PERSONALITY DIMENSIONS ACROSS CULTURES

Jüri Allik, PhD

In order to generalize the dimensional structure of personality—relatively independent groups of covarying traits—across languages and cultures, a large number of cultures must be studied. Until recently only a few worldwide personality datasets have been available. The first large–scale studies indicate that the pattern of covariation between personality traits is universal and is relatively easily generalizable across languages and cultures. In contrast to the structure of personality, the comparison of the mean trait scores across cultures is much more problematic because cross–cultural differences turned out to be very small in their magnitude, about one–third of the magnitude of individual differences within culture. More integral (e.g., the similarity between personality profiles) or subtle (e.g., the disparity between positively and negatively worded items) measures can reveal more systematic relationships with relevant socio-economic and geographic variables than the mean scores themselves. Relatively modest sizes of cross–cultural differences in the mean values may imply that a reasonable scalar equivalence can be achieved, and all individuals, irrespective of their language and culture, can be represented in a common metric.

Many popular psychological assessment instruments, originally developed in English, have been translated into numerous languages and are now commonly used throughout the world. Most of these translations were made with an explicit or at least tacit assumption that the core psychological constructs assessed by the measures substantively transcend human language and culture. The generalizability across languages and cultures, however, was in most cases presumed, not demonstrated. Therefore it is not surprising that some researchers have expressed concern with this assumption, and especially with practices guided by this assumption (Shweder, 1991). Skeptics have questioned, for instance, whether the uncritical extension of "Western" ways of thinking to the rest of the world should serve as standard procedure in psychological science. From a cultural constructionist point of view, all personality models are based on conceptions of personhood and

Address correspondence to Jüri Allik, Department of Psychology, University of Tartu, Tiigi 78, Tartu 50410, Estonia; E-mail: juri.allik@ut.ee.

standards of culturally appropriate behavior that have evolved, as a rule, in Anglo–American societies. It may be even more so in regard to personality disorders which are, as some researchers believe, quintessential cultural products that owe their meaning and descriptive content to a distinctive cultural tradition (Fabrega, 1994). It is also claimed that studies of mental health are culturally insensitive partly due to the uncritical transfer of tools and concepts from one language and culture, typically English and the United States, to other countries (Rogler, 1999).

Eysenck was one of the pioneers who started to think that the pattern of covariation between personality traits—the personality structure—might be universal and rises above the diversity of human cultures (cf. Allik, 2002). One of the main reasons for this conclusion was his study of twins that demonstrated a considerable genetic impact on all personality traits and almost no influence from the shared environment (Eysenck, 1990). In 1997, McCrae and Costa made a bold proposal about the cross–cultural generalizability of the Five–Factor Model of Personality. By studying six translations of the Revised NEO Personality Inventory (Costa & McCrae, 1992) into German, Portuguese, Hebrew, Chinese, Korean, and Japanese they observed that all translations showed similar structures after varimax rotation of five factors and almost identical factor structure when the varimax solution was targeted toward the original American factor structure. For some reason, people from different places around the world are inclined to think that individuals who, for example, talk a lot are at the same time optimistic and very happy, or those who often get disgusted with other people are also inclined to feel inferior to others. Because the samples studied represented highly diverse cultures with languages from five distinct language families, McCrae and Costa proposed that personality trait structure is universal (McCrae & Costa, 1997). The proposal that the personality structure is largely or even completely independent from the culture (McCrae & Costa, 1996; McCrae & Costa, 1999) was rather unorthodox and was expectedly met with skepticism and denial (Allik & McCrae, 2004a; Toomela, 2003). It is also important to notice that although the replicability of the pattern of covariations across cultures turned out to be a rather robust phenomenon (cf. Rolland, 2002), this fact does not necessarily imply that all of the finest details are always replicated in every new culture. Universality by itself is a relative concept implying that something must be variable as well; it is not an all–or–nothing phenomenon, but a degree of invariance.

PROBLEMS IN COMPARING PERSONALITY TRAIT SCORES ACROSS CULTURES

For psychologists seeking to investigate personality traits across cultures, one of the more inconvenient problems has centered on whether personality trait scales possess conceptual equivalence across cultures or not (van de Vijver & Leung, 1997; van de Vijver & Leung, 2000). It has been particularly troublesome to establish whether the mean scores across different cultures

show metric or scalar equivalence (Byrne & Campbell, 1999; Little, 2000). That is, when comparing the mean scores of different cultures on a personality trait scale, any observed differences may be due not only to a real cultural disparity on some personality trait, but also to inappropriate translations, biased sampling, or the nonidentical response styles of people from different cultures. All of these factors can be difficult to control fully, making some methodologists extremely skeptical about achieving true metric comparability of scores on the same test in different languages or cultures (Bijnen & Poortinga, 1988; Heine, Lehman, Peng, & Greenholtz, 2002; van de Vijver & Leung, 1997). Although much of this skepticism is certainly warranted, a recent, large–scale intercultural comparison provided evidence that personality scales may be more robust than methodologists had initially presumed (McCrae, 2001; McCrae, 2002). Perhaps because the many sources of error cancel out, it is possible to obtain meaningful results when scores are compared across cultures (Allik & McCrae, 2002).

Among the more common methods for establishing the cross–cultural comparability of personality trait measures is first to show that the trait scales contained in the measures are internally reliable across all targeted languages and cultures. A second frequently employed technique is to demonstrate a high degree of factorial structure invariance across different linguistic and cultural contexts (Barrett, Petrides, Eysenck, & Eysenck, 1998; Caprara, Barbaranelli, Bermudez, Maslach, & Ruch, 2000; Rolland, 2002). Metric equivalence can also be established through differential item analysis and bilingual administrations. The number of bilingual studies is surprisingly small, however, and theses studies rarely demonstrate consistent differences between languages (e.g., Konstabel, 1999). However, if the questionnaire format is replaced with autobiographical narratives, then indeed language may, to some extent, dictate how individuals describe themselves (Marian & Kaushanskaya, 2004). Historically, if trait scales from a personality measure showed high internal reliability, invariant factor structure, and item equivalence across different languages and cultures, comparing the mean scores across cultures was often deemed a reasonable next step (Steel & Ones, 2002; van de Vijver & Leung, 2001). However, even with evidence of reliability, factor invariance, and item equivalence, problems can remain in how to metrically interpret mean–level differences in personality traits across cultures. For example, it is possible that people from different countries have different internal standards and make their judgments in the relation to these standards (cf. Heine et al., 2002).

Another way to increase confidence in the cross–cultural comparability of personality measures is to show that the mean levels of different assessment instruments that intend to measure the same construct, or approximately the same construct, are highly correlated across multiple languages or cultures. For example, if two conceptually similar personality trait scales are used in a large number of different cultures, a positive association between the mean levels of those trait scales across the broad set of cultures would provide evidence that both measures are tapping the same underlying construct (Campbell & Fiske, 1959). Of course, in order to analyze the compara-

bility of personality measures using this cross–cultural convergent validation strategy, large numbers of cultures must be studied using conceptually similar measures of personality.

LARGE–SCALE CROSS–CULTURAL STUDIES

There have been many studies in which pairs of cultures have been compared on personality measures (cf. Katigbak, Church, & Akamine, 1996; cf. Katigbak, Church, Guanzon–Lapena, Carlota, & del Pilar, 2002), but there have only been only a few in which a sufficiently broad sample of cultures was examined. Although many personality instruments like the Temperament and Character Inventory (Cloninger, 2004) have successfully translated into different languages, only few worldwide personality datasets have been available for statistically meaningful comparisons to be made.

One of the first comprehensive personality trait measures to enjoy worldwide popularity and a fairly large number of translations into different languages was Eysenck's Personality Questionnaire (EPQ; Eysenck & Eysenck, 1975). In 1984, mean–level trait scores from 25 countries were made available (Barrett & Eysenck, 1984). Ten years later, the number of countries in which three broad personality traits—Neuroticism, Extraversion, and Psychoticism—were measured by the EPQ was expanded to 37 (Lynn & Martin, 1995). Soon it was found that the internal reliability and factorial structure of the EPQ across languages and cultures appeared to be replicable (Barrett et al., 1998). Although evidence of the cross–cultural generalizability of the EPQ seemed overwhelming, there was still some room for reservations. Some of these doubts were met by van Hemert and colleagues (van Hemert, van de Vijver, Poortinga, & Georgas, 2002) who critically reanalyzed available EPQ data both at the individual and the country level. Many previous studies were dropped because of insufficient information (e.g., the number of women and men was not specified). The final set contained studies in 38 countries with a total of 68,374 respondents. Using the multilevel factor analysis, they found that the EPQ has a different factorial structure at both levels. In some countries (China, India, Japan, and Uganda) the pattern of correlation between EPQ scales, differed from those found elsewhere. From the exploration of the meaning of the EPQ scales, country–level correlations were found with a variety of country characteristics, such as Hofstede's measures of cultural differences and Diener's subjective well–being. A disappointing result of this analysis was that Lynn and Martin's (1995) findings were not replicated, as no expected correlations were found, for example, between extraversion and the death rate.

However, because no other large personality datasets were available for comparison, it remained unclear as to whether mean–level differences in EPQ scores across cultures converged with other similar measures. Again, such cross–cultural construct validity evidence would have made it more likely that national differences in personality, as measured by the three broad trait scales of the EPQ, were due to real cultural disparities, and not some other biasing factors.

The most comprehensive instrument thus far designed to measure the Big Five or FFM is the Revised NEO Personality Inventory (NEO–PI–R; Costa & McCrae, 1992). Recently, the NEO–PI–R was translated into many different languages and administered to samples from over two dozen countries. In 2001, NEO–PI–R data from 26 countries or cultural regions became available to the research community (McCrae, 2001) and the database was soon expanded by 10 additional cultures covering 5 major language families: Indo–European, Uralic, Altaic, Dravidian, and Sino–Tibetian (McCrae, 2002). In every culture and language that has been studied, the trait scales of the NEO–PI–R have displayed adequate levels of internal reliability, and the factorial structure of the NEO–PI–R has been considered robust (McCrae, 2001; McCrae, 2002).

Direct comparisons of the NEO–PI–R with the EPQ have suggested that translations of both instruments provide reasonably comparable estimates of mean levels of extraversion and neuroticism across cultures. For example, the mean–level scores of neuroticism and extraversion as measured by the NEO–PI–R and the EPQ were significantly correlated across 18 nations, $r = .80$ and $r = .51$, respectively (McCrae, 2002). Thus, if a nation scored relatively high on the EPQ Neuroticsm and Extraversion scales, it was likely to score high on the NEO–PI–R Neuroticism and Extraversion scales as well. For this result, however, it was necessary to omit EPQ data from India as a suspected outlier. These empirical findings, though limited to 18 cultural regions, can be taken as supportive evidence that at least two of the Big Five dimensions, neuroticism and extraversion, can be comparably measured across human languages and cultures.

Recently McCrae and Terracciano initiated a large cross–cultural project to extend findings on the universality of personality traits in several important ways (McCrae et al., 2005). Most of cross–cultural studies have relied exclusively on self–report methods, leaving the possibility that, for example, the self–serving bias distorts findings. For this reason the third–person version of the NEO–PI–R was used instead. College students in 50 cultures identified an adult or college–age man or woman who they knew well and rated the 11,985 self-nominated target persons. Factor analyses within cultures showed that the normative American self–report structure was clearly replicated in most cultures, and was at least recognizable in all 50 cultures. The other–report findings replicated some regularities that were identified on the basis of self–reports. For example, like earlier self–report results, sex differences were the most pronounced in Western and less in Eastern and African cultures (cf. Costa, Terracciano, & McCrae, 2001). Cross–sectional age differences for three factors followed the pattern identified in self–reports, with moderate rates of change during college age and very slow changes after age 40. The authors conclude that with a few exceptions, their data support the hypothesis that features of personality traits are common to all human groups (McCrae et al., 2005).

Perhaps David Schmitt conducted the largest cross–cultural study of personality so far as a part of the International Sexuality Description Project, a collaborative effort of over 100 social, behavioral, and biological scientists

from 56 countries (Schmitt et al., 2003). All studied samples were administered the Big Five Inventory (BFI) of personality traits (Benet–Martinez & John, 1998). The BFI, a short version of the Big Five family of questionnaires, was translated from English into 29 languages and administered to 17,837 individuals from 56 nations. Results indicated that the five–dimensional structure of the BFI was highly robust across major regions of the world (Schmitt, Allik, McCrae, Benet–Martínez et al., submitted). However, the correlation between BFI scales and their counterparts from the EPQ and NEO–PI–R was disappointingly low. Although the EPQ and the BFI do not conceptualize Neuroticism and Extraversion in a completely identical way, it is reasonable to expect at least modest convergent correlations. There were 26 countries for which the mean scores of Neuroticism and Extraversion were measured by both the EPQ (Lynn & Martin, 1995; van Hemert et al., 2002) and the BFI. As expected, the Neuroticism scales of the two instruments were significantly correlated, $r = 0.49$, $p = .01$, but the correlation between the BFI Extraversion scale and its EPQ counterpart was disappointingly low and did not reach statistical significance, $r = 0.18$. The problems with convergent correlations were not only due to conceptual differences between instruments of measure. In some cases, the convergent correlation was absent when two conceptually similar the Big Five instruments, the BFI and NEO–PI–R, were compared. There was an overlapping set of 27 countries (i.e., Austria, Belgium, Canada, Croatia, Czech Republic, Estonia, France, Switzerland, Germany, Hong Kong, India, Indonesia, Italy, Japan, Malaysia, Netherlands, Peru, Philippines, Portugal, South Africa, South Korea, Spain, Taiwan, Turkey, United States, Serbia, and Zimbabwe) in which the correlations between the BFI and NEO–PI–R scales were possible. Although the convergent correlations of BFI and NEO–PI–R nation–level scores on Neuroticism, $r = .45$, $p < .05$; Extraversion, $r = .44$, $p < .05$; and Conscientiousness, $r = .45$, $p < .01$, were significant, two remaining convergent correlations, Agreeableness (0.22) and Openness (0.27), failed to reach the level of statistical significance. There were even more problems with the discriminant correlations. For example, the BFI Openness scale was very strongly correlated with the NEO–PI–R Extraversion scale, $r = .73$, $p < .001$, not with its intended counterpart. Thus, two independent measures of the Big Five (the BFI and the NEO–PI–R) demonstrated only partial cross–cultural agreement. Indeed, in some cases parallel measures were rather consistent. According to both the BFI and NEO–PI–R, for example, Japan's level of neuroticism was among the highest of all cultures, and according to the EPQ, Japan's neuroticism was the third highest (Lynn & Martin, 1995). In most other cases, however, agreement between parallel personality measures across cultures was modest at best (Schmitt et al., submitted).

These cross–cultural convergent correlations between the BFI and the NEO–PI–R domain scales were noticeably smaller than cross–instrument convergence at the individual level (i.e., when the same individuals simultaneously complete both measures). At the individual level, even the smallest convergent correlations typically exceed the .60 level with the BFI scales (Benet–Martínez & John, 1998). Apparently, biases and measurement er-

rors prevented the convergent correlations between two instruments of measure from being more substantial at the intercultural level. An indication that the low cross–instrument agreement may be caused by a bias became apparent after controlling for acquiescence bias. Indeed, after controlling for acquiescence, the partial correlations increased slightly; in the case of the BFI and NEO–PI–R Openness scales the association rose from $r = .27$, ns to $r = .40$, $p < .05$. Agreeableness cross–instrument correlations were also affected by partialing out the acquiescence from the BFI, shifting from $r = .22$ to $r = .27$. Thus, the acquiescence bias was likely one of the causes for the lowered convergent correlations between parallel instruments across 27 nations (Schmitt et al., submitted).

GEOGRAPHICAL DISTRIBUTION OF PERSONALITY TRAITS

How are personality traits distributed throughout the world? Until recently, it was impossible to give any informed answer to this question. However, the first known attempt to examine a systematic pattern in the worldwide distribution of personality traits has already proven to be rather rewarding (Allik & McCrae, 2004b). Although the translation quality of the NEO–PI–R varied considerably, and some of the studied cultures were represented by very small (< 100) and convenient (e.g., only college students) samples, the NEO–PI–R dataset provided strong and reliable evidence that the geographical distribution of the mean–level trait scores produced meaningful patterns. Not only the mean–level personality trait scores were predictably related to other culture–level indicators, such as Hofstede's dimensions of culture (Hofstede & McCrae, 2004; McCrae, 2002), but their distribution in geographic space seemed to have regular, systematic patterns. Neighboring countries tended to have, as a rule, similar personality means, and regions separated geographically or historically had less similar means on personality trait scales (Allik & McCrae, 2004b). Figure 1 shows the result of multidimensional scaling in which a clear contrast of European and American cultures with Asian and African cultures is visible: The former were higher in Extraversion and Openness to Experience and lower in Agreeableness. A second dimension reflected differences in psychological adjustment with the largest correlations with Neuroticism and Hofstede's Uncertainty avoidance.

It seems that the distribution of self–reported personality traits is organized geographically. However, the question about the cause of this regularity is without proper answer. One explanation is that the response style to personality questionnaires varies from country to country producing an appearance of geographic regularity. Indeed, it was found that acquiescence response bias is higher in nations that are high on family collectivism and tend to avoid uncertain situations (Smith, 2004). Another possibility is that cross–cultural differences in the personality trait levels are substantial. Although psychologists have a long–standing bias toward cultural explanations, genetical factors also cannot be ignored. Unfortunately, the separation of cultural and genetic causes from all other factors is a compli-
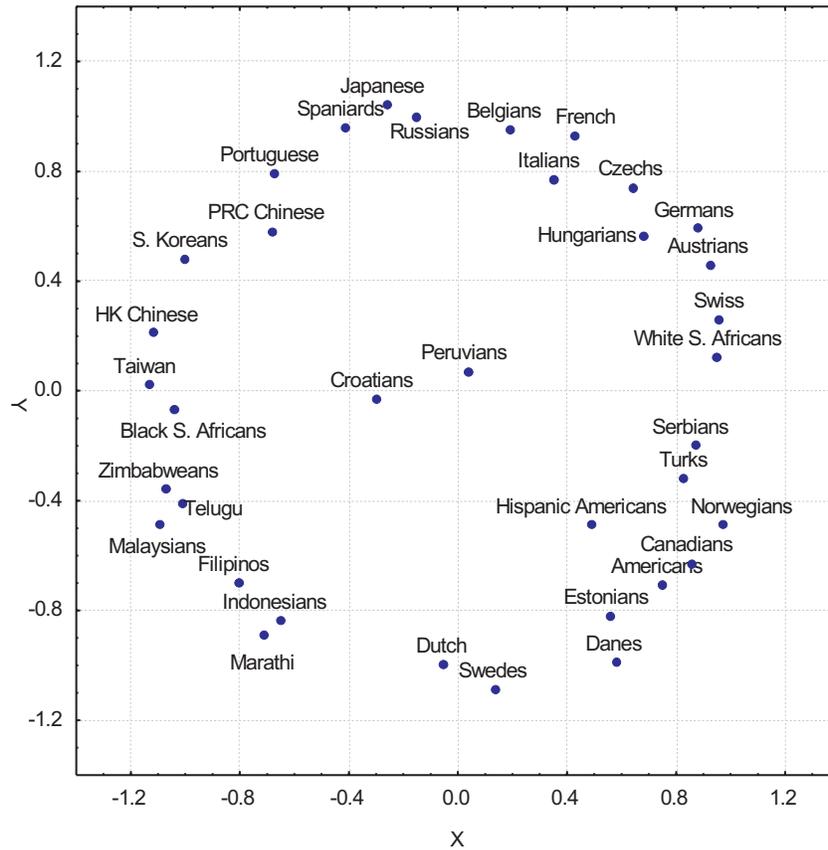
FIGURE 1. Multidimensional scaling plot of 36 cultures. As a mnemonic, it can be noted that "North" in the Figure is associated with N (Neuroticism) and "East" with E (Extraversion). Reproduced from Allik, J., & McCrae, R.R. (2004).

cated and sometimes impossible task. Geographically adjacent cultures often share both geophysical features, like climate, and cultural features, like religion or language family. In addition, however, they are likely to share ancestry and thus have overlapping gene pools (Cavalli–Sforza, Menozzi, & Piazza, 1994). A direct test of this view provided modest support (Allik & McCrae, 2004b). Cavalli–Sforza et al. (1994) provide data on *genetic distances* between cultures, which are determined by similarity in the distribution of a variety of alleles. By and large, these genetic distances correspond to known patterns of human migration. From the data in Table 3 of McCrae (2002) we can calculate personality profile distances, most simply as the Euclidean distance between the five factor scores. Genetic distances are given by Cavalli–Sforza et al. for 16 of the cultures in McCrae; the correlation between genetic and personality profile distances across the 120 pairs of cultures is $r = .19$, $p < .05$. When Serbia, which is identified as an outlier by Cavalli–Sforza et al. (1994, p. 268), is omitted, the correlation rises to $r = .24$,

$N = 105$, $p < .05$. These correlations are modest in part because there is relatively little genetic differentiation within European countries, and correspondingly little variation in mean personality levels. If the analysis were conducted on samples from around the world, larger associations would probably be found. Of course, even very large correlations would not prove causal associations. Genetically related groups tend to share geography, history, and culture as well as genes, and genetic distance may simply serve as a marker of cultural distance (Allik & McCrae, 2004b).

PERSONALITY DIMENSIONS ACROSS CULTURES: PROBLEMS AND CHALLENGES

THE CULTURE–LEVEL FACTOR STRUCTURE

When the number of studied cultures becomes large enough, it is possible to aggregate data within one culture and to treat aggregates as single cases for further cross–cultural analysis. Hofstede referred to this as *ecological factor analysis* and used it to identify dimensions of culture (Hofstede, 1983). However, even a few dozen cultures may be not enough for a meaningful interpretation of the results of the factor analysis. One potential method of increasing the number of studied cases, for example, is to aggregate men and women separately and also treat different age groups as separate cases (McCrae, 2001; 2002). Besides reflecting sex and age differences the factors also reflect the covariation of traits across cultures. It is typical to believe that there is no logical requirement that the culture level or ecological factor structure will bear any resemblance to the factor structure observed when individual data are analyzed. The main reason for this belief is that correlations that emerge from aggregated data do not necessarily repeat individual correlations: they can be stronger than, weaker than, or equal to individual correlations (Ostroff, 1993). This may mean, for example, that talkativeness is related to optimism on the individual–level but not on the culture–level analysis. Cultural norms may impose restrictions on how extraversion is expressed and in some cultures it is impossible to judge on the basis of talkativeness how energetic and optimistic somebody is. In fact, the observed culture–level factors closely resemble individual–level factors, and can readily be identified with the FFM (McCrae, 2001; 2002).

Initially, it was thought that the close resemblance of the culture–level factor structure to the individual–level factor structure is surprising and tells us something meaningful about generalizability across cultures (Allik & McCrae, 2002; McCrae, 2002). In reality, however, the resemblance of these two different levels' factor structures is quite expectable due to purely statistical reasons. For example, when in a large dataset all subjects are reassigned randomly to arbitrary groups ("cultures"), it is very likely that such a grouping retains the individual–level factor structure. This result becomes less mysterious if we think that when two variables covary, groups that happen for any reason to be high on one will tend to be high on the other as well. When group–level data are analyzed, these two variables will still covary.

Thus, the replication of the individual–level factor structure on the ecological level of analysis, where each culture is represented as a single case is not surprising, but all deviations from the exact replication are. These are deviations from the individual–level factor structure that are informative, suggesting that grouping—in this case, belonging to different cultural groups—adds something beyond interindividual variation. This addition, of course, can be due to measurement error (e.g., different sampling procedures in different cultures) or due to meaningful cultural differences. The culture–level factor analysis of personality traits has revealed only modest deviations from the individual–level factor structure (McCrae, 2001; 2002). This seems to indicate that if the culture–level aggregation adds anything that goes beyond interindividual variation then it is relatively modest in its size.

COMPREHENSIVENESS OF THE BIG FIVE

It has been argued that the administration of an instrument like the NEO–PI–R in nonWestern and especially illiterate societies will not necessarily lead to the similar factor structure. The main reason is that nonWestern cultures may lack the notion that an individual is clearly distinctive from others. Indeed, Luria's expeditions to the mountain regions of Uzbekistan and Kyrgyzstan in 1931 to 1932 showed that inhabitants of villages who had no formal schooling were very inept or even incapable of analyzing their subjective qualities and personality (Luria, 1976). They frequently found it much easier to characterize other people than themselves and instead of describing their own personality they referred to the group to which they belonged, or they described external circumstancesinstead of their personality traits or character (p. 144 ff.) However, illiteracy by itself is not an obstacle for an appropriate personality description because questionnaires can be administered orally to the participants (e.g., Lima, 2002). Also, the inability to describe one's own personality does not imply that personality is absent or that external observers who have mastered the analysis of personality with sufficient sophistication cannot provide consistent descriptions of their targets. Like children, one needs to have the minimally required cognitive capacities to analyze and report of her on his own personality traits (Allik, Laidra, Realo, & Pullmann, 2004).

It is also questioned whether the Big Five is sufficient or not to describe personality traits in all cultures. Cultures may differ in traitedness and some traits typical to Western cultures are not appropriate, for example, to Eastern ones (cf. Church, 2000). This proposal is a cultural-level equivalent of Allport's idiographic hypothesis (Allport, 1937): not all traits are equally applicable to all cultures. For instance, it was found that among Chinese students there is a unique Interpersonal Relatedness factor that is absent in the Big Five and could not be consistently explained by a combination of these factors (Cheung et al., 2001). However, this supposedly indigenous interpersonal relatedness factor was replicated in European American samples (Cheung, Cheung, Leung, Ward, & Leong, 2003; Lin & Church, 2004)

indicating that these dimensions are not unique to Chinese populations. Contrary to previous interpretations of the Interpersonal Relatedness dimension in terms of interdependent self–construals, the dimension was only modestly correlated with the relational and collective aspects of the self, two aspects of interdependent self–construals (Lin & Church, 2004).

STEREOTYPES AND EXPERT OPINIONS

Can one interpret mean personality trait scores as an indicator of common personality features, and can one compare cultures on these traits? The aggregate personality scores vary systematically across cultures and demonstrate interpretable correlations with other culture–level indicators. This seems to suggest that aggregate personality scores make sense. However, these scores did not match the intuitive assessments of both laypersons and a panel of expert cross–cultural judges: Japan, for example, showed a low score for Conscientiousness, despite the widespread perception that the Japanese are an industrious people. Perhaps only the lack of sufficient ethnographic knowledge prevents us from being too surprised that the most purposeful and strong–willed individuals, according to their own self–reports at least, are the Telugu and Marathi Indians (McCrae, 2002). In order to collect expert opinions, McCrae (2001) asked eight prominent cross–cultural psychologists to identify the personality factor that had been used to rank 26 cultures based on their mean NEO–PI–R scores. He asked, for example, which personality factor is lowest among Hong Kong Chinese and South Koreans, but highest among Norwegians and Americans? Rather surprisingly, these experts all considered this a difficult task and were unable to identify factors at a better–than–chance level.

The most focused and systematic attempt to compare national stereotypes with the means of self–reported personality traits was undertaken by Church and Katigbak (2002). In this study, 43 judges who had lived in both the Philippines and the United States for a considerable time, rated whether Filipinos or Americans would tend to show a particular trait more. They wrote, "The backgrounds of our bicultural judges would seem to be fairly ideal for their task. Thus, if their judgments of average cultural differences are not valid, it is not clear whose judgments would be" (p. 149). These bicultural judges were in high agreement with each another, but were not consistent with the mean NEO–PI–R profiles (Church & Katigbak, 2002).

These negative findings suggest that mean trait levels and national stereotypes, even those held by experts, are not necessarily compatible. This is, of course, puzzling although there are several obvious reasons for the discrepancy. One attractive interpretation, which has already been proposed, is to discount beliefs about national character as a form of mythology. National stereotypes may be historical accidents or self–serving attributions or totems that serve the function of cementing group identity, rather than veridical accounts of aggregate personality traits (Allik & McCrae, 2002). This interpretation runs against numerous facts where national stereotypes are well motivated and may reflect actual differences between groups (Pea-

body, 1985). Another possibility is that except in extreme cases, it is in principle complicated if not impossible to reach correct judgments of mean personality profiles. The actual differences are rather small for most traits, and it may be impossible for individuals to conceive of them accurately. Studies show that cross–cultural differences in the mean values are rather small compared to interindividual differences within each culture (McCrae, 2001, 2002). The distance between the standings of any two individuals, who were selected randomly from the same culture, on the basic personality dimensions is almost certainly larger than the difference between any two who were arbitrarily chosen on the same personality dimension.

SUBSTANCE VERSUS STYLE

Although cross–cultural personality assessment has been concerned with establishing the validity of self-reports, it is more complicated to achieve than, for example, in clinical psychology. Besides careless responding, some people are more inclined to present themselves in a false or disproportionately favorable light. This self–enhancement is not necessarily a result of deliberate distortion. Socially desirable responding may also be a consequence of a general disposition to be cooperative and helpful, about which the test taker is not even fully aware. It is also very likely that different cultures endorse this general disposition to a different extent. Researchers and clinicians, who were particularly concerned with defensive or self–denial responding, have invented several methods on how to identify dishonest test takers. One strategy for taking this social desirability into account was the creation of validity scales. Edwards asked judges to rate the social desirability of items on the Minnesota Multiphasic Personality Inventory (MMPI) and selected a set of highly evaluated items (Edwards, 1957). He reasoned that people who answer affirmatively to the desirable items are likely those who are wishing to make a good impression. In order to extend the validity scales beyond clinical application, Crowne and Marlowe looked for everyday examples that were highly desirable but unlikely to be true (Crowne & Marlowe, 1964). It was also assumed that pretending to be honest is a relatively stable disposition that generalizes from one testing situation to another. Surprisingly, however, the use of validity scales did very little to improve the accuracy of self–report personality measures. Many studies have confirmed that when social desirability scores (Edwards's scale) were partialed out the correlation between self–ratings and ratings of knowledgeable others decreased, not increased (Dicken, 1963; McCrae & Costa, 1983; Piedmont, McCrae, Riemann, & Angleitner, 2000). This result seems to support an interpretation according to which social desirability scales measure substantive personality characteristics rather than a superficial style of responding (Block, 1965; McCrae & Costa, 1983; Smith & Ellingson, 2002). This interpretation is also consistent with meta–analytic findings that various social desirability scales demonstrate a considerable overlap with personality measures, particularly with Neuroticism (on average $r = -.37$),

Conscientiousness ($R = .20$), and Agreeableness ($r = .14$) (Ones, Viswesvaran, & Reiss, 1996).

A regularity seems to emerge from recent large–scale cross–cultural studies of personality. The strongest correlation with culture-level indicators like Gross National Product (GNP) is observed not with the mean scores of personality traits but with the scales supposedly measuring social desirability and different indicators of response bias. For example, van Hemert and colleagues (2002) discovered that the most striking finding was a substantial negative correlation between the EPQ Lie scale and indicators of economic prosperity: people from more wealthy countries are less prone to conformity and their answers depend less on the approval of other people. Thus, it is likely that socially desirable responding is not a style of responding that jeopardizes the cross–cultural equivalence of other personality dimensions, but social desirability may be a relatively stable personality characteristic that is influenced by cultural and socioeconomic conditions (Poortinga, van de Vijver, & van Hemert, 2002).

In order to evaluate the quality and consistency of the NEO–PI–R data collected from 50 cultures, McCrae and Terracciano developed an index based on 6 indicators (e.g., the number of response protocols with more than 40 missing responses, acquiescence bias, whether the native or the second language was used, etc.; McCrae et al., 2005). When countries were listed in descending order on the data quality index, the entries at the top of the list were from affluent, mostly Western nations, whereas those at the bottom were from underdeveloped nations. It seems that in less affluent countries careless or acquiescent responding and failure to understand the nuances of language are more frequent than in better-educated and more affluent countries. Indeed, the rank–order correlations between the data quality index and the Cronbach's alphas for the Big Five dimensions were in the range of from .63 to .81.

Perhaps the most informative is a recent study in which the Rosenberg Self–Esteem Scale (RSES) was translated into 28 languages and administered to 16,998 participants across 53 nations (Schmitt & Allik, 2005). It was shown that with few exceptions (primarily due to a single item), the factor structure of the RSES was invariant across nations. As expected, the RSES scores significantly correlated negatively with neuroticism and positively with extraversion within nearly all nations, providing some support for the cross–cultural equivalence of global self–esteem. All nations scored, on average, above the theoretical midpoint of the RSES, indicating that generally positive self–evaluation may be culturally universal. Although psychometric studies have generally supported the unidimensionality of the RSES (Carmines & Zeller, 1979; Marsh, 1996; O'Brien, 1985), there is still a pervasive tendency to respond to negatively worded items slightly differently than to positively worded items (Benson & Hocevar, 1985; Marsh, 1986). If people from certain types of cultures respond differently to the phrasing of negative items, this would indicate that direct cross–cultural comparisons on the RSES are confounded by a negative item bias. Although positively and negatively worded items of the RSES were strongly correlated on both

intracultural and intercultural levels, and were similarly related to external personality variables, the difference between aggregates of positive and negative items was clearly smaller in more developed nations. Although national self–esteem levels had no relation to the Human Development Index (HDI), nor to its three principal components—life expectancy at birth, adult literacy rate, and standard of living measured by GDP per capita—the difference between sums of positively and negatively worded items was strongly correlated with HDI, $r = -.53$, $p < .001$ and several other national–level indicators. Thus, in developed countries, where people have better education and live longer, negatively worded statements have slightly different meanings compared with the same message formulated in the positive wording. This type of inappropriate response to negatively worded items can be interpreted as a method artifact and may be responsible for the appearance of separate factors associated with positively and negatively worded items in previous exploratory factor analyses (Spector, Van Katwyk, Brannick, & Chen, 1997). Marsh showed that the size of the negative item bias varied substantially with age among several student samples. For example, the correlation between aggregates of positively and negatively worded items varied from close to zero for the youngest students to about .6 for the oldest students (Marsh, 1986).

The results of this study suggest that comparing the raw scores of the RSES across cultures has limited value, unless the inherent bias related to the different functioning of positively and negatively worded items has been taken into account (Schmitt & Allik, 2005). Together with other findings, this result seems to suggest that not always the mean scores of personality traits *per se* do not always have a strong and systematic relation to the national–level indicators. In many cases these are the "secondary" or higher-order properties of the personality questionnaires that have pervasive associations with how much people in their respective countries are educated, how long they are expected to live, and to what extent their economic life is secured. For example, McCrae (2001, 2002) observed that the standard deviations around the national mean scores are clearly ordered: the variation between individuals' mean scores is higher in economically advanced nations and lower in stagnated countries. Interestingly, the magnitude of gender differences follows a similar pattern, with the largest gender differences in European countries and the smallest in Asian nations (Costa et al., 2001). Analogously, measures of social desirability (van Hemert et al., 2002), acquiescence bias (Smith, 2004), and negative item bias (Schmitt & Allik, 2005) were strongly and systematically related to national indicators of socioeconomic development, often even more powerfully than the mean levels of personality traits themselves. This fact can be interpreted as a nuisance of intercultural comparisons that impedes comparability of cultures with respect to personality traits (Poortinga et al., 2002). According to another interpretation, however, responding in a socially favorable manner and other response biases have substantive cultural meaning and cannot be eliminated after correcting for the response style (Smith, 2004). Although cultures may differ in response style or self–presentation strategies, it is also

possible that the observed regularities represent real differences in personality. It is possible, for example, that readiness to agree with the expressed opinions and caution toward disparaging self–descriptions are, like sociability and conservativism, some aspects of the basic personality traits. For currently poorly understood reasons, these traits are more accurately "leaking out" through what is known as the acquiescence or the negative item bias than through the mean levels themselves. Unfortunately, our knowledge concerning this question is seriously limited by the lack of relevant facts and only future studies can build a stronger basis for conclusions.

SOME CONCLUSIONS

The data reviewed above seems to point in the direction of the psychic unity of humankind, at least that which concerns underlying personality dispositions. Although cultural differences appear to be enormous, anthropologists are arguing that there is a common basic culture of all humanity in all places and in all periods: "I would argue that in a real sense there is only one culture—the culture of humankind—and that societal differences with respect to cultural items are small" (D'Andrade, 2001). It is likely that cultural unity is based, partly at least, on the psychic unity of all people. Perhaps due to this unity, the quality of the translation of personality questionnaires, conditions of their administration, and sampling that seldom represents a given culture as a whole had little effect on the results: the dimensional structure of personality—five relatively independent groups of covarying traits—can be easily recovered from data that are not necessarily optimal for it, not only in Western cultures but also in nonWestern cultures, including Zimbabwe, Philippines, Malaysia, and China (McCrae, 2002). Today there is much more justification than there was 10 years ago to agree with a hypothesis that the pattern of covariation among personality traits is universal and extends across languages and cultures (McCrae & Costa, 1997). This also means that the FFM, which is the best summary of the pattern of covariation, is generalizable to psychiatric samples as well (Widiger & Costa, 2002; Widiger & Trull, 1992). Unfortunately, the number of cross–cultural comparisons of psychiatric samples is still very small. The first results, however, are promising and demonstrate that the FFM retains its main properties also in a nonWestern psychiatric sample like China (Yang et al., 1999).

In contrast to the generalizability of the FFM across languages and cultures, the comparison of the mean trait scores is much more problematic. The first wave of the large-scale cross–cultural projects revealed several paradoxes. Although the mean trait scores for different cultures were geographically organized, the pattern that emerged was not so easily interpretable. It not only contradicts a layperson's intuition but the knowledge of experts as well (Church & Katigbak, 2002). Take for example an extravert who according to the NEO–PI–R manual is sociable, assertive, talkative, and likes large groups and gatherings (Costa & McCrae, 1992). Which nation fits best into

this psychological portrait? It is not very likely that any of the Scandinavian countries are among the candidates from the most extraverted nations; according to widespread stereotypes, Scandinavians are typically quiet and reserved. In fact, according to self–reports the most extraverted individuals live in Denmark, Norway, and Sweden (McCrae, 2002). Thus, stereotypes do not agree with the ranking of the mean scores on personality traits.

Another surprising result that emerged from the first wave of studies is a relatively low agreement between parallel studies and instruments. The convergent correlations were disappointingly low even between two instruments belonging to the same family of personality measures. Particularly problematic were, however, discriminate correlations: it was not uncommon that one scale from one instrument has the strongest correlation not with its counterpart from another questionnaire but with some other scale. Even the pattern of convergent–discriminant correlations between two of the Big Five questionnaires, NEO–PI–R and BFI, was rather confused (Schmitt et al., submitted).

One possible explanation for poor agreement is that the mean-level differences in personality traits across cultures are very small in their magnitude. For example, the NEO–PI–R factor means of 36 cultures had standard deviations equal to about one–third of the magnitude of individual differences within culture (McCrae, 2002). Approximately the same ratio of between to within culture variation was obtained among 56 cultures where the BFI was administered (Schmitt et al., submitted). This means that even if the personality measures were perfectly accurate, knowing a person's nationality or cultural background would tell us rather little about his or her own personality (Allik & McCrae, 2002). The magnitude of the cross–cultural differences in the trait mean scores is smaller or comparable to the overall measurement accuracy that can be compromised by many factors, including biased sampling and small differences in translation, to say nothing about random measurement errors. In the result it is not very likely that the same "true" ordering of cultures on any of the Big Five personality traits would be exactly preserved in two parallel studies. If these differences are small, it is also improbable that people can accurately notice these tiny differences between different groups and aggregate them into stereotypes that reflect a true ordering on personality dimensions. Perhaps it may be possible in few extreme cases when two compared cultures are extraordinarily polarized, like, for example, North American and Japanese on the scale of self–esteem. But in all other cases the differences in the mean levels of personality traits may be too small to be consistently reproduced on two parallel occasions. Nevertheless, some more integral (e.g., the similarity between personality profiles) or subtle (e.g., the disparity between positively and negatively worded items) features can reveal a systematic relationship with relevant socioeconomic and geographic variables.

Fortunately, this may be good news for cross–cultural studies in general and for cross–cultural studies of personality disorders in particular. The prospect that for a proper psychological assessment of both normal and psychiatric samples it would be necessary to develop culture– or even subcul-

ture–specific norms looks frightening—not only due to the enormous amount of work that is required to develop these norms but even more due to the increasing fragmentation and particularization of the psychological measurement. A relatively modest size of cross–cultural differences in the mean values may imply that a reasonable scalar equivalence can be achieved, and all individuals, irrespective of their language and culture, can be represented in a common metric.

REFERENCES

Allik, J. (2002). Toward a theory of personality. In C. von Hofsten & L. Bäckman (Eds.), *Psychology at the turn of the millennium: Vol. 2: Social, developmental, and clinical psychology* (pp. 179–200). New York: Psychology Press.

Allik, J., Laidra, K., Realo, A., & Pullmann, H. (2004). Personality development from 12 to 18 years of age: Changes in mean levels and structure of traits. *European Journal of Personality, 18,* 445–462.

Allik, J., & McCrae, R. R. (2002). A Five–Factor Theory perspective. In R. R. McCrae & J. Allik (Eds.), *The Five–Factor Model of Personality across cultures* (pp. 303–322). New York: Kluwer Academic/Plenum Publishers.

Allik, J., & McCrae, R. R. (2004a). Escapable conclusions: Toomela (2003) and the universality of trait structure. *Journal of Personality and Social Psychology, 87,* 261–265.

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

Allport, G. W. (1937). *Personality: A psychological interpretation* New York: Holt, Rinehart & Winston.

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

Barrett, P. T., Petrides, K. V., Eysenck, S. B. G., & Eysenck, H. J. (1998). The Eysenck Personality Questionnaire: An examination of the factorial similarity of P, E, N, and L across 34 countries. *Personality and Individual Differences, 25* 805–819.

Benet–Martínez, V., & John, O. P. (1998). Los cinco grandes across cultures and ethnic groups: Multitrait multimethod analysis of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75,* 729–750.

Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement, 22* 231–240.

Bijnen, E. J., & Poortinga, Y. H. (1988). The questionable value of cross–cultural comparisons with the Eysenck Personality Questionnaire. *Journal of Cross Cultural–Psychology, 19*(2), 193–202.

Block, J. (1965). *The challenge of response set.* New York: Appleton–Century–Crofts.

Byrne, B. M., & Campbell, T. L. (1999). Cross–cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross–Cultural Psychology, 30,* 555–574.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Caprara, G. V., Barbaranelli, C., Bermudez, J., Maslach, C., & Ruch, W. (2000). Multivariate methods for the comparison of factor structures in cross–cultural research: An illustration with the Big Five Questionnaire. *Journal of Cross–Cultural Psychology, 31,* 437–464.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* Beverly Hills, CA: Sage.

Cavalli–Sforza, L. L., Menozzi, P., & Piazza, A. (1994). *The history and geography of human genes.* Princeton, NJ: Princeton University Press.

Cheung, F. M., Cheung, S. F., Leung, K., Ward, C., & Leong, F. (2003). The Eng-

lish version of the Chinese Personality Assessment Inventory. *Journal of Cross–Cultural Psychology, 34*, 433–452.

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

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

Cloninger, C. (2004). *Feeling good: The science of well–being* New York: Oxford University Press.

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, 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.

Crowne, D., & Marlowe, D. (1964). *The approval motive.* New York: Wiley.

D'Andrade, R. (2001). A cognitivist's view of the units debate in cultural anthropology. *Cross–Cultural Research, 35*, 242–257.

Dicken, C. (1963). Good impression, social desirability, and acquiescence as suppressor variables. *Education and Psychological Measurement, 23*, 699–720.

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research.* New York: Dryden.

Eysenck, H. J. (1990). Genetic and environmental contributions to individual differences: The three major dimensions of personality. *Journal of Personality, 58*, 245–261.

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

Fabrega, H. (1994). Personality disorders as medical entities: A cultural interpretation. *Journal of Personality Disorders, 8*, 149–167.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross–cultural comparisons of subjective Likert scales: The reference–group problem. *Journal of Personality and Social Psychology, 82*, 903–918.

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

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*(1), 99–114.

Katigbak, M. S., Church, A. T., Guanzon–Lapena, M. A., Carlota, A. J., & del Pilar, G. H. (2002). Are indigenous personality dimensions culture specific? Philippine inventories and the Five–Factor Model. *Journal of Personality and Social Psychology, 82*(1), 89–101.

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

Lima, M. P. (2002). Personality and culture: The Portuguese case. In R. R. McCrae & J. Allik (Eds.), *The Five–Factor Model of Personality across cultures* (pp. 249–260). New York: Kluwer Academic/Plenum Publishers.

Lin, E. J. L., & Church, A. T. (2004). Are indigenous Chinese personality dimensions culture–specific? An investigation of the Chinese Personality Assessment Inventory in Chinese American and European American samples. *Journal of Cross–Cultural Psychology, 35,* 586–605.

Little, T. D. (2000). On the comparability of constructs in cross–cultural research: A critique of Cheung and Rensvold. *Journal of Cross–Cultural Psychology, 31,* 213–219.

Luria, A. R. (1976). *Cognitive development: Its cultural and social foundations.* Cambridge, MA: Harvard University Press.

Lynn, R., & Martin, T. (1995). National differences for thirty–seven nations in extraversion, neuroticism, psychoticism and economic demo-

graphic and other correlates. *Personality and Individual Differences, 19,* 403–406.

Marian, V., & Kaushanskaya, M. (2004). Self–construal and emotion in bicultural bilinguals. *Journal of Memory and Language, 51,* 190–201.

Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive–developmental phenomenon. *Developmental Psychology, 22,* 37–49.

Marsh, H. W. (1996). Positive and negative global self–esteem: a substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70,* 810–819.

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

McCrae, R. R. (2002). NEO–PI–R data from 36 cultures. 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., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51,* 882–888.

McCrae, R. R., & Costa, P. T., Jr. (1996). Toward a new generation of personality theories: Theoretical context for the Five–Factor Theory. In J. S. Wiggins (Ed.), *The Five–Factor Model of Personality: Theoretical perspective* (pp. 51–87). New York: Guilford.

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

McCrae, R. R., & Costa, P. T., Jr. (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 Press.

McCrae, R. R., Terracciano, A., Khoury, B., Nansubuga F., Kne evič, G. D. J. D., Ahn, H., Ahn, C., De Fruyt, F., Gülgöz, S., Ruch, W., Arif Ghayur, M., Avia, M. D., Sánchez–Bernardos, M. L., Rossier, J., Dahourou, D., Fischer, R., Shakespeare–Finch, J., Yik, M. S. M., Smith, P. B., Angleitner, A., Ostendorf, F., Halim, M. S., Hřebíčková, M., Martin, T. A., Sineshaw, T., Sekowski, A., Klinkosz, W., Prentice, G., McRorie, M., Flores–Mendoza, C., Shimonaka, Y., Nakazato, K., Mastor, K. A., Barbaranelli, C., Alcalay, L., Simonetti, F., Pramila, V. S., Falzon, R., Lauri, M. A., Borg Cunen, M. A., Calleja, S. S., Pedroso de Lima, M., Bratko, D., Marušič, I., Allik, J., Realo, A., Abdel Khalek, A. M., Alansari, B. M., del Pilar, G. E. H., Ojedokun, A. O., Munyae, M., Budzinski, L., Oishi, S., Diener, E., Chittcharat, N., Wang, L., Beer, A., Humrichouse, J., Mortensen, E. L., Jensen, H. H., Jónsson, F. H., Ficková, E., Adamovová, L., Rus, V. S., Podobnik, N., Diaz–Loving, R., Leibovich, N. B., Schmidt, V., Reátegui, N., Brunner–Sciarra, M., Ayearst, L. E., Trobst, K. K., Matsumoto, D., Neubauer, A., Porrata, J., Rolland, J.–P., Petot, J.–M., & Camart, N. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology, 88,* 547–561.

O'Brien, E. J. (1985). Global self–esteem scales: Unidimensional or multidimensional? *Psychological Reports, 57,* 383–389.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing. *Journal of Applied Psychology, 81,* 660–679.

Ostroff, C. (1993). Comparing correlations based on individual–level and aggregated data. *Journal of Applied Psychology, 78,* 569–582.

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

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

Poortinga, Y. H., van de Vijver, F. J. R., & van Hemert, D. A. (2002). Cross–cultural equivalence of the Big Five. In R. R. McCrae & J. Allik (Eds.), *The Five–Factor Model of Personality across cultures* (pp. 281–302). New York: Kluwer Academic/Plenum Publishers.

Rogler, L. H. (1999). Methodological sources of cultural insensitivity in mental health research. *American Psychologist, 54,* 424–433.

Rolland, J. –P. (2002). The cross–cultural generalizability of the Five–Factor Model of Personality. In R. R. McCrae & J. Allik (Eds.), *The Five–Factor Model of Personality across cultures* (pp. 7–28). New York: Kluwer Academic/Plenum Publishers.

Schmitt, D., Allik, J. , McCrae, R. R., Benet–Martínez, V. et al. (submitted). The geographic distribution of Big Five personality traits: Patterns and profiles of human self–description across 56 nations. *Journal of Cross–Cultural Psychology.*

Schmitt, D. H., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self–Esteem Scale in 53 nations: Exploring the universal and culture–specific features of global self–esteem. *Journal of Personality and Social Psychology.*

Schmitt, D. P., Alcalay, L., Allik, J., Ault, L., Austers, I., Bennett, K. L., Bianchi, G., Boholst, F., Borg Cunen, M. A., Braeckman, J., Brainerd, E.G., Jr., Caral, L. G. A., Caron, G., Casullo, M. M., Cunningham, M., Daibo, I., De Backer, C., De Souza, E., Diaz–Loving, R., Diniz, G., Durkin, K., Echegaray, M., Eremsoy, E., Euler, H. A., Falzon, R., Fisher, M. L., Foley, D., Fry, D. P., Fry, S., Ghayur, M. A., Golden, D. L., Grammer, K., Grimaldi, L., Halberstadt, J., Herrera, D., Hertel, J., Hoffman, H., Hradilekova, Z., Hudek–Kene–evi, J., Jaafar, J., Jankauskaite, M., Kabangu–Stahel, H., Kardum, I., Khoury, B., Kwon, H., Laidra, K., Laireiter, A., Lakerveld, D., Lampart, A., Lauri, M., Lavallée, M., Lee, S., Leung, L. C., Locke, K. D., Locke, V., Luksik, I., Magaisa, I., Marcinkeviciene, J., Mata, A., Mata, R., McCarthy, B., Mills, M. E., Moreira, J., Moreira, S., Moya, M., Munyea, M., Noller, P., Opre, A., Panayiotou, A., Petrovic, N., Poels, K., Popper, M., Poulimenou, M., P'yatokha, V., Raymond, M., Reips, U., Reneau, S. E., Rivera–Aragon, S., Rowatt, W. C., Ruch, W., Rus, V. S., Safir, M. P., Salas, S., Sambataro, F., Sandnabba, K. N., Schulmeyer, M. K., Schütz, A., Scrimali, T., Shackelford, T. K., Shaver, P. R., Sichona, F., Simonetti, F., Sineshaw, T., Sookdew, R., Speelman, T., Spyrou, S., Sümer, H. C., Sümer, N., Supekova, M., Szlendak, T., Taylor, R., Timmermans, B., Tooke, W., Tsaousis, I., Tungaraza, F. S. K., Vandermassen, G., Vanhoomissen, T., Van Overwalle, F., Vanwesenbeek, I., Vasey, P. L., Verissimo, J., Voracek, M., Wan, W. W. N., Wang, T., Weiss, P., Wijaya, A., Woertment, L., Youn, G., & Zupanèiè, A. (2003). Universal sex differences in the desire for sexual variety: Tests from 52 nations, 6 continents, and 13 islands. *Journal of Personality and Social Psychology, 85,* 85–104.

Shweder, R. A. (1991). *Thinking through cultures: Expeditions in cultural psychology.* Cambridge, MA: Harvard University Press.

Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at a social desirability in motivating contexts. *Journal of Applied Psychology, 87,* 211–219.

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

Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management, 23,* 659–677.

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

Toomela, A. (2003). Relationships between personality structure, structure of word meaning, and cognitive ability: A study of cultural mechanisms of personality. *Journal of Personality and Social Psychology, 85,* 723–735.

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

van de Vijver, F. J. R., & Leung, K. (2000). Methodological issues in psychological research on culture. *Journal of Cross–Cultural Psychology, 31,* 33–51.

van de Vijver, F. J. R., & Leung, K. (2001). Personality in cultural context: Methodological issues. *Journal of Personality, 69,* 1007–1031.

van Hemert, D. A., van de Vijver, F. J. R., Poortinga, Y. H., & Georgas, J. (2002).

Structural and functional equivalence of the Eysenck Personality Questionnaire within and between countries. *Personality and Individual Differences, 33,* 1229–1249.

Widiger, T. A., & Costa, P. T. (2002). Five factor model personality disorder research. In P. T. Costa, & T. A. Widiger (Eds.), *Personality Disorders and the Five Factor Model of Personality* (2nd ed., pp. 59–87). Washington, DC: American Psychological Association.

Widiger, T. A., & Trull, T. J. (1992). Personality and psychopathology: An application of the Five–Factor Model. *Journal of Personality, 60,* 363–93.

Yang, J., McCrae, R. R., Costa, P. T., Jr., Dai, X., Yao, S., & Cai, T. G. B. (1999). Cross–cultural personality assessment in psychiatric populations: The NEO–PI–R in the People's Republic of China. *Psychological Assessment, 11,* 359–368.