The Cross-Cultural Generalizability of Zuckerman’s
Alternative Five-Factor Model of Personality

Jérôme Rossier
Institute of Psychology
University of Lausanne, Switzerland

Anton Aluja and Luis F. García
Department of Pedagogy and Psychology
University of Lleida, Spain

Alois Angleitner
Department of Psychology
University of Bielefeld, Germany

Vilfredo De Pascalis
Department of Psychology
University of Rome “La Sapienza,” Italy

Wei Wang
Department of Medical Psychology
Zhejiang University, China

Michael Kuhlman and Marvin Zuckerman
Department of Psychology
University of Delaware

The aim of this study was to analyze the cross-cultural generalizability of the Alternative five-factor model (AFFM; Zuckerman, Kuhlman, & Camac, 1988). The total sample was made up of 9,152 subjects from six countries: China, Germany, Italy, Spain, Switzerland, and the United States. The internal consistencies for all countries were generally similar to those found for the normative American sample. Factor analyses within cultures showed that the normative American structure was replicated in all cultures; however, the congruence coefficients were slightly lower in China and Italy. A similar analysis at the facet level confirmed the high cross-cultural replicability of the AFFM. Mean-level comparisons did not always show the hypothesized effects. The mean score differences across countries were very small.

The Five-factor model (FFM; McCrae & John, 1992) is a hierarchical model based on an empirical generalization about the covariation of personality traits (Digman, 1990). This model postulates that five dimensions named Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness adequately map these personality traits (McCrae & Costa, 1985, 1999). A large consensus exists about the FFM (Rossier, Meyer de Stadelhofen, & Berthoud, 2004), and these five dimensions are similar to the Big Five identified in numerous lexical studies (De Raad, 2000). According to the Five-factor theory (McCrae & Costa, 1999), these dimensions are biologically rooted, pointing to evidence that the five factors and their structure are heritable (McCrae, Jang, Livesley, Riemann, & Angleitner, 2001). Because all people share the same human genome, Five-factor theory claims that certain characteristics of traits,
including their structure and development, should be universal. Indeed, cross-cultural studies have shown that the FFM is replicable across cultures (e.g., Rolland, 2002). Universality of personality traits does not extend to mean levels, however, but McCrae (2002) suggested that these mean levels might vary across cultures in a predictable way (Rossier, Dahourou, & McCrae, 2005). Much research on the biological correlations of personality traits have used other dimensions, however, such as sensation seeking or impulsivity, which are not well represented in the FFM according to Zuckerman, Kuhlman, and Camac (1988), who developed an Alternative five-factor model (AFFM). The goal of this research is to evaluate the cross-cultural equivalence of the AFFM and to study if mean level differences according to the AFFM are similar to those observed according to the FFM.

To develop their AFFM, Zuckerman and colleagues (1988) studied the structure underlying 46 scales selected from 8 tests, which had been used as measures of temperament or involved in psychobiological studies of personality, and identified five replicable factors (Zuckerman, Kuhlman, Thornquist, & Kiers, 1991). Zuckerman, Kuhlman, Joireman, Teta, and Kraft (1993) developed the Zuckerman–Kuhlman Personality Questionnaire (ZKPQ) to measure the five dimensions of the AFFM: Impulsive Sensation Seeking (ImpSS), Neuroticism-Anxiety (N-Anx), Aggression-Hostility (Agg-Host), Activity (Act), and Sociability (Sy). These five dimensions were almost orthogonal as theoretically postulated \( r \leq |0.37| \). Zuckerman (2002) also suggested a hierarchical structure and three global factors that might be divided into two facet scales: The ImpSS factor contains a Sensation Seeking and an Impulsivity facet, the Act factor contains a Need for General Activity and a Need for Work Activity facet, and the Sy factor contains a Liking Lively Parties and Friends and an Intolerance of Social Isolation facet.

Aluja, García, and García (2002) compared the AFFM with the FFM using the ZKPQ and the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992). Considering only correlations above .40, they found that ImpSS correlated negatively with Conscientiousness \( r = -0.53 \), N-Anx correlated positively with Neuroticism \( r = 0.81 \), Agg-Host correlated negatively with Agreeableness \( r = -0.59 \), and Sy correlated positively with Extraversion \( r = 0.66 \). Thus, four dimensions of the AFFM and of the FFM appear quite similar, and the difference between these two models is the presence of a single dimension in each model that has no conceptual equivalent counterpart (Joireman & Kuhlman, 2004; Zuckerman et al., 1993). Indeed, the Openness to Experience domain is represented only poorly in the ZKPQ and the Activity factor is only poorly represented in the NEO-PI-R. Angleitner, Riemann, and Spinath (2004) studied a sample of 338 pairs of adult twins reared together and found that the heritability of the factors measured by the ZKPQ ranged from .43 to .51. These results are similar to those found for the FFM (McCrae et al., 2001).

Zuckerman (2002) reported several personality differences according to gender. In the American normative sample, men scored higher on ImpSS, Agg-Host, and Act, and women scored higher on N-Anx and Sy. The most robust differences were the ones for ImpSS and N-Anx. Gender differences for personality measurement were found to be highly consistent across cultures for the FFM (Costa, Terracciano, & McCrae, 2001). Concerning the AFFM, Aluja and colleagues (2002) reported gender differences similar to those reported by Zuckerman (2002) for a Spanish sample. However, De Pascalis and Russo (2003) observed only a difference on N-Anx in an Italian sample. Concerning adult trait development no data are available for the AFFM but for the FFM, Neuroticism, Extraversion, and Openness show usually moderate declines, whereas Agreeableness and Conscientiousness increase between age 18 and 30 (McCrae et al., 2002).

Cross-cultural equivalence is an important criterion of validity for personality models like the FFM and the AFFM that claim personality is biologically rooted and should be universal (Rossier, Dahourou, & McCrae, 2005). This equivalence might be studied at a structural level or at a metric or scalar level (Rossier, 2005). Over the past decade, a large number of cross-cultural studies have been conducted and confirmed the high structural equivalence of the FFM (Allik, 2005; McCrae et al., 2005). A review by Rolland (2002) based on a reanalyses of data from 16 cultures (from Asia, Europe, and America) compared the factorial structure of these countries with the American normative structure and found that the mean congruence coefficients were above .90 for all five dimensions of the FFM.

Comparison of mean scores for cultures should be carried out when structural, metric, and scalar equivalence have been demonstrated. The structural equivalence of the NEO-PI-R was found to be high, and several studies have shown that distortion due to translation is very small studying bilingual samples (McCrae, Yik, Trapnell, Bond, & Paulhus, 1998). Metric and scalar equivalence, however, are difficult to assess. Nevertheless, the attempt of comparing mean scores for cultures generally lead to meaningful results (Rossier, Dahourou, & McCrae, 2005). McCrae (2002, p. 112) observed, for example, that Americans score slightly lower on Neuroticism than Chinese, Germans, Italians, Spaniards, or German Swiss. However, mean score differences across cultures are generally small. These analyses usually have been done using the NEO-PI-R, and no study has tried to replicate these results using the ZKPQ.

The cross-cultural structural equivalence of the AFFM was never studied systematically although several translations of the ZKPQ are available (Zuckerman, 2002). Gomà-i-Freixanet, Valero, Punti, and Zuckerman (2004) made a first study of the structural replicability by comparing the structure obtained in their Catalan sample with the normative American structure. For women, congruence coefficients of all factors were above the factorial replication.
Mean and standard deviation for each country are reported in Table 1. The sample from China, Italy, Spain, Switzerland, and the United States consisted mainly of university students. For the United States, age was not available but is believed to range from 19 to 21, with a mode around 19 years. The German subjects pooled a sample of university students with participants in the Bielefeld twins study.

Measure

The ZKPQ (Zuckerman, 2002; Zuckerman et al., 1993) is made up of 99 true–false items and measures the following major scales: Impulsive Sensation Seeking (ImpSS), Neuroticism-Anxiety (N-Anx), Aggression-Hostility (Agg-Host), Activity (Act), Sociability (Sy), and Infrequency (Inf). ImpSS consists of 19 items (2 reversed keyed items) and can be divided according to Zuckerman (2002) in two facets Sensation Seeking (SS), which describes a general need for change, novelty, and excitement (i.e., “I like doing things just for the thrill of it”), and Impulsivity (Imp), which describes a lack of planning and acting on impulse (i.e., “I very seldom spend much time on the details of planning ahead”). N-Anx consists of 19 items (2 reversed keyed items) describing emotional upset and instability, tension, anxiety, lack of self-confidence, and sensitivity to criticism (i.e., “I sometimes feel edgy and tense”). Agg-Host consists of 17 items (6 reversed keyed items) describing the readiness to express verbal aggression or rude and antisocial behavior (i.e., “When I get mad, I say ugly things”). Act consists of 17 items (5 reversed keyed items) and can be divided into two facets: need for General Activity (GA), which describes a need for activity and the tendency to experience impatience and restlessness when being inactive (i.e., “I like to keep busy all the time”), and need for Work Effort (WE), which describes the tendency to invest a lot of energy for hard work and other tasks (i.e., “I like a challenging task more than a routine one”). Sy consists of 17 items (7 reversed keyed items) and can be divided into two facets scales: liking lively Parties and Friends (PF), which describes the preference of interacting with many people and of having many friends (i.e., “I am a very sociable person”), and Isolation Intolerance (II), which describes an intolerance for social isolation (i.e., “I need to feel that I am a vital part of a group”). Finally, Inf consists of 10 items (no reversed keyed item) and is a control scale used to eliminate subjects with possibly invalid records. Contents of these items are exaggerated socially desirable and are unlikely to be true (i.e., “I have always told the truth”). According to Zuckerman (2002) a score higher than 3 on this scale might indicate an invalid record. For this research Mandarin (Wu et al., 2000), English (Zuckerman et al., 1993), French (Rossier, Verardi, Massoudi, & Aluja, in press), German (Ostendorf & Angleitner, 1994), Italian (De Pascalis, & Russo, 2003), and Spanish (Aluja et al., 2002) versions of the ZKPQ have been used.

### TABLE 1

**Characteristics of the Samples**

<table>
<thead>
<tr>
<th>Country</th>
<th>Language</th>
<th>N</th>
<th>% female</th>
<th>% male</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>Chinese</td>
<td>1,247</td>
<td>69.5</td>
<td>30.5</td>
<td>21.5</td>
<td>8.0</td>
</tr>
<tr>
<td>Germany</td>
<td>German</td>
<td>985</td>
<td>80.1</td>
<td>19.9</td>
<td>37.1</td>
<td>13.3</td>
</tr>
<tr>
<td>Italy</td>
<td>Italian</td>
<td>808</td>
<td>62.4</td>
<td>37.6</td>
<td>25.1</td>
<td>6.2</td>
</tr>
<tr>
<td>Spain</td>
<td>Spanish</td>
<td>2,665</td>
<td>63.2</td>
<td>36.8</td>
<td>26.7</td>
<td>10.8</td>
</tr>
<tr>
<td>Switzerland</td>
<td>French</td>
<td>1,064</td>
<td>60.2</td>
<td>39.8</td>
<td>23.4</td>
<td>7.1</td>
</tr>
<tr>
<td>United States</td>
<td>English</td>
<td>2,383</td>
<td>61.8</td>
<td>38.2</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*Note. Dashes in cells indicate data were not available.*
Translations

Translations followed established guidelines regarding the translation of self-report inventories (Geisinger, 1994). For each language, the 99 ZKPQ items were translated by a team of experts in personality psychology and back-translated into English by an independent psychologist. These back-translations were checked by the authors of the ZKPQ and amendments were made, translated, back-translated, and reviewed. This process continued until the authors agreed with the different translations of the ZKPQ.

Procedure

In each country, subjects completed anonymously the appropriate language-version of the ZKPQ. This research complies with the ethical rules of the American Psychological Association.

Analyses

To compare the cross-cultural generalizability of the AFFM with the cross-cultural generalizability of the FFM, the analyses done were similar to the ones usually carried out for the NEO-PI-R (see, for example, Rolland 2002). For each country, reliability was assessed using Cronbach’s alphas, and gender differences and correlations with age were analyzed. The factorial structures were assessed by conducting exploratory factor analyses with varimax rotations at the factor and facet levels. To assess structural replicability, orthogonal Procrustes rotations were carried out with the American loadings as the target. Mean score comparisons were conducted on gender corrected z-scores and data restricted to the college-age subjects (18 to 25 years) in order to avoid bias due to age (see, for example, McCrae, 2002). All analyses were conducted on the entire sample including subjects with a score higher than 3 on the Inf scale. Participants had an Inf score > 3 with the following frequencies: China = 1.8%, Germany = 13.7%, Italy = 18.6%, Spain = 14.8%, Switzerland = 19.3%, and United States = 5.6%. This control scale seems very sensitive to culture and might thus not be used in cross-cultural studies (Rossier, 2005).

RESULTS

Internal Consistency and Influence of Gender and Age

In the full sample, the internal consistencies were similar to those found in the American sample. For the five global scales they were .79, .85, .72, .73, and .76 for ImpSS, N-Anx, Agg-Host, Act, and Sy, respectively. For the six facet scales they were .76, .66, .71, .56, .69, and .69 for SS, Imp, GA, WE, PF, and II respectively. Table 2 presents the internal consistencies of each scale for each country. For the five main factors, the internal consistencies seemed slightly lower in China (M = .68) and Italy (M = .73) compared with Germany (M = .80), Spain (M = .79), Switzerland (M = .78), and the United States (M = .79). In Italy the internal consistency was lower especially for Agg-Host. At the facet level the same trend was observed with lower internal consistencies in China (M = .56) and Italy (M = .63) compared with Germany (M = .71), Spain (M = .70), Switzerland (M = .78).

<table>
<thead>
<tr>
<th>Sample</th>
<th>ImpSS</th>
<th>Imp</th>
<th>Total</th>
<th>N-Anx</th>
<th>Agg-Host</th>
<th>GA</th>
<th>WE</th>
<th>Total</th>
<th>PF</th>
<th>II</th>
<th>Total</th>
<th>Inf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistencies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>.63</td>
<td>.59</td>
<td>.68</td>
<td>.77</td>
<td>.64</td>
<td>.57</td>
<td>.49</td>
<td>.64</td>
<td>.45</td>
<td>.63</td>
<td>.65</td>
<td>-.10</td>
</tr>
<tr>
<td>Germany</td>
<td>.79</td>
<td>.67</td>
<td>.81</td>
<td>.88</td>
<td>.74</td>
<td>.77</td>
<td>.61</td>
<td>.78</td>
<td>.71</td>
<td>.71</td>
<td>.78</td>
<td>.52</td>
</tr>
<tr>
<td>Italy</td>
<td>.80</td>
<td>.65</td>
<td>.83</td>
<td>.85</td>
<td>.57</td>
<td>.67</td>
<td>.52</td>
<td>.71</td>
<td>.64</td>
<td>.52</td>
<td>.68</td>
<td>.51</td>
</tr>
<tr>
<td>Spain</td>
<td>.79</td>
<td>.71</td>
<td>.82</td>
<td>.86</td>
<td>.74</td>
<td>.74</td>
<td>.54</td>
<td>.75</td>
<td>.68</td>
<td>.73</td>
<td>.78</td>
<td>.52</td>
</tr>
<tr>
<td>Switzerland</td>
<td>.75</td>
<td>.64</td>
<td>.79</td>
<td>.87</td>
<td>.72</td>
<td>.74</td>
<td>.59</td>
<td>.75</td>
<td>.71</td>
<td>.68</td>
<td>.76</td>
<td>.47</td>
</tr>
<tr>
<td>United States</td>
<td>.76</td>
<td>.67</td>
<td>.79</td>
<td>.85</td>
<td>.77</td>
<td>.74</td>
<td>.59</td>
<td>.75</td>
<td>.75</td>
<td>.71</td>
<td>.79</td>
<td>.55</td>
</tr>
<tr>
<td>Mean</td>
<td>.75</td>
<td>.66</td>
<td>.79</td>
<td>.85</td>
<td>.70</td>
<td>.71</td>
<td>.56</td>
<td>.73</td>
<td>.66</td>
<td>.66</td>
<td>.74</td>
<td>.41</td>
</tr>
<tr>
<td>Correlations with age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>-.20</td>
<td>-.03</td>
<td>-.16</td>
<td>.00</td>
<td>-.05</td>
<td>.19</td>
<td>-.05</td>
<td>.09</td>
<td>-.09</td>
<td>-.17</td>
<td>-.16</td>
<td>.23</td>
</tr>
<tr>
<td>German</td>
<td>-.23</td>
<td>-.01</td>
<td>-.17</td>
<td>-.21</td>
<td>-.12</td>
<td>.20</td>
<td>.02</td>
<td>.15</td>
<td>-.12</td>
<td>-.14</td>
<td>-.16</td>
<td>.22</td>
</tr>
<tr>
<td>Italy</td>
<td>-.15</td>
<td>-.04</td>
<td>-.13</td>
<td>-.13</td>
<td>-.04</td>
<td>.05</td>
<td>-.01</td>
<td>.02</td>
<td>-.11</td>
<td>-.15</td>
<td>-.15</td>
<td>.15</td>
</tr>
<tr>
<td>Spain</td>
<td>-.33</td>
<td>-.14</td>
<td>-.30</td>
<td>-.12</td>
<td>-.17</td>
<td>.02</td>
<td>-.11</td>
<td>-.04</td>
<td>-.23</td>
<td>-.27</td>
<td>-.30</td>
<td>.20</td>
</tr>
<tr>
<td>Switzerland</td>
<td>-.17</td>
<td>-.09</td>
<td>-.16</td>
<td>-.09</td>
<td>-.05</td>
<td>.07</td>
<td>.12</td>
<td>.11</td>
<td>-.11</td>
<td>-.22</td>
<td>-.19</td>
<td>.02</td>
</tr>
<tr>
<td>United States</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>-.22</td>
<td>-.06</td>
<td>-.18</td>
<td>-.11</td>
<td>-.07</td>
<td>.11</td>
<td>-.01</td>
<td>.07</td>
<td>-.13</td>
<td>-.19</td>
<td>-.19</td>
<td>.16</td>
</tr>
</tbody>
</table>

Note. ImpSS: Impulsive Sensation Seeking; SS: Sensation Seeking; Imp: Impulsivity; N-Anx: Neuroticism-Anxiety; Agg-Host: Aggression-Hostility; Act: Activity; GA: General Activity; WE: Work Effort; Sy: Sociability; PF: Parties and Friends; II: Isolation Intolerance; Inf: Infrequency; Dashes in cells indicate data were not available.
.69), and the United States (M = .70). Interestingly, in China the internal consistency was especially low for Inf, which is not a personality dimension but a control scale.

For analyzing gender differences, the effect of age was controlled by restricting the data to the college-age subjects (18 to 25 years). This restriction resulted in a small number of men in Germany (N = 37). Globally across all samples men had a significantly higher score on ImpSS, t(6539) = 11.23, p < .001 (d = .29). This difference was significant in Spain (d = .32), Switzerland (d = .25), and the United States (d = .35), but not in China (d = .07), Italy (d = .16), and Germany (d = .20). These discrepancies are due to the fact that men scored consistently higher on SS (ds ranged from .24 to .30) in all countries and that the results were more mixed concerning Imp. Indeed, men scored significantly higher on Imp in Spain (d = .22) and the United States (d = .34), but women scored significantly higher on Imp in China (d = .21), and no significant difference was found in Germany (d = .09), Italy (d = .04), and Switzerland (d = .10). Globally, women scored significantly higher on N-Anx, t(6547) = 25.47, p < .001 (d = .67), and this difference was strongly consistent across countries (ds ranged from .47 to .92). Globally, the difference for Agg-Host, where men scored slightly higher, was associated with a negligible effect size, t(6540) = 3.88, p < .001 (d = .10).

Actually, men scored significantly higher only in the United States (d = .24). Globally, the difference for Act, where men scored slightly higher, was associated with a negligible effect size, t(6546) = 5.43, p < .001 (d = .14). This difference was nonsignificant in Germany (d = .30) and Italy (d = .08), significant but associated with a negligible effect size in Spain (d = .12) and the United States (d = .12), and significant with modest size in China (d = .29) and Switzerland (d = .28). These discrepancies were due to the fact that men scored consistently higher on WE (ds ranged from .22 to .62) in all countries and that no significant difference was observed on GA in all countries (ds ranged from .00 to .11). Globally, no significant difference was observed for Sy, t(6545) = -1.66, p > .05 (d = .04), and men had only significantly higher scores in China (d = .33) and women in Spain (d = .19) on this factor. Finally, men scored consistently higher on Inf, t(6548) = 17.18, p < .001 (d = .43), and this occurred in all countries (ds ranged from .31 to .53). To summarize these gender differences, women scored consistently higher on N-Anx, and men scored consistently higher on SS, WE, and Inf.

Globally, age correlated negatively with ImpSS (r = -.26, due to a negative correlation with SS, r = -.31), N-Anx (r = -.16), and Sy (r = -.24, due to correlations with both PF [r = -.22] and II [r = -.17]), and positively with Inf (r = .14). All other correlations were negligible (r < .110). Correlations for each country except the United States are presented in Table 2. Correlations between age and ImpSS, SS, Sy, PF, and II, were especially consistent across countries.

Replicability of the Factor Structure Across Countries

On the full sample, a principal component factor analysis with varimax rotation of the 89 items (excluding items from the Inf scale) allowed extracting five factors using Cattell’s criterion. These factors explained 24.94% of the total variance. The first six eigenvalues were 6.52, 6.14, 3.73, 3.16, 2.66, and 2.26. A one-to-one association was observed between the five factors and the five main scales of the ZKPQ. Factor 1 correlated with N-Anx (r = .97), factor 2 with ImpSS (r = .94), factor 3 with Act (r = .94), factor 4 with Sy (r = .90), and factor 5 with Agg-Host (r = .93).

In order to assess the structural replicability of the AFFM, a principal component factor analysis with varimax rotation of the 89 items was conducted for each country. For China, the five-factor solution explained 20.97% of the total variance and the first six eigenvalues were 6.09, 4.07, 3.68, 2.65, 2.18, and 1.82. For Germany, the five-factor solution explained 28.48% of the total variance and the first six eigenvalues were 7.97, 6.97, 3.98, 3.38, 3.05, and 2.40. For Italy, the five-factor solution explained 25.83% of the total variance and the first six eigenvalues were 7.40, 6.39, 3.82, 2.72, 2.67, and 2.09. For Spain, the five-factor solution explained 26.94% of the total variance and the first six eigenvalues were 7.04, 6.77, 3.92, 3.53, 2.72, and 2.20. For Switzerland, the five-factor solution explained 27.06% of the total variance and the first six eigenvalues were 7.49, 5.97, 4.19, 3.46, 2.97, and 2.21. For the United States, the five-factor solution explained 26.36% of the total variance and the first six eigenvalues were 7.06, 5.76, 4.10, 3.54, 3.00, and 2.32. A one-to-one association was observed between the five factors and ZKPQ’s five main scales for each country.

To compare the structure observed for the six countries, each loading matrix (except the American matrix) was subjected to an orthogonal Procrustes rotation using the American matrix as the target. Convergence after rotation was determined by congruence coefficients, which range from -1 to +1 (Haven & ten Berge, 1977). Congruence coefficients higher than .90 indicated a high structural equivalence, congruence coefficients ranging from .81 to .90 indicated a borderline structural equivalence, and congruence coefficients equal to or below .80 indicated a poor structural equivalence (McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). A congruence coefficient was calculated for the complete matrix, giving an estimation of the global structural equivalence (total congruence coefficient) for each column, giving an estimation of the structural equivalence for each factor (factor congruence coefficient), and for each line, giving an estimation of the structural equivalence for each item (variable congruence coefficient; e.g., Rossier, Dahourou, & McCrae, 2005). We found a mean total congruence coefficient (CC) of .92 (see Table 3). The mean CCs for factors ranged from .88 to .94 (Mdn = .92). The mean CCs for items ranged from .69 to .98 (Mdn = .94). Sixty-three items
were associated with a high CC, 19 items were associated with a borderline CC, and 7 items were associated with a poor CC.

For China, the total CC was .88, the factor CCs ranged from .84 to .91 (Mdn = .86), and item CCs ranged from −.46 to .99 (Mdn = .92). Fifty-three items were associated with a high CC, 18 with a borderline CC, and 18 with a poor CC. For Germany, the total CC was .95, the factor CCs ranged from .93 to .97 (Mdn = .94), and item CCs ranged from .81 to 1.00 (Mdn = .97). Eighty-two items were associated with a high CC and 9 with a borderline CC. For Italy, the total CC was .89, the factor CCs ranged from .78 to .93 (Mdn = .91), and item CCs ranged from .16 to 1.00 (Mdn = .94). Sixty-four items were associated with a high CC, 11 with a borderline CC, and 14 with a poor CC. For Spain, the total CC was .92, the factor CCs ranged from .88 to .95 (Mdn = .91), and item CCs ranged from .56 to 1.00 (Mdn = .95). Sixty-three items were associated with a high CC, 19 with a borderline CC, and 7 with a poor CC. For Switzerland, the total CC was .96, the factor CCs ranged from .93 to .97 (Mdn = .96), and item CCs ranged from .73 to 1.00 (Mdn = .97). Eighty-two items were associated with a high CC, 6 with a borderline CC, and one with a poor CC.

For China, 8 items seemed to be particularly problematic (CC ≤ .70): items 9, 14, 21, 38, 47, 53, 82, and 85. When removing these 8 items, the total CC was .91, the factor CCs ranged from .87 to .93 (Mdn = .91), and item CCs ranged from .71 to 1.00 (Mdn = .93). For Italy, 11 items seemed to be problematic: items 12, 14, 36, 53, 54, 67, 72, 77, 82, 88, and 89. When removing these 11 items, the total CC was .92, the factor CCs ranged from .88 to .95 (Mdn = .91), and item CCs ranged from .75 to .99 (Mdn = .95). For Spain, 3 items seemed to be problematic: items 14, 82, and 89. When removing these 3 items, the total CC was .96, the factor CCs ranged from .94 to .97 (Mdn = .95), and item CCs ranged from .82 to 1.00 (Mdn = .97). No problematic item was found for Germany and Switzerland.

Factor replicability is usually assessed by factoring facet-level scales, not items (e.g., Rolland, 2002). Because of the brevity and unbalanced nature of single items, it is often difficult to replicate factor structures at the item level. That makes the observed replicability even more striking.

### Replicability of the Facet Scales Structures Across Countries

To assess cross-cultural structural replicability at the facet level, three separate series of principal component factor analysis with varimax rotation of the items of ImpSS, Act, and Sy were carried out. For each factor and for each country, two factors were extracted. The loading matrix for each factor and country except the United States was subjected to an orthogonal Procrustes rotation using the American matrix as the target. Concerning ImpSS, the mean total CC for the two-factor solutions and the mean CCs for facets were above .90 (see Table 4). The mean CCs for items ranged from .92 to 1.00 (Mdn = .98). All items were associated with a high CC. Concerning Act, the mean total CC and the mean CCs for facets were above .90. The mean CCs for items ranged from .84 to 1.00 (Mdn = .98). Fifteen items were associated with a high CC. Concerning Sy, the mean total CC and the mean CCs for facets were also above .90. The mean CCs for items ranged from .61 to 1.00 (Mdn = .97). Twelve items were associated with a high CC, 4 with a borderline CC, and 1 with a poor CC. Across countries, except for PF in China and II in Italy, all factor congruence coefficients suggested replication by Haven and ten Berge’s criterion (1977).

### TABLE 3
**Congruence Coefficients for Factors, Comparing the American Sample With the 5 Other Samples**

<table>
<thead>
<tr>
<th>Sample</th>
<th>ImpSS</th>
<th>N-Anx</th>
<th>Agg-Host</th>
<th>Act</th>
<th>Sy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>.86</td>
<td>.86</td>
<td>.88</td>
<td>.91</td>
<td>.84</td>
<td>.88</td>
</tr>
<tr>
<td>Germany</td>
<td>.95</td>
<td>.97</td>
<td>.93</td>
<td>.94</td>
<td>.94</td>
<td>.95</td>
</tr>
<tr>
<td>Italy</td>
<td>.91</td>
<td>.93</td>
<td>.78</td>
<td>.91</td>
<td>.82</td>
<td>.89</td>
</tr>
<tr>
<td>Spain</td>
<td>.91</td>
<td>.95</td>
<td>.88</td>
<td>.90</td>
<td>.92</td>
<td>.92</td>
</tr>
<tr>
<td>Switzerland</td>
<td>.96</td>
<td>.97</td>
<td>.93</td>
<td>.96</td>
<td>.94</td>
<td>.96</td>
</tr>
<tr>
<td>Mean</td>
<td>.92</td>
<td>.94</td>
<td>.88</td>
<td>.92</td>
<td>.89</td>
<td>.92</td>
</tr>
</tbody>
</table>

Note. ImpSS: Impulsive Sensation Seeking; N-Anx: Neuroticism-Anxiety; Agg-Host: Aggression-Hostility; Act: Activity; Sy: Sociability.

### TABLE 4
**Congruence Coefficients for Facet Scales, Comparing the American Sample With the 5 Other Samples**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sensation seeking</th>
<th>Impulsivity</th>
<th>General activity</th>
<th>Work effort</th>
<th>Parties and Friends</th>
<th>Isolation intolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>.95</td>
<td>.93</td>
<td>.88</td>
<td>.93</td>
<td>.83</td>
<td>.90</td>
</tr>
<tr>
<td>Germany</td>
<td>.98</td>
<td>.97</td>
<td>.99</td>
<td>.97</td>
<td>.95</td>
<td>.96</td>
</tr>
<tr>
<td>Italy</td>
<td>.96</td>
<td>.92</td>
<td>.97</td>
<td>.88</td>
<td>.94</td>
<td>.80</td>
</tr>
<tr>
<td>Spain</td>
<td>.98</td>
<td>.97</td>
<td>.99</td>
<td>.94</td>
<td>.97</td>
<td>.97</td>
</tr>
<tr>
<td>Switzerland</td>
<td>.98</td>
<td>.97</td>
<td>.98</td>
<td>.97</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td>Mean</td>
<td>.97</td>
<td>.95</td>
<td>.96</td>
<td>.94</td>
<td>.94</td>
<td>.92</td>
</tr>
</tbody>
</table>
Mean Score Comparison

In order to control the effect of age and gender, the data were restricted to the college-age subjects (18 to 25 years) and all scores were standardized by calculating z-scores using the Ms and SDs from the gender-appropriate American norms. The gender corrected standardized scores for each factor and facet can be seen in Table 5. The mean standardized SDs were all smaller than the one observed in the American sample. The mean standardized SD was especially low for China. This value was also lower for Italy compared with the other European countries.

The difference between countries was assessed with one-way ANOVAs on the gender-corrected standardized scores. ImpSS was significantly affected by the country variable, F(5, 6484) = 24.89, p < .001, and was lower in China, Germany, and Italy than in the United States. This difference seemed mainly due to the SS facet, F(5, 6488) = 32.40, p < .001, compared with the Imp facet, F(5, 6494) = 9.30, p < .001. The country variable had a significant impact on N-Anx, F(5, 6491) = 12.20, p < .001, which was most notably lower in Germany. Agg-Host was significantly affected by the country variable, F(5, 6484) = 38.31, p < .001, and was lower in China, Italy, and Switzerland than in the United States. Act was significantly affected by the country variables, F(5,6490) = 34.58, p < .001, and was lower in Switzerland and higher in China, Germany, Italy, and Spain than in the United States. This difference seemed mainly due to GA, F(5, 6495) = 60.45, p < .001, compared with WE, F(5, 6494) = 17.71, p < .001. GA was lower in Switzerland and higher in China, Italy, and Spain, and WE was higher in Germany. The country variable also had a significant impact on Sy, F(5, 6491) = 55.11, p < .001. Sy was lower in China, Germany, Italy, and Switzerland than in the United States. This difference seemed mainly due to II, F(5, 6495) = 91.99, p < .001, compared with PF, F(5, 6493) = 30.99, p < .001. II was especially low in China, Italy, and Switzerland, whereas PF was especially low in China and Germany. Inf was higher in all countries compared with the United States and was significantly affected by the country variable, F(5, 6492) = 98.14, p < .001.

The impact of the country variable on the factors and facets were generally small. Indeed, all differences were associated with a small or negligible effect size except II, which was associated with a medium effect size. On average, the country variable accounted for 2.48% of the variability of the five main personality factors and for 2.98% of the variability of the six facet scales.

DISCUSSION

The observed internal consistencies across countries were all similar to those found in the American sample (Zuckerman, 2002). However, they were slightly lower in China and in Italy. Several previous studies using the NEO-PI-R with Asian or African samples also have found lower internal consistencies. These have been attributed to smaller standard deviations, which can lower alphas (McCrae, 2002). In this study, standard deviations were indeed smaller in the Chinese sample. Rossier and colleagues (2005) have suggested that this range restriction could be due to a bias toward more socially desirable responding or that, in collectivistic cultures, behavior might be determined more by social context than by traits. For all countries, the scale with the highest internal reliability was N-Anx. The scale with the lowest internal reliability was Act in the United States and Agg-Host in Germany, Italy, Spain, and Switzerland. In China, Agg-Host and Act shared the lowest internal reliability. Gomá-i-Freixanet and colleagues (2004) reported that the internal reliability for Agg-Host is usually lower in translated versions of the ZKPQ. This seems to be the case in China and in Italy. At the facet level, reliabilities were also similar to those found in the American sample, with lower values for China and Italy.

The main gender differences observed in this study, higher scores on N-Anx for women and higher scores on SS, WE, and Inf for men, were very consistent across countries, confirming previous results about gender differences and the AFFM (Aluja, García, & García, 2002; De Pascalis & Russo, 2003; Zuckerman, 2002). Moreover, this consistency across cultures confirmed results obtained for the FFM (Costa, Terracciano, & McCrae, 2001). For example, women scored
CROSS-CULTURAL GENERALIZABILITY OF THE AFFM

This research was supported by a Swiss National Science Foundation grant number 100013-108290.
REFERENCES


Jérôme Rossier
Institute of Psychology
University of Lausanne
Anthropole-3127
CH-1015 Lausanne, Switzerland
Email: Jerome.Rossier@unil.ch

Received June 22, 2006
Revised November 27, 2006