European Journal of Personality
Eur. J. Pers. **16**: S25–S41 (2002)
DOI: 10.1002/per.450

Replicable Types and Subtypes of Personality: Spanish NEO-PI Samples

BETTINA BOEHM[1]*, JENS B. ASENDORPF[1] and MARIA D. AVIA[2]

[1] *Humboldt-Universität zu Berlin, Germany*
[2] *Universidad Complutense, Madrid, Spain*

Abstract

Three major personality prototypes were derived on the basis of a Big Five instrument (NEO-PI) by means of replicated cluster analysis in two Spanish samples (a sample from the general population and a student sample). The replicability of the three prototypes within and their consistency between the two samples were evaluated. In addition, subtypes were analysed in a similar way. Finally, the relation between prototype assignment and level of education was examined in the sample from the general population. Within-study replicability was satisfactory only for the student sample. Comparison with the results for a similar instrument (NEO-PI-R) applied to a German sample showed satisfactory consistency only for the student sample. Discussion centres on the strong sample dependency of the results. Copyright © 2002 John Wiley & Sons, Ltd.

INTRODUCTION

From a person-centred profile perspective, personality prototypes describe types of individuals that share the same basic personality profile. Individual personality profiles are derived on the basis of the covariation of a number of variables, for example the Big Five factors of personality, and are subsequently classified into groups of similar profiles. It is unlikely that an individual's profile will correspond exactly to a prototypical profile because a prototypical profile is an average of many profiles.

Research on empirically derived personality types first began at the beginning of the 1970's with Block's pioneering work *Lives through time* (1971). In this work Block reported on a longitudinal study carried out in a sample of men and women from early adolescence to adulthood over a period of 22 years. By means of Q-factor analysis, Block found five personality prototypes for the male and six for the female sample. In the male sample three prototypes showed high stability over time, namely the resilient, the

*Correspondence to: Jens B. Asendorpf, Institut für Psychologie, Humboldt-Universität zu Berlin, Oranienburger Strasse 18, 10178 Berlin, Germany. E-mail: asen@rz.hu-berlin.de

Received 30 October 2001
Accepted 31 December 2001

overcontrolled, and the undercontrolled prototypes. The resilient type was characterized by a pattern of well adapted behaviour and social competence, the overcontrolled type by high emotional constraint, and the undercontrolled type by high impulsiveness and an inability to delay gratification.

Robins, John, Caspi, Moffitt, and Stouthamer-Loeber (1996) were the first to replicate the three prototypes. This was done in a sample of boys aged 12–13 years by means of replicated Q-factor analysis on the basis of Q-sorts, and showed that the three prototypes were sufficiently stable within the sample. When they related the three prototypes to the Big Five, they found that the resilient boys scored low on Neuroticism and high on all other Big Five factors, that the overcontrolled boys scored high on Neuroticism and Agreeableness and low on Extraversion, and that the undercontrolled boys had elevated scores on Neuroticism and particularly low scores on Conscientiousness. The three prototypes and their relation to the Big Five was also replicated in a longitudinal German sample of boys and girls aged 4–12 years (Asendorpf & van Aken, 1999). In addition, this study found that the three prototypes were sufficiently consistent with the three prototypes found in samples of Dutch and Icelandic children (Hart, Hofmann, Edelstein, & Keller, 1997; Haselager, van Lieshout, & van Aken, unpublished data).

Research on replicable prototypes in adulthood was initiated by Asendorpf, Borkenau, Ostendorf, and van Aken (2001), who used the method of replicated cluster analysis to search for prototypes on the basis of two Big Five instruments, the German version of the NEO-FFI (Borkenau & Ostendorf, 1993) and bipolar adjective scales derived from a larger pool of German personality-describing adjectives (Ostendorf, 1990). Replicability within samples and consistency across samples were measured by a quantitative coefficient, Cohen's kappa. In two German adult studies they found three replicable prototypes, which were identified as the resilient, overcontrolled, and undercontrolled types. They could also show that the three prototypes were consistent across the two adult samples and instruments and even with a sample of children that had been analysed by a different method.

In their article, Asendorpf et al. (2001) suggested that future studies should evaluate potential factors that may influence the derivation of personality prototypes (e.g. culture, method, instrument, rater, and age of subjects), by varying one factor and holding all other factors constant. The present research is in line with this suggestion. The influence of culture on the three prototypes is studied here by contrasting prototypes and their subtypes derived using similar methods from German and Spanish samples. Two Spanish adult samples aged 20 to 30 years are analysed by means of replicated cluster analysis of profiles based on a Spanish version of the NEO-PI (Costa & McCrae, 1985). The Spanish types are compared with those derived by Schnabel, Asendorpf, and Ostendorf (this special issue) from German NEO-PI-R data, which were similar in most respects (i.e. method, instrument, rater and age), and the consistency between the Spanish and the German results is examined.

A prerequisite for the search for replicable prototypes on the basis of the Big Five in Spanish samples is that the five-factor model of personality (see e.g. Goldberg, 1981; John & Srivastava, 1999; McCrae & Costa, 1987) shows sufficient construct validity in this sample. In a study on lexically derived factors from the Spanish lexicon, it could be shown that when highly evaluative terms were included in the variable pool a seven factor solution was optimal in factor analytic procedures (Benet-Martínez & Waller, 1997). Four of the seven factors could broadly be identified as Big Five factors (Extraversion, Conscientiousness, Agreeableness, and Culture). Another Big Five factor, Neuroticism,

was spread across the Extraversion and Conscientiousness factors. Saucier, Hampson, and Goldberg (2000) suggested analysing a three-factor solution in the Spanish data because the 'Big Three' (Extraversion, Agreeableness, and Conscientiousness; Peabody & Goldberg, 1989) were clearly recognizable. This hypothesis is in line with a general agreement among personality psychologists that the Big Three can be generalized to most modern languages better than the Big Five (Saucier et al., 2000).

Studies evaluating the Big Five in personality questionnaires in Spain have shown that the internal structure of the instruments has been successfully kept in Spanish versions (see e.g. Avia, Sanz, Sánchez-Bernardos, Martínez-Arias, Silva, & Graña, 1994; Benet-Martínez & John, 1998; Silva, Avia, Sanz, Martínez-Arias, Graña, & Sánchez-Bernardos, 1994). Factor analyses with the Spanish version of the NEO-PI (Silva et al., 1994; Avia et al., 1994), for example, showed that the Spanish five-factor structure was very similar to that obtained by Costa and McCrae (1985). Simultaneous component analysis, which was used to calculate the optimal factor variance for both the Spanish and the North-American NEO-PI data, showed that the total variance accounted for by factors was virtually identical in these two languages. Drifts of individual facets from the intended factor, in particular Extraversion, were not specific to the Spanish samples but could also be observed in two North-American samples (McCrae & Costa, 1989; Church & Burke, 1994) and one German sample (Borkenau & Ostendorf, 1989), except for the facet 'actions' (Openness), which in the Spanish samples received high loadings also on Conscientiousness.

Thus, findings on the cross-cultural consistency of the five-factor model have revealed that dimensions found in the Spanish language do not entirely converge with the factor structure found in a number of Germanic languages, but that the Big Five measured through personality inventories has replicated well in the Spanish language. The last finding justifies the search for prototypes on the basis of a Big Five instrument in the two Spanish samples. We therefore expected that the three replicable prototypes from the German NEO-PI-R data can be also found in Spanish samples.

In a second analysis, replicable subtypes were analysed. In the German NEO-PI-R study, replicable subtypes have been found for the resilient type, but not for the overcontrolled or undercontrolled type (see Schnabel et al., this special issue). The two resilient subtypes have been labelled the 'well adapted' and the 'assertive', the first being characterized by a Big Five profile of low Neuroticism and medium scores on Extraversion, Openness, and Agreeableness, and the second by low Neuroticism and above-average scores on Extraversion and Openness. We therefore assumed that at least two replicable subtypes can be identified in the Spanish samples, and that these subtypes are consistent with the ones found in the German NEO-PI-R data.

In a third analysis, relations between level of education and the Spanish personality prototypes were examined. The German adult samples (Asendorpf et al., 2001; Schnabel et al., this special issue) were biased toward a higher level of education because they were dominated by students. It is therefore important to determine the relation between prototypes and level of education in a general population sample. Previous research relating education and personality prototypes have been confined to childhood. Both Robins et al. (1996) and Asendorpf and van Aken (1999) found that the IQ of resilient children was significantly higher than the IQ of the overcontrolled or undercontrolled children, and that resilient children achieved higher school grades. We therefore assumed that resilient adults should have reached, on average, a higher level of education than under- or overcontrolled adults.

METHOD

Samples

The *Spanish student sample* originally consisted of 1162 subjects, recruited from the University of Oviedo and from the University Complutense of Madrid. To ensure comparability of the sample with the German NEO-PI-R sample, we excluded all participants who were younger than 20 or older than 30 years of age, and the sample was then balanced for sex, thereby reducing it to 758 subjects. Mean age was 21.9 years (SD 1.8 years).

The *Spanish general population sample* was recruited by students of the University Complutense of Madrid, who asked family and friends to participate. The original sample consisted of 1171 subjects. After reducing the sample to include only ages from 20 to 30 years and balancing for sex, the sample's final size was 460 subjects. Mean age was 24.0 years (SD 2.9 years). Level of education was measured on a five-point scale. From the 363 subjects who had answered this item, 4% had received only basic education (4 years of school), 13% had received only basic and primary education (8 years of school), 36% had received only basic, primary, and secondary education (11–12 years of school), 25% had a bachelor's degree, and 22% a master's degree. The last two categories (totalling 47%) were substantially over-represented compared with 24% in a representative sample of the Spanish general population of the same age range (1998 data from the Spanish National Institute of Statistics).

Instrument

The two samples were examined with the NEO-PI (original version by Costa & McCrae, 1985; adaptation into Spanish by Avia et al., 1994). The NEO-PI consists of five scales that measure the Big Five factors Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness with a total of 181 items. The first three scales (Neuroticism, Extraversion, and Openness) consist of 48 items each and are subdivided into six facets. Agreeableness and Conscientiousness are not subdivided into facets and comprise 18 items each.

Internal consistencies of the scales (Cronbach's alpha) ranged between 0.71 and 0.90 for the Spanish student sample (mean reliability 0.84), and between 0.65 and 0.90 for the Spanish general population sample (mean reliability 0.82). In both samples, the internal consistency was lowest for Agreeableness. A significant difference in mean scores for the Big Five scales was only found for Openness, with the Spanish student sample scoring slightly higher, $t(1216) = 2.62$, $p < 0.01$. Significant differences in variance were found for Extraversion, $F(2, 1216) = 5.07$, $p < 0.05$, which was higher in the Spanish general population sample. Comparing mean scores of the Big Five scales with the German NEO-PI-R data (Schnabel et al., this special issue) showed that there were significant differences in Openness between the German sample and the Spanish samples ($t(1244) = 13.25$, $p < 0.001$, for the general population sample, and $t(1542) = 12.78$, $p < 0.001$, for the student sample). The positive t-values represent higher scores in the German sample.

Due to measurement inequivalence between the NEO-PI and NEO-PI-R, significant differences in mean scores on Agreeableness and Conscientiousness were not tested although the Spanish samples scored much higher on both of them. The difference in Agreeableness seems to be however due to different items rather than different language or culture. When only identical NEO-PI and NEO-PI-R items in Agreeableness were considered, differences in mean scores between the Spanish and German data disappeared, $F < 1$.

Method of deriving prototypes and subtypes

The three prototypes were derived by applying the same two-step clustering procedure that was used by Asendorpf et al. (2001) and Schnabel et al. (this special issue). In a first hierarchical analysis (Ward method), individual profiles of Big Five mean scores were clustered into groups on the basis of the squared Euclidean distances. Because the Euclidean distance is sensitive to the mean and standard deviation of the scales, which differed between the NEO-PI scales due to the smaller number of Agreeableness and Conscientiousness items, this analysis was based on Big Five scale mean scores rather than Big Five scale sum scores.

The clustering process started by each subject initially forming its own cluster. Step by step those clusters were fused that least raised their squared Euclidean distance. This process was repeated until a three-cluster solution was reached. As clusters that were once fused remain together in all further steps, it is possible that some profiles are not assigned to the most similar cluster. Therefore, in a second step a non-hierarchical cluster analysis (*k*-means) was performed, shifting the profiles into better-fitting clusters according to their Euclidean distance. Subsequently, the cluster centres were recalculated. This process was repeated up to 20 times, or until the largest change in the cluster centres was less than 2% of the smallest distance between the initial centres (the default in the SPSS procedure Quick Cluster, option NOUPDATE). Subtypes were derived by applying the same two-step clustering analysis to each cluster separately.

Replicability of the three-cluster solution was evaluated with the same procedure that was used by Asendorpf et al. (2001) and Schnabel et al. (this special issue). The samples were split into two halves that were balanced for both sex and age. A full two-step clustering analysis was performed with each half. Subsequently, the subjects' profiles were assigned to the best-fitting cluster centre of the other sample's half according to their Euclidean distance, using the SPSS procedure Quick Cluster, option CLASSIFY (this time, the cluster centres were not recalculated, of course). Convergence between the first and the second assignment was tested by Cohen's kappa, and the two resulting kappas were averaged. This process was repeated ten times, and the median kappa of the ten random splits was regarded as the final replicability coefficient. A kappa of 0.60 was considered acceptable.

Consistency across studies was assessed with the procedure used by Schnabel et al. (this special issue). The *z*-standardized individual Big Five patterns of one study were assigned to (i) the best-fitting original *z*-standardized cluster centres that were derived on the basis of raw scores from this study and (ii) the best-fitting *z*-standardized cluster centres of the other study. This resulted in two classifications per subject, which were then cross-classified. Because this procedure could be applied to both studies, the resulting κs were averaged. A mean κ-value of at least 0.60 was considered acceptable. As explained in the Editorial in a footnote, this procedure deviated slightly and for good reasons from the procedure of Asendorpf et al. (2001) of computing cross-study consistencies.

RESULTS

Table 1 reports the replicabilities for the three-cluster solution in the Spanish student and the Spanish general population samples. It also shows consistencies for the three-cluster solution between the German NEO-PI-R sample and the Spanish NEO-PI samples.

According to the cut-off criterion of $\kappa = 0.60$, the three-cluster solution in the Spanish student sample replicated well (0.69) although the replicabilities for the individual sample

Table 1. Replicability and consistency of Spanish three-cluster solutions

	Spanish NEO-PI samples	
	Students ($N = 758$)	General population ($N = 460$)
German NEO-PI-R sample ($N = 786$)	0.63	0.49
Spanish NEO-PI student sample	*0.69*	0.66
Spanish NEO-PI general population sample	—	*0.49*

Note: Reported are κ coefficients. Replicability coefficients within samples are reported in Italics.

splits varied widely (0.22–0.88). The three clusters in the Spanish student sample also showed adequate consistency with the German NEO-PI-R data (0.63). In contrast, the three-cluster solution in the Spanish general population sample was neither replicable (0.49, individual replicabilities ranging from 0.21 to 0.79), nor consistent with the German NEO-PI-R data (0.49). However, sufficient consistency with the three-cluster solution in the Spanish student sample was found (0.66).

A more detailed analysis revealed a high sensitivity of the clustering method to sample composition. Table 2 shows the consistencies between the German NEO-PI-R three-cluster solution and the Spanish general population sample and the pooled Spanish overall sample (comprising the data of both the general population sample and the student sample). In addition, consistencies are shown with two slightly modified versions of the general population sample and the pooled overall sample, when 26 subjects from another, small Spanish NEO-PI general population study were added.

Consistency for the three-cluster solution between the German NEO-PI-R and the Spanish general population sample substantially improved when the 26 subjects were included in the general population data. In contrast, this improvement could not be observed for the modified Spanish overall sample, where the consistency with the German NEO-PI-R data declined slightly. It was quite surprising for us to see that adding as few as 26 subjects to much larger samples had notable effects on cross-study consistencies.

In Figure 1, mean z-standardized NEO-PI scale scores for the three prototypes in the Spanish student sample, the general population sample, and the overall sample are

Table 2. Consistency of the three German prototypes across various Spanish samples

	Spanish NEO-PI samples			
	General population ($N = 460$)	Modified general population ($N = 486$)	Pooled overall sample ($N = 1218$)	Modified pooled overall sample ($N = 1244$)
German NEO-PI-R sample ($N = 786$)	0.49	0.64	0.59	0.54

Note: Reported are κ coefficients.

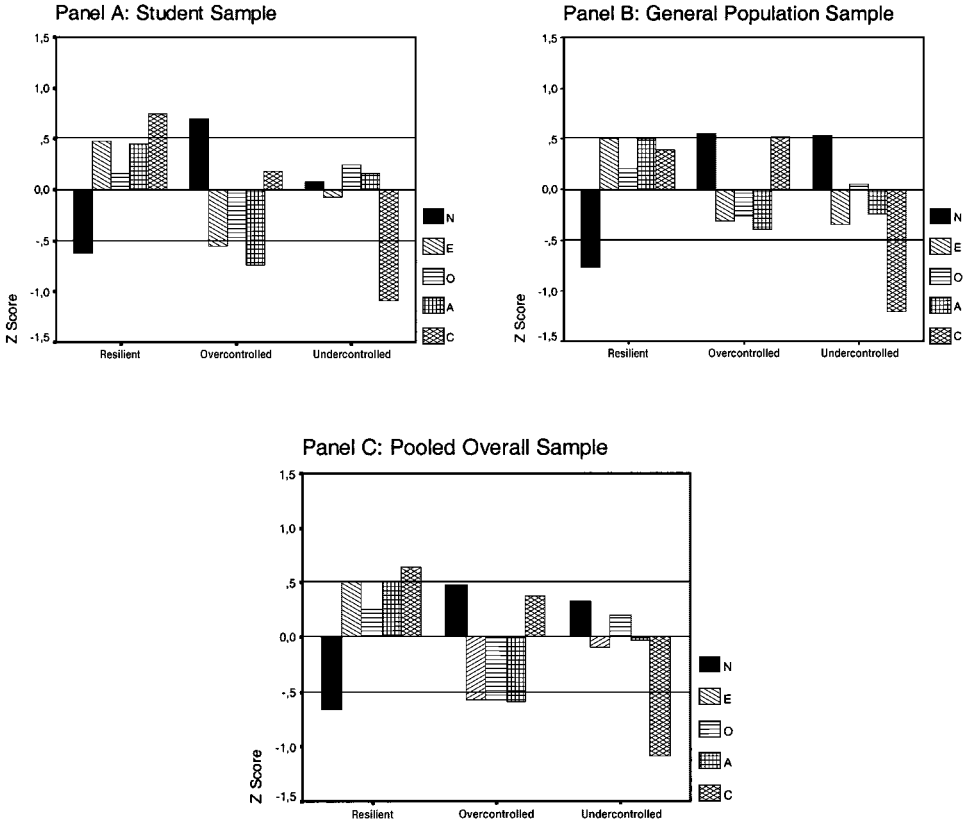

Figure 1. Three personality prototypes characterized by their Big Five patterns in two Spanish NEO-PI samples.

presented. As expected, the Spanish samples clearly replicated the characteristic profiles of the three prototypes found in the German NEO-PI-R data, with the resilient type scoring low on Neuroticism and high on most other scales, the overcontrolled type scoring high on Neuroticism and low on Extraversion, and the undercontrolled type scoring low on Conscientiousness.

Significant sample differences in standardized mean scores were tested by applying a MANOVA to the German and Spanish data for each cluster separately. Significant sample differences in z-standardized mean scores could be identified in Neuroticism, Agreeableness, and Conscientiousness for the resilient type, $F(2, 782) > 12.80$, $p < 0.001$, in all scales except for Extraversion for the overcontrolled type, $F(2, 626) > 5.70$, $p < 0.005$, and in all scales for the undercontrolled type, $F(2, 596) > 8.00$, $p < 0.001$. *Post hoc* Bonferroni t-tests were performed in order to see which samples differed significantly from each other for each scale and each cluster. In Table 3, the means and standard deviations for the z-transformed Big Five scale scores are shown for the German sample and the two Spanish samples separately for each cluster. Non-identical superscripts indicate significant sample differences.

Table 3 shows that for the resilient type, differences between the two Spanish samples and the German NEO-PI-R samples were particularly strong for Agreeableness. The

Table 3. Significant sample differences in prototypic Big Five profiles

Type	Sample	N	E	O	A	C
Resilient	German NEO-PI-R ($n = 299$)	-0.89^{a} (0.60)	0.33^{a} (0.84)	0.08^{a} (0.92)	-0.03^{a} (0.94)	0.68^{a} (0.74)
	Spanish general population ($n = 190$)	-0.77^{a} (0.66)	0.49^{a} (0.82)	0.19^{a} (0.92)	0.49^{b} (0.84)	0.39^{b} (0.69)
	Spanish students ($n = 293$)	-0.60^{b} (0.81)	0.48^{a} (0.85)	0.15^{a} (1.0)	0.44^{b} (0.87)	0.74^{a} (0.64)
Overcontrolled	German NEO-PI-R ($n = 257$)	0.71^{a} (0.71)	-0.89^{a} (0.77)	-0.62^{a} (0.92)	-0.03^{a} (1.08)	-0.03^{a} (0.81)
	Spanish general population ($n = 144$)	0.55^{a} (0.73)	-0.32^{b} (0.95)	-0.28^{b} (1.01)	-0.42^{b} (0.94)	0.53^{b} (0.65)
	Spanish students ($n = 225$)	0.70^{a} (0.87)	-0.55^{c} (0.88)	-0.47^{ab} (0.98)	-0.73^{c} (0.86)	0.18^{c} (0.72)
Undercontrolled	German NEO-PI-R ($n = 230$)	0.36^{a} (0.81)	0.56^{a} (0.72)	0.58^{a} (0.77)	0.08^{a} (0.99)	-0.84^{a} (0.82)
	Spanish general population ($n = 126$)	0.54^{a} (0.93)	-0.37^{b} (1.0)	0.03^{b} (1.04)	-0.26^{b} (0.97)	-1.20^{b} (0.68)
	Spanish students ($n = 240$)	0.08^{b} (0.88)	-0.07^{c} (1.0)	0.25^{b} (0.86)	0.15^{a} (0.89)	-1.07^{b} (0.57)

Note: Reported are mean *z*-scores (standard deviations in parentheses) within types. For the same type and scale, means with the same superscript are not significantly different across samples according to Bonferroni tests ($p < 0.05$).

Spanish resilients scored substantially higher on Agreeableness than the German ones. In comparison, significant sample differences in Neuroticism and Conscientiousness were less strong. As expected, all three samples scored below average on Neuroticism, and above average in Conscientiousness.

The most salient sample difference for the overcontrolled type was once again found in Agreeableness. The overcontrollers in both Spanish samples scored lower on Agreeableness than the overcontrollers in the German sample. The overcontrollers in the German sample scored lower on both Extraversion and Openness, though in the latter scale they did not significantly differ from the Spanish student sample. All three samples differed significantly from each other in Conscientiousness, with the general population sample scoring highest.

Given these findings, it was less surprising to see that the German NEO-PI-R scores on Extraversion and Openness in the undercontrolled type were substantially higher than in the two Spanish samples, and that scores on Openness did not differ significantly between the German undercontrollers and the Spanish student undercontrollers. These effects can be explained by the fact that scores on these two scales did not substantially differ in the resilient type and were lower for the German NEO-PI-R sample in the overcontrolled type. Thus, the effects for the undercontrollers mirrored the effects for the overcontrollers.

Another finding was that the Spanish student sample was somewhat lower on Neuroticism than both the German NEO-PI-R sample and the Spanish general population,

Table 4. Replicability of subtypes within Spanish samples

Subtypes	Spanish NEO-PI samples	
	Students	General population
Resilient		
2-subtype solution	0.64	0.60
3-subtype solution	0.56	0.35
Overcontrolled		
2-subtype solution	0.65	0.61
3-subtype solution	0.46	0.35
Undercontrolled		
2-subtype solution	0.49	0.13
3-subtype solution	0.29	0.20

Note: Reported are κ coefficients.

and that the two Spanish samples were significantly lower on Conscientiousness than the German sample, although the German undercontrollers still scored very low. Regarding Agreeableness, the Spanish general population had the lowest score, but all three scores were still close to the sample mean.

The relative frequencies of the three types in the two Spanish samples were highly similar: 40% were resilients, 30.5% were overcontrollers, and 29.5% were undercontrollers. These frequencies did not differ significantly from those found in the German NEO-PI-R sample, $\chi^2(n = 2004, df = 4) = 3.81, p > 0.40$.

Sex differences were analysed using chi-square tests that compared the relative frequency of men and women across the three types. Differences between men and women were not significant, either in the Spanish student sample, $\chi^2(n = 758, df = 2) = 1.03$, $p > 0.50$, or in the Spanish general population sample, $\chi^2(n = 460, df = 2) = 2.06$, $p > 0.35$.

Replicabilities of the subtypes were analysed for the two- and three-cluster solutions in both Spanish samples. They are reported in Table 4.

Table 4 shows that the two-cluster solutions in the resilient and overcontrolled type, but not in the undercontrolled type, showed sufficient replicability in both Spanish samples. In each of the two types, the student's kappa was slightly higher than the general population's kappa. The three-cluster solutions in the resilient, overcontrolled, and undercontrolled types were not sufficiently replicable in either of the two Spanish samples.

Although subtypes do not replicate well within a sample they may show sufficient consistency across samples. Therefore, consistencies between the German NEO-PI-R and the Spanish NEO-PI as well as the pooled Spanish overall data were analysed for the two- and the three-cluster solutions of each prototype. Consistencies are reported in Table 5.

Consistencies between the German NEO-PI-R sample and the Spanish student sample were satisfactory for the two-cluster solutions for both the resilient and overcontrolled type. The fact that the two-cluster solution for the overcontrolled type showed sufficient consistency between the two samples was quite surprising given that it had not been replicable within the German sample. The same observation could be made for the three-cluster solution in the resilient type: it showed substantial consistency between the

Table 5. Consistency of subtypes across Spanish samples and a German sample

Subtypes	German NEO-PI-R sample versus Spanish NEO-PI samples			Spanish NEO-PI students versus
	Students	General population	Overall sample	Spanish NEO-PI general population
Resilient				
2-subtype solution	0.67	0.37	0.52	0.62
3-subtype solution	0.28	0.71	0.40	0.39
Overcontrolled				
2-subtype solution	0.69	0.45	0.77	0.56
3-subtype solution	0.57	0.31	0.42	0.48
Undercontrolled				
2-subtype solution	0.42	0.47	0.49	0.22
3-subtype solution	0.23	0.36	0.39	0.14

Note: Reported are κ coefficients.

NEO-PI-R sample and the Spanish general population sample, though it did not replicate satisfactorily within either of the two samples.

Regarding the consistency of the subtypes between the Spanish NEO-PI samples, only the two-cluster solution in the resilient type was sufficiently consistent between the two of them. The consistency between the German sample and the pooled Spanish sample was quite high for the two-cluster solution in the overcontrolled type, whereas all other cluster solutions showed no sufficient consistencies.

The standardized mean scores for the two-cluster solutions in the resilient, overcontrolled, and undercontrolled types are shown in Figure 2.

As expected, the characteristic German profiles of the well adapted and assertive subtype clearly replicated in the Spanish samples. The well adapted subtype was characterized by low scores on Neuroticism and high scores on Conscientiousness. Scores on all other scales were average. The assertive subtype was characterized by low scores on Neuroticism and mostly high scores on the remaining scales. In contrast to the German sample, differences in Neuroticism between the well adapted and the assertive subtype were more pronounced in the Spanish samples, with the well adapted subtype scoring particularly low. In line with findings in the Spanish resilient prototype, the Spanish subtypes scored higher on Agreeableness than the German resilient subtypes too. It was also interesting to see that the Spanish general population, which had scored significantly lower on Conscientiousness in the resilient type as compared to the other two samples, scored lower on Conscientiousness in the assertive type, but not in the well adapted subtype.

In contrast to the German NEO-PI-R data, the two subtypes in the overcontrolled type replicated well within both Spanish samples. The first subtype was characterized by below average scores on Extraversion and Openness and average scores on the remaining scales, with the exception that the Spanish student sample also had below average scores on Agreeableness. This subtype closely resembled the subtype labelled ‘restricted’ in the German NEO-PI-R sample.

The second Spanish overcontrolled subtype was above all characterized by much higher scores on Neuroticism than the restricted subtype. The Spanish general population differed

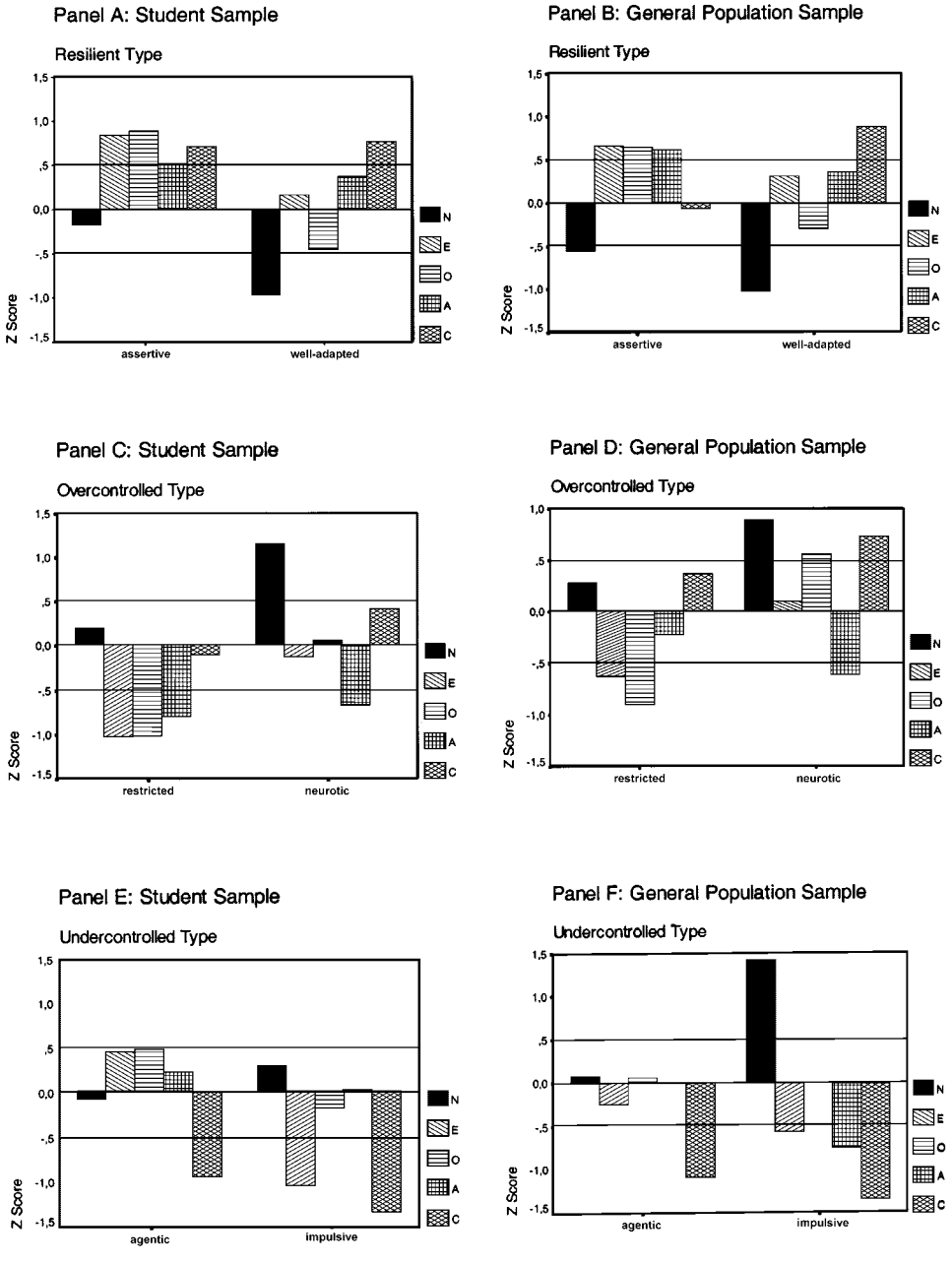

Figure 2. Two resilient, overcontrolled, and undercontrolled subtypes in two Spanish NEO-PI samples.

from the Spanish student sample insofar as scores on Openness and also Conscientiousness were above average in the general population, but were not in the student sample. The subtype in the Spanish student sample resembled very much the subtype labelled 'insecure' in the German NEO-PI-R sample, though there was one major difference between the two: the insecure subtype in the German sample was very low on Extraversion

while the Spanish student subtype had average scores. Average scores on Extraversion for this subtype were also observed in the Spanish general population. Therefore, the label 'neurotic subtype' instead of 'insecure subtype', as in the German sample, seems to be more appropriate for the second Spanish overcontrolled subtype.

The two subtypes in the undercontrolled type, which were not replicable within the samples, also differed very much in their Big Five profiles across the two Spanish samples. It was surprising to note, however, that one of the two undercontrolled subtypes in both Spanish samples resembled closely one of the two undercontrolled subtypes in the German NEO-PI-R sample. The German subtype labelled 'impulsive' came close to a subtype identified in the Spanish general population sample with high scores on Neuroticism and low scores on Conscientiousness. Differences, however, could be noted for Extraversion, Openness, and Agreeableness, where the Spanish general population scored lower. Conversely, the German subtype labelled 'agentic' had its counterpart in the Spanish student sample with its characteristic high scores on Extraversion and Openness and low scores on Conscientiousness.

Frequencies are reported for the replicable cluster solutions in the resilient and overcontrolled types. In the student sample, 65% of the resilients were assigned to the well adapted subtype and 35% to the assertive subtype. In the general population sample, frequencies were 47% for the well adapted subtype and 53% for the assertive subtype. In the overcontrolled type, frequencies for the neurotic subtype amounted to 53% in the student sample and to 44% in the general population sample, and for the restricted subtype frequencies were 47% and 56%, respectively.

Sex differences were analysed using chi-square tests. In the Spanish student sample sex differences were significant in the resilient and overcontrolled subtypes. For the resilient subtypes, the proportion of women in the assertive subtype was larger than in the well adapted subtype, where there was a higher proportion of men, $\chi^2(n = 293, df = 1) = 9.35$, $p < 0.005$. For the overcontrolled subtypes, the proportion of men was larger in the restricted subtype than in the neurotic subtype, $\chi^2(n = 225, df = 1) = 12.61$, $p < 0.001$. In the general population sample, only one significant sex difference was found for the two resilient subtypes, $\chi^2(n = 190, df = 1) = 3.79$, $p < 0.05$. Unequal proportions of men and women were identical to those identified in the Spanish student sample.

Finally, the relationship between level of education and prototype assignment in the Spanish general population sample was explored. After excluding all subjects with missing data on the level of education scale and balancing the sample for sex, the remaining 350 subjects were assigned to four groups of different educational levels (primary education, secondary education, bachelor's degree, and master's degree). Differences in level of education were analysed using a chi-square test. A significant overall difference between educational level and prototype assignment was found, $\chi^2(n = 350, df = 3) = 23.41$, $p < 0.005$. *Post hoc* comparisons were performed, testing the significance of singular cells by using the adjusted residual method (Haberman, 1974). The strongest result was that the proportion of the subjects with only primary education was significantly higher than expected in the overcontrolled type, 55% versus 32%, $z = 4.1$, $p < 0.05$. As expected, the proportion of the subjects with a master's degree was significantly higher than expected in the resilient type, 54% versus 41%, $z = 2.7$, $p < 0.05$. A somewhat weaker result was that the proportion of the subjects with a bachelor's degree was significantly higher than expected in the undercontrolled type, 35% instead of 27%, $z = 2.1$, $p < 0.05$.

DISCUSSION

Replicability and Consistency

The hypothesis that three prototypes would be replicable within the sample and consistent across samples was confirmed for the Spanish student data. Contrary to expectation, however, findings for the Spanish general sample were mixed. On the one hand, the three prototypes did not show sufficient replicability within the sample, and the consistency with the German NEO-PI-R data was unsatisfactory. On the other hand, the three prototypes in the general population sample were consistent with those found in the Spanish student sample, and when as few as 26 subjects were added to the general population sample consistency with the NEO-PI-R data became acceptable, too. These findings suggest that the clustering procedure is highly sensitive to sample composition, and cluster solutions may be consistent across samples although they are not replicable within a sample with the random-split procedure suggested by Asendorpf et al. (2001).

Sensitivity to sample composition could also be observed in the Spanish overall sample (note that the sample size exceeded 1000 subjects), where consistency with the German NEO-PI-R data decreased when 26 subjects were added. One should keep in mind, though, that findings regarding sensitivity to sample composition included the Spanish general population sample, where the prototypes were not replicable within the sample.

One case where clustering is highly sensitive to sample composition is when there are different but similarly optimal cluster solutions. When we analysed the replicability of the three-cluster solution in the Spanish general population sample from this perspective, two alternative three-cluster solutions turned out to be optimal. When we split the sample into halves and inspected the distribution of the kappas for more than the standard ten random splits, there was a notable bimodal distribution of the kappas over 50 random splits. One peak was found in the replicable area with kappas above 0.60 that referred to similar three-cluster solutions. Another peak concerned the non-replicable area with kappas clearly below 0.60 that referred to two different, rival three-cluster solutions. In contrast, only one optimal solution was found for the Spanish student sample, and the kappas had a unimodal distribution in this case. The replicability did not decrease even when the sample size was reduced to 460 subjects.

The finding of two different but similarly optimal solutions in the Spanish general population sample was supported by the fact that a bimodal distribution of the kappas could also be observed in the German NEO-PI-R data, though it was somewhat less marked. Interestingly, cluster solutions that were not replicable within the sample and yet consistent across samples were only found for the Spanish general population sample and the German NEO-PI-R sample, but not for the Spanish student sample. In other words, consistency evaluation that considered the full sample obtained more robust results than replicability evaluation that within a sample. This finding suggests that sample size can be particularly important when a sample comprises two optimal cluster solutions.

That the within-study replicability was often lower than the cross-study consistency may appear contraintuitive at a first glance because two random halves of one sample are, on average, more homogeneous than samples that vary in educational level or culture. It should be noted, however, that the cross-study consistencies were based on the full samples whereas the within-study replicabilities were based only on halves of the sample, which makes comparably high replicabilities less likely. Alternative methods for evaluating within-study replicability such as the bootstrapping procedure proposed

by Barbaranelli (this special issue) are not affected by this problem of reduced sample size.

Prototypes

The resilient, overcontrolled, and undercontrolled types with their characteristic patterns in the Big Five were by and large replicated in the Spanish samples. Noteworthy differences in cluster mean scores between the German NEO-PI-R data and the Spanish samples were found in Agreeableness for the resilient and the overcontrolled types. The Spanish resilients described themselves as more agreeable than the German resilients whereas the Spanish overcontrollers showed unexpectedly low scores on Agreeableness in comparison with the German overcontrollers. As Agreeableness is not equivalently measured in the NEO-PI-R and NEO-PI, cluster differences could be due to the unequal number of items. A new cluster analysis using the same number of items in the NEO-PI-R and NEO-PI partly supported the results obtained under the condition of unequal measurement. Differences in mean *z*-scores between the Spanish overcontrollers and the German overcontrollers further increased (from an average of 0.55 to 0.96), whereas differences in the resilient type, though still significant, decreased (from an average of 0.50 to 0.22).

A possible explanation for low scores on Agreeableness in the Spanish overcontrolled type comes from a study examining the cultural script of 'simpatía' found in Hispanic populations (Triandis, Lisansky, Marín, & Betancourt, 1984). This concept refers to a personal quality where an individual is perceived as 'likeable, attractive, fun to be with, and easy going' (Triandis et al., 1984, p. 1363). Accordingly, the profile of the overcontrolled type with its high scores on Neuroticism and low scores on Extraversion contradicts Spanish expectations of positive interpersonal behaviour, and overcontrollers may therefore perceive themselves as less agreeable.

Another difference was that Spanish undercontrollers described themselves as less extraverted than the German undercontrollers, and that scores on Openness were also somewhat lower, particularly in the general population sample. A similar pattern of medium to low Extraversion and Openness scores in the undercontrolled type was found in child samples (Robins et al., 1996; Asendorpf et al., 2001).

Subtypes

The hypothesis that the two-cluster solution would be replicable within the Spanish samples was confirmed. Consistency of the two-cluster solution was demonstrated between the German NEO-PI-R and the Spanish student sample and across the two Spanish samples. The typical Big Five profiles of the well adapted and assertive subtype were similar to the German profiles. Contrary to our expectation, the two-cluster solution in the overcontrolled type was also replicable within the Spanish samples and consistent between the German NEO-PI-R and the Spanish student sample. The Spanish restricted subtype had a very similar Big Five profile to that of the NEO-PI-R sample. The neurotic subtype in the Spanish samples, however, differed from the German insecure subtype above all in Extraversion, i.e. the Spanish subtype was mainly neurotic, but not introverted.

The two overcontrolled subtypes identified in the Spanish samples and in the German NEO-PI-R data resemble two corresponding types in the classic study by Block (1971). He described two different overcontrolled types for the male and female sample. The overcontrolled type identified in the female sample—the hyper-feminine

repressives—was characterized by a highly neurotic and low-agreeable character and resembles the Big Five neurotic subtype. Conversely, the overcontrolled type identified in the male sample was characterized by high social insecurity, conformity, and low intellectual Openness and resembles closely the Big Five restricted subtype. In addition, sex differences found in the Big Five overcontrolled subtypes were consistent with the distinction between Block's male and female overcontrollers, as in both the Spanish student and the German NEO-PI-R sample more women than men were assigned to the neurotic subtype and more men than women to the restricted subtype.

Education

Differences in level of education in prototype assignment were found, confirming the hypothesis that subjects with a higher level of education would be over-represented in the resilient type. Comparing this finding to the results by Robins et al. (1996) and Asendorpf and van Aken (1999), there seems to be a continuity from IQ in childhood to educational level in adulthood, as resilients on average scored higher in IQ tests as children and had a higher level of education as adults.

Another finding was that the proportion of subjects with lower attained level of education was larger than expected in the overcontrolled type and not, as it had alternatively been assumed, in the undercontrolled type. One possible explanation could be that Openness differentiates over- from undercontrollers. It might be that the undercontrollers' higher scores on Openness ($z = 0.05$) in comparison to the overcontrollers' scores ($z = -0.26$) create more favourable conditions for reaching a higher level of education, independent of the IQ.

CONCLUSION

In the present study, similarities and differences in Spanish and German personality prototypes and subtypes were analysed. Regarding methodology, a strong dependency of the results on sample composition, and a better consistency across samples than replicability within samples were found. The resilient, overcontrolled, and undercontrolled prototypes, and to some degree their subtypes, above all the assertive and well adapted subtypes of the resilient type, showed modest to good cross-cultural consistency in their characteristic Big Five patterns. Cross-cultural comparisons on the basis of the Big Five are limited, however, to the extent that the Big Five represents a concept of basic dimensions of personality that can be found in a number of Germanic languages, but not to the same extent in other languages such as Spanish.

If one is interested in testing the consistency of personality prototypes between Germany and Spain, it is of course possible to compare German and Spanish types on the basis of the Big Five as people in both Germany and Spain reliably differ on these dimensions, but one cannot claim to have identified generalizable cross-cultural similarities or differences in personality types because the types are based on dimensions that vary in their centrality for Germans and Spanish people.

In order to achieve a truly generalizable typology, it is necessary to try to find consistency both at the level of personality dimensions (variable-centred perspective) and at the level of personality types (person-centred perspective). It was said at the beginning that the Big Three may be factors that can be found in most cultures that have been studied

today. Although prototypes based on the Big Three would represent very rough distinctions of personality profiles on a very high level of abstraction, most cultures could perhaps reliably be compared according to these prototypes. Future studies may investigate this hypothesis.

REFERENCES

Asendorpf, J. B., Borkenau, P., Ostendorf, F., & van Aken, M. A. G. (2001). Carving personality description at its joints: Confirmation of three replicable personality prototypes for both children and adults. *European Journal of Personality, 15,* 169–198.

Asendorpf, J. B., & van Aken, M. A. G. (1999). Resilient, overcontrolled and undercontrolled personality prototypes in childhood: Replicability, predictive power, and the trait-type issue. *Journal of Personality and Social Psychology, 77,* 815–832.

Avia, M. D., Sanz, J., Sánchez-Bernardos, M. L., Martínez-Arias, R., Silva, F., & Graña, J. L. (1995). The five-factor model: II. Relations of the NEO-PI with other personality variables. *Personality and Individual Differences, 19,* 81–97.

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

Benet-Martínez, V., & Waller, N. G. (1997). Further evidence for the cross-cultural generality of the 'Big Seven' model: Imported and indigenous Spanish personality constructs. *Journal of Personality, 65,* 567–598.

Block, J. (1971). *Lives through time.* Berkeley, CA: Bancroft.

Borkenau, P., & Ostendorf, F. (1989). Untersuchungen zum Fünf-Faktoren-Modell der Persönlichkeit und seine diagnostische Erfassung [Studies on the Five-Factor Model of personality and its assessment]. *Zeitschrift für differentielle und diagnostische Psychologie, 10,* 239–251.

Church, A. T., & Burke, P. J. (1994). Exploratory and confirmatory tests of the Big Five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology, 66,* 93–114.

Costa, P. T., & Mc Crae, R. R. (1985). *The NEO-PI/NEO-FFI manual supplement.* Odessa, FL: Psychological Assessment Resources.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141–165). Beverly Hills, CA: Sage.

Haberman, S. J. (1974). *The analysis of frequency data.* Chicago, IL: University of Chicago Press.

Hart, D., Hofmann, V., Edelstein, W., & Keller, M. (1997). The relation of childhood personality types to adolescent behavior and development: A longitudinal study of Icelandic children. *Developmental Psychology, 33,* 195–205.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin, & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford.

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52,* 81–90.

McCrae, R. R., & Costa, P. T. (1989). Rotation to maximize the construct validity of factors in the NEO Personality Inventory. *Multivariate Behavior Research, 24,* 107–124.

Ostendorf, F. (1990). *Sprache und Persönlichkeitsstruktur: Zur Gültigkeit des Fünf-Faktor-Modells der Persönlichkeit* [*Language and personality structure: On the validity of the Five-Factor Model of personality*]. Regensburg: Roderer.

Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology, 57,* 552–567.

Robins, R. W., John, O. P., Caspi, A., Moffitt, T. E., & Stouthamer-Loeber, M. (1996). Resilient, overcontrolled, and undercontrolled boys: Three personality types in early adolescence. *Journal of Personality and Social Psychology, 70,* 157–171.

Saucier, G., Hampson, S. E., & Goldberg, L. R. (2000). Cross-language studies of lexical personality factors. In S. E. Hampson (Ed.), *Advances in Personality Psychology* (Vol. 1, pp. 1–37). Hove: Psychology.

Silva, F., Avia, M. D., Sanz, J., Martínez-Arias, R., Graña, M. L., & Sánchez-Bernardos, M. L. (1994). The Five Factor Model. I: Contributions to the structure of the NEO-PI. *Personality and Individual Differences, 17,* 741–753.

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

