation are reported in Table 5. Even in this small sample, the normative, adult, individual-level structure is reasonably replicated for Neuroticism, Extraversion, Agreeableness, and Conscientiousness

Table 5
Culture-Level Factor Structure of NEO-PI-3 Facet Scales

NEO-PI-3 Facet Scale	Procrustes-Rotated Principal Component					
	N	E	O	A	C	VC[a]
N1: Anxiety	**.83**	.11	.03	.22	.19	.90[b]
N2: Angry Hostility	**.80**	− .01	− .03	− .19	− .17	.91[c]
N3: Depression	**.81**	− .14	.04	.19	.01	.91[c]
N4: Self-Consciousness	**.77**	− .20	− .10	− .06	.07	.95[b]
N5: Impulsiveness	**.48**	.29	.23	− .33	− **.49**	.94[b]
N6: Vulnerability	**.77**	− .13	.29	.10	− .28	.89[c]
E1: Warmth	− .12	**.64**	.27	**.41**	− .20	.91[c]
E2: Gregariousness	.03	**.70**	**.43**	.20	.00	.84
E3: Assertiveness	− .24	**.53**	.39	− .21	.39	.96[b]
E4: Activity	− .18	**.44**	**.60**	− .35	.16	.76
E5: Excitement Seeking	− .18	**.51**	− .16	− **.51**	− .17	.88[c]
E6: Positive Emotions	.06	**.53**	**.45**	.26	.03	.87[c]
O1: Fantasy	.38	**.65**	− .01	− .09	− .34	.49
O2: Aesthetics	.31	− .03	**.46**	.38	**.58**	.76
O3: Feelings	.27	**.59**	**.44**	.27	.17	.91[c]
O4: Actions	− .24	− .03	**.73**	− .11	.17	.87[c]
O5: Ideas	− .14	− .09	− .02	− .18	**.50**	.24
O6: Values	− .19	**.51**	− .13	.22	− .18	.04
A1: Trust	− .05	**.46**	.15	**.40**	− .03	.82
A2: Straightforwardness	− .07	− .17	.26	**.59**	− .01	.80
A3: Altruism	.00	**.71**	− .18	.29	.16	.90[c]
A4: Compliance	.05	− **.47**	.08	**.48**	.23	.71
A5: Modesty	.03	.10	− .14	**.47**	− .18	.89[c]
A6: Tender-Mindedness	.19	**.41**	− .18	**.46**	**.52**	.65
C1: Competence	− .26	.39	− .11	− .01	**.75**	.91[c]
C2: Order	− .02	− .13	.22	.28	**.78**	.79
C3: Dutifulness	.02	.02	.15	**.40**	**.83**	.94[b]
C4: Achievement Striving	− .10	.15	.24	− .11	**.88**	.99[b]
C5: Self-Discipline	− .13	.04	.12	.22	**.87**	.92[c]
C6: Deliberation	− .09	− .38	− .02	**.41**	**.72**	.96[b]
Factor congruence[d]	.93[b]	.88[b]	.47	.86[b]	.88[b]	.83[b]

Note. These are principal components from 108 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]Congruence higher than that of 99% of rotations from random data. [c]Congruence higher than that of 95% of rotations from random data. [d]Congruence with American normative factor structure.

factors (congruence > .85; Lorenzo-Seva & ten Berge, 2006), and 26 of the 30 facets show loadings above .40 on the intended factor. Comparisons to randomly permuted data from an earlier study of the NEO-PI-R (McCrae, Zonderman, Costa, Bond, & Paunonen, 1996) suggested that 4 factor congruences and 19 of the 30 variable congruence coefficients exceeded chance values.

However, the Openness factor is clearly not replicated. Three of its intended facets are unrelated to the factor, and three of the definers of the observed factor are facets of Extraversion. There appear to be two reasons for these deviations from the usual structure. First, Openness to Ideas loads on the Conscientiousness factor. This finding at the culture level is expected, given that, in these data, Openness to Ideas loads strongly (.48 to .51) on the Conscientiousness factor at the individual level (De Fruyt et al., 2009). Although sometimes seen in self-reports (Hřebíčková, 2008), this phenomenon appears chiefly in observer ratings of adolescents. Costa et al. (2008) reported a loading of .39 for Openness to Ideas on the Conscientiousness factor when middle-school-aged respondents rated another child of the same age, but only .24 when they rated themselves. In observer ratings of college students and adults (McCrae, Terracciano, & 78 Members, 2005), the loading of O5: Ideas on Conscientiousness is .31; in self-reports from adults (Costa & McCrae, 1992), it is .16. It thus appears that high loadings of O5: Ideas on Conscientiousness are a joint function of method and target age: When outside observers assess intellectual curiosity in school children, they are apt to confuse it with academic success, which is also associated with Conscientiousness. Teachers, for example, attribute academic self-esteem to students they rate as high in both Conscientiousness and Openness (Graziano & Ward, 1992). By contrast, when American adolescents rate themselves, they can distinguish between intrinsic intellectual interest and academic achievement orientation (Costa et al., 2008).

The Openness factor is also poorly defined because O1: Fantasy and O6: Values have their major loadings on the Extraversion factor. This is not unique to analyses of adolescents or of observer ratings; instead, it appears to be a culture-level phenomenon. Modern Western nations tend to be high on Extraversion, and they also tend to embrace such self-expressive values as imagination and tolerance (Inglehart, 1997). Raters from such cultures are thus more likely to describe their compatriot targets as high both in Extraversion and in traits like Fantasy and Values. As data simulations show (McCrae & Terracciano, 2008),

the effect is to broaden the culture-level Extraversion factor to represent something more like individualism.

This is, however, only part of the story. In adult data from PPOC, Openness to Fantasy and Values had joint loadings on the culture-level Extraversion and Openness factors (McCrae, Terracciano, & 79 Members, 2005), whereas Table 4 shows no loadings at all for these facets on the Openness factor. At least with regard to Openness to Values, this may be because young adolescents do not yet have a clearly defined ideology, leading to very low internal consistency for this facet (Costa et al., 2008; De Fruyt et al., 2009).

Conclusion

The present study, using college students' ratings of adolescents aged 12 to 17 on a modified version of the NEO-PI-R in 24 cultures, provides further evidence for three conclusions. First, there is general agreement about characterizations of cultures based on personality assessments of individuals: Adult self-reports, observer ratings of adults, and now observer ratings of adolescents all show similar patterns, whether one considers each trait across all cultures or the profile of all traits within each culture or the clustering of culture profiles in multidimensional space. Second, there is no consistent agreement between these aggregate characterizations of cultures and the corresponding collective beliefs about traits of the "typical" culture member: National character stereotypes again appear to be largely unfounded. Finally, there is further evidence that the culture-level factor structure differs from the individual-level structure with regard to the Extraversion factor. In ratings of young adolescents, as in observer ratings and self-reports of college students and adults, Openness to Fantasy and Values, Competence, and low Compliance are associated with the Extraversion factor, but only at the culture level. This robust finding requires a culture-level explanation.

The repeated finding that national character stereotypes are unrelated to assessed aggregate personality has seemed counterintuitive to some psychologists (e.g., Perugini & Richetin, 2007), but it makes sense if national stereotypes are, in fact, determined chiefly by such nonpsychological features as a nation's wealth or mean temperature (McCrae et al., 2007a). This finding is not of merely academic interest: Beliefs about national character can have an important influence on political and social views and affect both ethnic and international relations.

Psychologists should educate the public on the dangers of stereotypic thinking, especially with regard to national stereotypes. At the same time, they need to conduct more research on the origins of these beliefs and how they might be changed (Terracciano & McCrae, 2007).

Other findings from the present study pose more purely intellectual challenges. At the individual level, aggregating facets to define broad domains generally leads to more reliable and valid scores. For example, among adolescents ages 14–20, the median cross-observer correlation for the five NEO-PI-3 domains is .53, whereas the median for the 30 facets is only .43 (McCrae, Costa, et al., 2005). That pattern is reversed at the culture level: In the present study, the median Form R cross-study correlation is .37 for the five domains but .50 for the 30 facets. It is possible that this finding is a fluke, attributable to the small number of cultures examined. Until that can be established, however, it would appear wise to conduct cross-cultural comparisons of aggregate traits chiefly at the facet level: We can have more confidence in the claim that a given culture is high in Altruism or Deliberation than that it is high in Agreeableness or Conscientiousness. Studies on the cultural origins or effects of personality traits should target specific facets.

The basic claim of the field of culture-level personality studies—that averaging the trait scores of a sample of culture members can yield meaningful information about the personality profile of the culture group itself—is far from indisputable, but it has shown itself to be a valuable working hypothesis. How far this hypothesis can be generalized to other individual difference variables (e.g., attitudes, interests, values) remains to be seen.

REFERENCES

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

Boster, J. S., & Maltseva, K. (2006). A crystal seen from each of its vertices: European views of European national characters. *Cross-Cultural Research*, **40**, 47–64.

Church, A. T. (2001). Introduction. *Journal of Personality*, **69**, 787–801.

Cohen, J. (1969). r_c: A profile similarity coefficient invariant over variable reflection. *Psychological Bulletin*, **71**, 281–284.

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

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.

De Fruyt, F., De Bolle, M., McCrae, R. R., Terracciano, A., Costa, P. T., Jr., & 43 Collaborators of the Adolescent Personality Profiles of Cultures Project (2009). Assessing the universal structure of personality in early adolescence: The NEO-PI-R and the NEO-PI-3 in 24 cultures. *Assessment*, **16**, 301–311.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, **82**, 878–902.

Graziano, W. G., & Ward, D. (1992). Probing the Big Five in adolescence: Personality and adjustment during a developmental transition. *Journal of Personality*, **60**, 425–439.

Griffin, D., & Gonzalez, R. (1995). Correlational analysis of dyad-level data in the exchangeable case. *Psychological Bulletin*, **118**, 430–439.

Guimond, S., Brunot, S., Chatard, A., Garcia, D. M., Martinot, D., Branscombe, N. R., et al. (2007). Culture, gender, and the self: Variations and impact of social comparison processes. *Journal of Personality and Social Psychology*, **92**, 1118–1134.

Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What do cross-national comparisons of personality traits tell us? The case of conscientiousness. *Psychological Science*, **19**, 309–313.

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.

Hřebíčková, M. (2008). NEO osobnostni inventar (NEO-PI-3) v diagnostice osobnosti deti a adolescentu [NEO Personality Inventory (NEO-PI-3) in diagnostics of children and adolescents]. *Ceskoslovenska Psychologie*, **52**, 425–442.

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

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The ''Big Five'' Inventory—Versions 4a and 54*. Berkeley: University of California, Berkeley, Institute of Personality and Social Research.

LeVine, R. A. (2001). Culture and personality studies, 1918–1960: Myth and history. *Journal of Personality*, **69**, 803–818.

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

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.

Madon, S., Jussim, L., Keiper, S., Eccles, J., Smith, A., & Palumbo, P. (1998). The accuracy and power of sex, social class, and ethnic stereotypes: A naturalistic study in person perception. *Personality and Social Psychology Bulletin*, **24**, 1304–1318.

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. (2008). A note on some measures of profile agreement. *Journal of Personality Assessment*, **90**, 105–109.

McCrae, R. R., & Costa, P. T., Jr. (2008). Empirical and theoretical status of the Five-Factor Model of personality traits. In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *Sage handbook of personality theory and assessment* (Vol. 1, pp. 273–294). Thousand Oaks, CA: Sage.

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

McCrae, R. R., & Terracciano, A. (2008). The Five-Factor Model and its correlates in individuals and cultures. In F. J. R. Van de Vijver, D. A. van Hemert, & Y. H. Poortinga (Eds.), *Multilevel analyses of individuals and cultures* (pp. 249–283). Mahwah, NJ: Erlbaum.

McCrae, R. R., Terracciano, A., & 78 Members of the Personality Profiles of Cultures Project. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, **88**, 547–561.

McCrae, R. R., Terracciano, A., & 79 Members of the Personality Profiles of Cultures Project. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology*, **89**, 407–425.

McCrae, R. R., Terracciano, A., Realo, A., & Allik, J. (2007a). Climatic warmth and national wealth: Some culture-level determinants of national character stereotypes. *European Journal of Personality*, **21**, 953–976.

McCrae, R. R., Terracciano, A., Realo, A., & Allik, J. (2007b). On the validity of culture-level personality and stereotypes score. *European Journal of Personality*, **21**, 987–991.

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.

Nye, C. D., Roberts, B. W., Saucier, G., & Zhou, X. (2008). Testing the measurement equivalence of personality adjective items across cultures. *Journal of Research in Personality*, **42**, 1524–1536.

Oishi, S., & Roth, D. P. (2009). The role of self-reports in culture and personality research: It is too early to give up on self-reports. *Journal of Research in Personality*, **43**, 107–109.

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

Perugini, M., & Richetin, J. (2007). In the land of the blind, the one-eyed man is king. *European Journal of Personality*, **21**, 977–981.

Rentfrow, P. J., Gosling, S. D., & Potter, J. (2008). A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics. *Perspectives on Psychological Science*, **3**, 339–369.

Schmitt, D. P., Allik, J., McCrae, R. R., Benet-Martínez, V., Alcalay, L., Ault, L., et al. (2007). The geographic distribution of Big Five personality traits: Patterns and profiles of human self-description across 56 nations. *Journal of Cross-Cultural Psychology*, **38**, 173–212.

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, **94**, 168–182.

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

Terracciano, A., Abdel-Khalak, A. M., Ádám, N., Adamovová, L., Ahn, C.-k., Ahn, H.-n., et al. (2005). National character does not reflect mean personality trait levels in 49 cultures. *Science*, **310**, 96–100.

Terracciano, A., & McCrae, R. R. (2007). Perceptions of Americans and the Iraq invasion: Implications for understanding national character stereotypes. *Journal of Cross-Cultural Psychology*, **38**, 695–710.

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.