Journal of Research in Personality 46 (2012) 442–449

Contents lists available at SciVerse ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

A multigroup multitrait-multimethod study in two countries supports the validity of a two-factor higher order model of personality

Ömer Faruk Şimşek [a,*], Selda Koydemir [b], Astrid Schütz [c]

[a] Izmir University of Economics, Department of Psychology, Izmir, Turkey
[b] Department of Psychology, Chemnitz University of Technology, Chemnitz, Germany
[c] Department of Psychology, University of Bamberg, Germany

ARTICLE INFO

Article history:
Available online 8 May 2012

Keywords:
Big Five
Stability
Plasticity
Personality
Informant reports
Multitrait-multimethod analysis

ABSTRACT

This study examined the factor structure of the Big Five Inventory and tested the hypothesis that the five personality dimensions could be summarized by two higher order factors, namely, plasticity and stability, using multigroup multitrait-multimethod confirmatory factor analyses. We tested the higher order model in two young adult samples drawn from Germany and Turkey. Adequate inter-rater agreement between self- and informant reports was obtained. Among the models tested, a two-factor model was the most parsimonious model in which the first factor included Agreeableness, Conscientiousness, and low Neuroticism; and the second factor included Extraversion and Openness to Experience. Invariance of this model was supported by multiple-group analyses, suggesting a lack of variability across samples.

1. Introduction

The Five Factor Model, also referred to as the Big Five, has been the most widely accepted and empirically validated classification and conceptualization of personality (Costa & McCrae, 1985; Goldberg, 1990; John, Angleitner, & Ostendorf, 1988). However, since the late 1990s, a variety of measurement issues have been discussed as caveats regarding this model, one of which concerns the orthogonality of the personality dimensions. For instance, although the Big Five dimensions were originally conceived as relatively independent and as representing the highest level of personality (Costa & McCrae, 1992; Goldberg, 1993), consistent relations among them have been found (Benet-Martinez, & John, 1998; Mount, Barrick, Scullen, & Rounds, 2005), which led researchers to question the extent to which the model indeed reflects the highest order constructs underlying individual differences. Consequently, several attempts have been made to examine whether broader constructs of personality exist such as two higher order factors (e.g., Ashton, Lee, Goldberg, & de Vries, 2009; DeYoung, Peterson, & Higgins, 2002; Digman, 1997) or a general factor of personality (e.g., Musek, 2007; Rushton et al., 2009).

A leading proposition regarding the existence of metatraits as an explanation for the nonorthogonality of the Big Five was made by Digman (1997), who, after studying the interrelations among the five factors of personality in large data sets, found that there were two orthogonal higher order traits, or the Big Two, which he named *alpha* and *beta*. Whereas the first factor merges Emotional Stability, Conscientiousness, and Agreeableness, the second factor is comprised of Openness and Extraversion. Digman interpreted *alpha* as reflecting the effects of socialization, which include qualities related to being responsible, productive, and a good person, and *beta* as related to personal growth and self-actualization. Becker (1999), similarly, found two higher order factors, namely, mental health and behavior control. Later, studies by DeYoung and colleagues (e.g., DeYoung, Hasher, Dijkic, Criger, & Peterson, 2007; DeYoung et al., 2002) revealed two similar factors, which were named *stability* and *plasticity* and represented *alpha* and *beta*, respectively. Using twin pairs, Jang et al. (2006) likewise demonstrated that the Big Two exist and are stable across three culturally diverse samples. Other replication studies (e.g., Anusic, Schimmack, Pinkus, & Lockwood, 2009) also found evidence for the Big Two.

As an attempt to provide an explanation for the correlations between the Big Five, another line of research examined the possibility of a single general factor of personality (GFP). For instance, Musek (2007) used different personality measures of the Big Five in three samples, documented a GFP named the Big One, and interpreted that factor as a combination of positive personality states (consisting of high Agreeableness, high Extraversion, low Neuroticism, high Conscientiousness, and high Openness). Basing on an evolutionary theory, Rushton and colleagues (Rushton & Irwing, 2008; Rushton et al., 2009) similarly showed evidence of a GFP. Using behavioral-genetic analyses, Vaselka, Schermer, Petrides,

* Corresponding author. Address: Izmir University of Economics, Department of Psychology, Sakarya Cad. No.: 156, Balcova, Izmir 35330, Turkey.
E-mail address: simsekof@gmail.com (Ö. Faruk Şimşek).

0092-6566/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jrp.2012.04.005

and Vernon (2009) quantified the genetic basis of the GFP. However, some studies have failed to show the superiority of that model (e.g., Ashton et al., 2009; de Vries, 2011).

Although the existence of superordinate traits—represented by either a two-factor or a general factor model—have inspired researchers to study personality in a different way, a number of issues remain unresolved. One of the most important criticisms has been the argument that a higher order factor may be a method artifact, indicating that it is a result of a common method bias factor (e.g., Biesanz & West, 2004; McCrae & Costa, 1999). Thus, attempts have been made to replicate the previous studies and to determine whether the relations among dimensions and factors that are extracted are the result of a methodological artifact (see Chang, Connelly, & Geeza, 2012, for a recent review). One option for approaching this question is to use a multimethod approach and to measure multiple traits—each trait on the basis of multiple methods—or as in this study, multiple informants (Campbell & Fiske, 1959). The multimethod approach is advantageous because it allows for partialling out common method variance from estimates of the relation between traits. Using a multimethod strategy, Biesanz and West (2004) found that the Big Five were orthogonal. Anusic at al. (2009) examined artifact interpretations and found that the correlations among the Big Five, which were assessed by self-reports only, were at least partly due to rating biases, but that two higher order factors emerged even after method effects had been controlled. DeYoung (2006) similarly showed some evidence that the Big Five were correlated even when the biases were accounted for and concluded that two orthogonal higher order factors existed. Chang et al. (2012) documented that method variance may have an influence on the interrelations among the Big Five; however, they also concluded that these traits were not orthogonal.

Besides the artifact interpretations of the hierarchical structure of the Big Five, another issue emerged after replication studies were conducted. In different studies, the composition of the two higher order factors was defined by different personality dimensions. For example, De Raad and Peabody (2005) found that Neuroticism was absent from the factor structure, whereas in Jang et al.'s (2006) study, *alpha* was defined primarily by Neuroticism and Conscientiousness and by the absence of Agreeableness. Another deviating example was observed in a study by Blackburn, Renwick, Donnelly, and Logan (2004), who showed that Openness added little to the *beta* factor.

Finally, previous studies testing higher order structures of the Big Five used mainly samples from North American and West European populations. We believe that it is essential to conduct cross-cultural studies and to test whether the findings hold across samples. In this way, previous research findings can be extended, and additional validity evidence can be provided. Moreover, none of the existing studies examined the viability of the higher order models across cultures by incorporating the issue of equality of factor structures within a multitrait-multimethod strategy. It is indisputable that a true comparison of different models across groups necessitates examining the equality of factor structures for the groups under investigation.

Therefore, in this study, we tested the validity of the Big Two across cultures. Given that earlier findings showed inconsistencies regarding the orthogonality of the personality dimensions and the nature of the hierarchical structure of personality, we made sure to use a sound methodology. We employed a multigroup multitrait-multimethod (MTMM) model using a confirmatory factor analysis (CFA) approach with multiple informant reports. MTMM assumes that although correlations based on the same method may be biased by shared method variance, correlations across independent methods (i.e., cross-method) are less biased (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003). Thus, when the relations among different traits are examined across different methods, trait effects can be separated from potential method effects. We tested higher order models of personality in German and Turkish samples, which are considered to be culturally diverse (Kagitcibasi, Ataca, & Diri, 2010; Schwartz, 2004). To our knowledge, the MTMM approach has not been employed in a cross-cultural study that examined the higher order structure of the Big Five. Thus, the present study is the first that used a cross-cultural approach in a multigroup methodology and an MTMM approach to tackle the issue of higher order factors.

In the analyses, we treated each informant (i.e., Informant 1 and Informant 2) as a different method and compared the results of the analyses obtained in each country. We expected that there would be high inter-rater agreement because the BFI usually produces consistent personality ratings from multiple sources (e.g., DeYoung, 2006) and because a similar cross-cultural study obtained good agreement between self- and informant reports (Koydemir & Schütz, 2012). High inter-rater agreement generally allows for higher correlations between the personality dimensions (Biesanz & West, 2004; DeYoung, 2006), and considerable evidence has been presented regarding the interrelations between the Big Five (e.g., Chang et al., 2012; DeYoung et al., 2002; Digman, 1997; Olson, 2005). Therefore, we expected similar findings with small to moderate correlations between the Big Five dimensions. Accordingly, we proposed that a Big Two model with Agreeableness, Conscientiousness, and low Neuroticism merging in the first factor and Extraversion and Openness to Experience included in the second factor would fit the data better than a first-order model with five factors. As in Chang, Connelly, and Geeza's (2012) study, we also expected that a Big Two model would fit the data better than a GFP.

2. Method

2.1. Participants

Two young adult samples were used in the present study. The first one was selected from a large university in Eastern Germany, and the second sample was drawn from a university located in the capital city of Turkey. Students' ages ranged between 20 and 31 ($M = 23.50$, $SD = 2.21$) in Germany and 18 and 29 ($M = 22.12$, $SD = 2.81$) in Turkey. In the present study, we used only data sets that included self-reports and data from two informants who knew the participants well. We ensured that students in the German sample were originally from Germany and in the Turkish sample, were originally from Turkey. The final sample size was 223 (151 participants from Germany and 72 participants from Turkey). Reports from a total of 446 informants—two informants for each participant—were collected.

2.2. Instruments and procedures

As part of a larger battery of personality and well-being measures, both samples completed *The Big Five Inventory* (BFI-44; John & Srivastava, 1999), which is a 44-item measure of the Big Five personality traits assessing Extraversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N), and Openness to Experience (O). The BFI is a widely used and well-validated instrument having scales with short statements rated on a 5-point scale ranging from 1 (*disagree strongly*) to 5 (*agree strongly*). Sample items are: "Outgoing, sociable" (E); "Likes to cooperate with others" (A); "Does a thorough job" (C); "Worries a lot" (N); "Is curious about a number of things" (O). The instrument was developed to measure the five dimensions using as few items as possible and simultaneously providing adequate reliability. John and Srivastava (1999) reported

that the scales showed good convergent validity with Trait Descriptive Adjectives and the NEO Five-Factor Inventory. Others (e.g., Benet-Martinez & John, 1998) have also demonstrated good reliability and validity evidence. Past research has also demonstrated that the BFI is a cross-culturally valid and reliable tool for measuring five personality traits (Schmitt, Allik, McCrea, & Benet-Martinez, 2007). To ensure validity, we used previously translated and validated German (Lang, Lüdtke, & Asendorpf, 2001) and Turkish (Alkan, 2006) versions of the inventory. In the current study, the alpha reliabilities ranged between .79 and .84 in Germany and between .74 and .85 in Turkey.

All students responded to an online survey including a demographic information form and the BFI. After completing the survey, they were requested to nominate two close acquaintances as informants whom they thought knew them well and who would be willing to respond to a web-based survey. We asked participants to provide the email address of each informant, and we ensured the confidentiality of the informant responses. Then the informants were contacted by email and were requested to fill out the same questionnaires online, which were modified to allow for the informant to rate the student on each scale. Follow-up emails were sent if informants did not respond. The informants were close acquaintances—either very close friends (66%) or family members (34%). We did not offer incentives to informants as compensation for participation; however, students in Germany received course credits, whereas students in Turkey received either money or feedback on their scores.

2.3. Strategy of analyses

First, an MTMM correlation matrix was examined to determine the level of inter-rater agreement (correlations for the same trait, different informant) for the Big Five factors. In the traditional MTMM analysis strategy provided by Campbell and Fiske (1959), a latent factor is defined for each trait assessed as well as for each method employed. However, this method has been known to result in many identification problems and convergence failures, which can cause unstable estimates (Eid et al., 2003; Marsh & Bailey, 1991). Thus, we used uniqueness models in which multiple methods are still used, but latent method factors are not defined. Rather, disturbance variables are allowed to intercorrelate, and this is considered to represent method effects (Lance, Noble, & Scullen, 2002). Moreover, uniqueness models allow for multidimensional method effects in that the model does not force each method to load on one factor (Kenny & Kashy, 1992). Thus, uniqueness models, which do not assume unidimentionality of method effects, seem to better account for common method bias and are in line with our theoretical reasoning.

Because we aimed to provide support for the validity of the higher order structure of personality for both groups, we conducted a series of multigroup uniqueness model analyses using LISREL 8.8 (Jöreskog & Sörbom, 2006) to determine whether the first-order and the second-order models of personality fit both the Turkish and German data. The models tested in this study are outlined below. In all of these tests, we assumed that the factor loadings of the informants would be equal on all personality dimensions because all of the informants were close acquaintances. Moreover, given that a latent variable with two indicators would result in identification problems (Kline, 2005), we also assumed that the paths from *plasticity* to E and O were equal.

M1. First-order Correlated Trait Correlated Uniqueness (CTCU) model for both Turkish and German samples.
M2. First-order Orthogonal Traits Correlated Uniqueness model for both Turkish and German samples.
M3. CTCU model of the Big Two for both Turkish and German samples, in which E and O create *plasticity*, and N, A, and C create *stability*, as identified by DeYoung et al. (2002).
M3A. CTCU model of a GFP for both Turkish and German samples, in which all personality factors create a general factor of personality.

The comparison of Models 1 and 2 thus provides a statistical test of the orthogonality of the Big Five. We expected that the first model would fit the data better than the second. In Model 3, we tested the measurement model in which *stability* and *plasticity* were defined as higher order factors of the Big Two. Model 3A is an alternative model of the Big Two, which defines a GFP as the higher order construct behind the Big Five factors. We expected that Model 3 would fit the data better than both the first-order model and the alternative second-order factor (Model 3A) in both countries. After assuring the validity of the Big Two model in both groups, we continued providing support for the equivalence of this model across groups.

M4. First-order multigroup CTCU model in which all parameters are assumed to be equal across groups.
M5. First-order multigroup CTCU model in which factor loadings are assumed to be different across groups.
M6. First-order multigroup CTCU model in which factor correlations are assumed to be different across groups.

Tests of Models 4 through 6 test the assumption that personality factor structures are equivalent across groups, which then allows for tests of the second-order CTCU model for both groups.

After examining the above-mentioned models, finally, we compared the higher order multigroup model in which no between-group constraints on any of the Big Two factor loadings (M7) were allowed to a model in which these factor loadings were considered to be different in both groups (M8).

3. Results

3.1. Preliminary analyses

MTMM correlation matrices, means, and standard deviations for the Turkish and German samples are shown in Table 1 and Table 2, respectively. Heteromethod correlations of the factors ranged in magnitude from .41 to .83 for the Turkish sample and from .40 to .60 for the German sample, indicating a high level of inter-rater agreement between self- and informant ratings.

3.2. The correlated uniqueness models analyses

As can be seen in Table 3, the test of M1,first-order CTCU, resulted in acceptable goodness of fit statistics in both the Turkish sample, $\chi^2(55, N = 72) = 57.77$, $p > .05$; CFI = .98; RMSEA = .027 (90% confidence interval for the RMSEA = .000–.080), and the German sample, $\chi^2(55, N = 151) = 74.06$, $p < .05$; CFI = .98; RMSEA = .048 (90% confidence interval for the RMSEA = .008–.077). The test of M2,first-order orthogonal traits correlated uniqueness model, resulted in a clear deterioration of the model fit in both the Turkish sample, $\chi^2(65, N = 72) = 119.68$, $p < .05$; CFI = .90; RMSEA = .109 (90% confidence interval for the RMSEA = .078–.140)], and the German sample, $\chi^2(65, N = 151) = 117.38$, $p < .05$; CFI = .95; RMSEA = .073 (90% confidence interval for the RMSEA = .052–.094). Chi-square difference tests showed that the first model was better than the second for both the Turkish (61.91,10: $p < .01$) and German (43.32,10: $p < .01$) samples, which

Table 1
Multitrait multi-informant correlation matrix for the Turkish sample.

	Self					Informant 1					Informant 2				
	E	O	A	N	C	E	O	A	N	C	E	O	A	N	C
Self															
E	–														
O	.46**	–													
A	.15	.00	–												
N	−.15	.12	−.59**	–											
C	.14	.03	.56**	−.47**	–										
Inf-1															
E	**.54****	.27*	.17	−.12	.06	–									
O	.29*	**.45****	.12	.04	−.03	.32**	–								
A	.06	−.04	**.56****	−.34**	.15	.25*	.28*	–							
N	−.11	.10	−.53**	**.65****	−.21	−.09	−.20	−.43**	–						
C	.01	−.10	.43**	−.15	**.41****	.23	.24	.54**	−.36**	–					
Inf-2															
E	**.74****	.35**	.16	−.15	.02	**.52****	.35**	.18	−.03	.00	–				
O	.15	**.56****	.01	.06	.11	.15	**.43****	.00	.01	.19	.18	–			
A	.09	−.02	**.83****	−.48**	.48**	.08	.09	**.55**	−.44**	.39**	.24*	.10	–		
N	.05	.27*	.37**	**.71****	−.25*	.12	.17	−.21	**.47****	−.05	.00	.05	−.35**	–	
C	.01	−.08	.45**	−.46**	**.69****	.00	.02	.24*	−.24*	**.43****	.08	.23	.54**	−.42**	–
M	29.27	34.37	33.95	22.11	32.13	29.66	33.93	34.51	21.87	34.33	29.55	33.54	34.11	21.90	32.58
SD	5.36	4.70	5.93	5.99	5.46	5.06	5.79	5.68	5.11	6.00	4.74	5.38	5.12	5.09	5.60

Note. $N = 72$; E = Extraversion, O = Openness, A = Agreeableness, N = Neuroticism, C = Conscientiousness; Heteromethod correlations are indicated bold.
* $p < .05$.
** $p < .01$.

Table 2
Multitrait multi-informant correlation matrix for the German sample.

	Self					Informant 1					Informant 2				
	E	O	A	N	C	E	O	A	N	C	E	O	A	N	C
Self															
E	–														
O	.26**	–													
A	.15	.05	–												
N	−.23**	−.01	−.34**	–											
C	.07	−.04	.16*	−.12	–										
Inf-1															
E	**.48****	.21**	.10	−.20*	.06	–									
O	.10	**.44****	−.05	.00	.07	.47**	–								
A	.09	.00	**.53****	−.22**	.20*	.23**	.18*	–							
N	−.10	−.08	−.31**	**.59****	−.19*	−.27**	−.09	−.48**	–						
C	.09	−.10	.14	−.16	**.59****	.16	.25**	.46**	−.34**	–					
Inf-2															
E	**.47****	.22**	−.02	−.01	.03	**.40****	.22**	−.04	−.08	−.06	–				
O	.12	**.56****	−.04	.12	.04	.23**	**.56****	.08	−.11	.08	.45**	–			
A	.05	.16	**.47****	−.17*	.17*	.25**	.17*	**.52****	−.43**	.22**	.22**	.24**	–		
N	−.03	−.07	−.21**	**.47****	−.10	−.16*	−.12	−.28**	**.60****	−.15	−.15	−.15	−.51**	–	
C	−.09	−.10	.01	.07	**.58****	.11	.14	.24**	−.12	**.57****	.05	.22**	.22**	.20*	–
M	28.09	37.29	33.88	24.31	32.09	30.25	38.09	34.92	22.47	34.23	29.98	37.95	34.13	22.30	33.54
SD	4.65	5.81	5.07	4.32	5.52	5.02	5.59	5.12	5.43	6.06	5.07	6.27	5.12	5.43	6.24

Note. $N = 151$; E = Extraversion, O = Openness, A = Agreeableness, N = Neuroticism, C = Conscientiousness; Heteromethod correlations are indicated bold.
* $p < .05$.
** $p < .01$.

provided support for rejecting the orthogonality of personality factors.

The test of M3, CTCU model of the Big Two, resulted in acceptable goodness of fit statistics in both the Turkish sample, $\chi^2(60, N = 72) = 63.84$, $p > .05$; CFI = .98; RMSEA = .030 (90% confidence interval for the RMSEA = .000–.080), and the German sample, $\chi^2(60, N = 151) = 75.18$, $p < .05$; CFI = .98; RMSEA = .041 (90% confidence interval for the RMSEA = .000–.068). In order to test whether there was a statistically significant difference between this higher order model and first-order CTCU model (M1), a chi-square difference test was performed in both groups. The results showed that there was no statistically significant difference between the models in the Turkish (6.07,4: $p > .05$) and German (1.12,4: $p < .05$) groups. Because the Big Two model was more parsimonious than the first-order model, we concluded that the higher order factors accurately describe the pattern of covariance among the first-order factors. Moreover, a test of the GFP model, M3A, resulted in reduced model fit for both the Turkish, $\chi^2(63, N = 72) = 116{,}26$, $p > .05$; CFI = .92; RMSEA = .109 (90% confidence interval for the RMSEA = .077–.140), and German samples, $\chi^2(63, N = 151) = 108.90$, $p > .05$; CFI = .96; RMSEA = .070 (90% confidence interval for the RMSEA = .040–.091).

Table 3
The results of the correlated uniqueness models tested.

Models	χ^2	*df*	CFI	RMSEA	90% CI for RMSEA
Within-culture analyses					
M1 (F-CTCU)					
Turkish	57.77	55	.98	.027	000–080
German	74.06	55	.98	.048	008–074
M2 (F-OTCU)					
Turkish	119.68	65	.90	.109	078–140
German	117.38	65	.95	.073	052–094
M3 (BT-CTCU)					
Turkish	63.84	60	.98	.030	000–080
German	75.18	60	.98	.041	000–068
M3A (GFP-CTCU)					
Turkish	116.26	63	.92	.109	077–140
German	108.90	63	.96	.070	047–091
Across-culture analyses					
M4 (F-MG-CTCU/APE)	263.30	175	.93	.068	050–084
M5 (F-MG-CTCU/FLD)	229.58	160	.94	.063	043–080
M4R (F-MG-CTCU/PI)	248.06	173	.94	.063	044–080
M6 (F-MG-CTCU/FCD)	254.78	165	.93	.070	053–087
M7 (BT-MG-CTCU/APE)	263.69	179	.93	.065	048–082
M8 (BT-MG-CTCU/FLD)	260.72	176	.93	.066	048–082

Note. $N = 72$ (Turkish); 151 (German); F-CTCU = First-order correlated traits correlated uniqueness model, F-OTCU = First-order orthogonal traits correlated uniqueness model, BT-CTCU = Correlated traits correlated uniqueness model of the Big Two, GFP-CTCU = Correlated traits correlated uniqueness model of General Factor of Personality, F-MG-CTCU/APE = First-order multigroup correlated traits correlated uniqueness model with all parameters are assumed to be equal across groups, F-MG-CTCU/FLD = First-order multigroup correlated traits correlated uniqueness model in which factor loadings are assumed to be different across groups, F-MG-CTCU/PI = Partial invariant version of M4, F-MG-CTCU/FCD = First-order multigroup correlated traits correlated uniqueness model in which factor correlations are assumed to be different across groups, BT-MG-CTCU/APE = Multigroup correlated traits correlated uniqueness model of the Big Two in which all parameters are assumed to be equal, BT-MG-CTCU/FLD = Multigroup correlated traits correlated uniqueness model of the Big Two in which second-order factor loadings are assumed to be different across groups.

After providing preliminary evidence that the Big Two model was reasonable in both groups, we tested the equality of covariance structures across groups in a multiple-group confirmatory factor analysis. Before testing the higher order model, we tested a series of first-order multiple-group analyses to provide thorough evidence for the equality of the personality factor structure across groups.

The test of M4, first-order multigroup CTCU which assumed invariance across groups, resulted in acceptable goodness of fit statistics, $\chi^2(175, N = 223) = 263.30$, $p < .05$; CFI = .93; RMSEA = .068 (90% confidence interval for the RMSEA = .050–.084). However, when factor loadings were estimated freely in groups (M5), results, $\chi^2(160, N = 223) = 229.58$, $p < .05$; CFI = .94; RMSEA = .063 (90% confidence interval for the RMSEA = 0.043–0.080), showed statistically significant difference between the models (33.72,15: $p < .05$). An examination of the modification indices suggested that the difference resulted mainly from the factor loadings of *N*. Indeed, freeing these parameters in groups resulted in a good fit, $\chi^2(173, N = 223) = 248.06$, $p < .05$; CFI = .94; RMSEA = .063 (90% confidence interval for the RMSEA = .044–.080). However, the standardized parameter for the participant loading of N in the Turkish sample went over |1.00, indicating a possibility of improper solution. This parameter's error variance was constrained to be equal to that for the first participant as suggested by Byrne (1998, p.282). Evidently, all these respecifications refer to the partial invariance defined by Kline (2005). A chi-square difference test (18.48,13: $p > .05$) showed that this revised model (M4R) was not worse than M5 and, thus, accepted as better because of its parsimony.

In addition, the results showed that M6 (first-order multigroup CTCU model in which factor correlations were freely estimated in both groups), $\chi^2(165, N = 223) = 254.78$, $p < .05$; CFI = .93; RMSEA = .070 (90% confidence interval for the RMSEA = .053–.087), was not statistically different from M4R (5.22,10: $p > .05$). Because the most parsimonious model among these three models was M4R (first-order multigroup CTCU model, assuming partial invariance between groups), it was considered to be the best model for accounting for the variance in the data.

These results indicated that except for the first-order factor loadings of *N*, all factor loadings and factor correlations were equivalent across groups, which provided strong support for the last step of the analyses in which a higher order model was tested in a multiple-group MTMM analyses. Correlations between the Big Five factors provided additional support for the higher order model. As displayed in Table 4, identifiers of *plasticity* (i.e., E and O) were moderately correlated with each other, whereas their correlations with other factors were all nonsignificant except for the weak correlation between E and A. Similarly, the correlations between A, C, and N were all significant.

A test of the higher order model in which all second-order factor loadings were considered to be equal across groups (M7) resulted in acceptable goodness of fit statistics, $\chi^2(179, N = 223) = 263.69$, $p < .05$; CFI = .93; RMSEA = .065 (90% confidence interval for the RMSEA = .048–.082). In order to understand whether there were differences concerning the paths from first-order factors to the higher order factors, we tested M8 in which these paths were estimated freely in both groups. The results of this model, $\chi^2(176, N = 223) = 260.72$, $p < .05$; CFI = .93; RMSEA = .066 (90% confidence interval for the RMSEA = .048–.082), were not statistically different from those of M7 (2.97,3: $p > .05$), which indicated that the higher order model was equivalent in the two groups.

Table 4
Intercorrelations of the Big-Five factors.

Variable	E	O	A	N
E	–			
O	.43**	–		
A	.17*	.09	–	
N	−.16	.00	−.53**	–
C	.01	.02	.38**	−.28**

Note. $N = 72$ (Turkish); 151 (German); E = Extraversion, O = Openness, A = Agreeableness, N = Neuroticism, C = Conscientiousness.
* $p < .05$.
** $p < .01$.

As can be seen in Fig. 1, the factor loadings for the five factors ranged from .64 to .85 and were all statistically significant, which provided evidence of convergent validity. The factor loadings of the self-rating method were the highest of all factors and provided strong support for the convergent validity of the self-rating method. The factor loadings of informant methods as well were considerably high for all factors, although lower than those of self-ratings.

Finally, error covariances of participant ratings ranged from .04 to .15, whereas they were much higher for both the first (which ranged from .06 to .27) and second (which ranged from .09 to .18) informant ratings, suggesting that the method effect was weaker for self-ratings.

4. Discussion

After Digman's (1997) proposal and empirical evidence that there are two higher order factors of personality, many attempts have been made to replicate the Big Two model. Recent studies have documented that the interpretation of the model is not an artifact and that the BigTwo can be found reliably (e.g., DeYoung, 2006). In the current study, we replicated and extended that research by using sound methodology and powerful statistical techniques in a cross-cultural setting. One potential problem with the exclusive use of self-report data in personality research is that self-reports leave a lot of room for response biases (Moskowitz, 1986), and there is difficulty in determining whether the factors obtained truly represent trait covariation or artifacts that arise from biases associated with self-reports (Campbell & Fiske, 1959). In using data from multiple informants in this study, we tried to address this limitation. In addition, we used two samples that are known to be culturally diverse; thus, we addressed another limitation of previous research that lacked data from non-Western countries and cross-cultural comparisons.

Consistent with our expectations, we found that the Big Five personality dimensions were not orthogonal. For instance, while we found significant moderate correlations between Extraversion and Openness, as well as between Agreeableness and Neuroticism; small but significant correlations were obtained for the relationships between Extraversion and Agreeableness, and between Neuroticism and Conscientiousness. Therefore, our study suggests that the five subscales are correlated in a manner that reflects two higher order dimensions proposed by earlier research. Such relations are consistent with the ones found in earlier studies using samples from Germany (e.g., Nezlek, Schütz, Schröder-Abe, & Smith, 2011) and Turkey (e.g., Şimşek & Yalınçetin, 2010). In contrast to earlier research (e.g., Biesanz & West, 2004; DeYoung, 2006), however, the correlations between the Big Five for self-ratings and informant ratings, as well as the cross-method correlations, were greater in the present study, suggesting that the correlations between the Big Five are not due to rating biases and that the correlations are relatively valid.

The results of the current study suggest that the Big Two, called *alpha* and *beta* by Digman (1997), and named *stability* and *plasticity* by DeYoung (2006), can be found reliably in the BFI-44. Using a multigroup MTMM approach, we provided strong support for the Big Two model of personality both in a German and a Turkish sample, indicating that even when method effects (variance specific to individual informants) were removed, the correlations between the Big Five were retained. Consistent with the model, in both Germany and Turkey, *plasticity* was defined by Extraversion and Openness, and *stability* by low Neuroticism, Agreeableness, and Conscientiousness. These results are in agreement with the findings of DeYoung (2006), who used the same personality inventory,

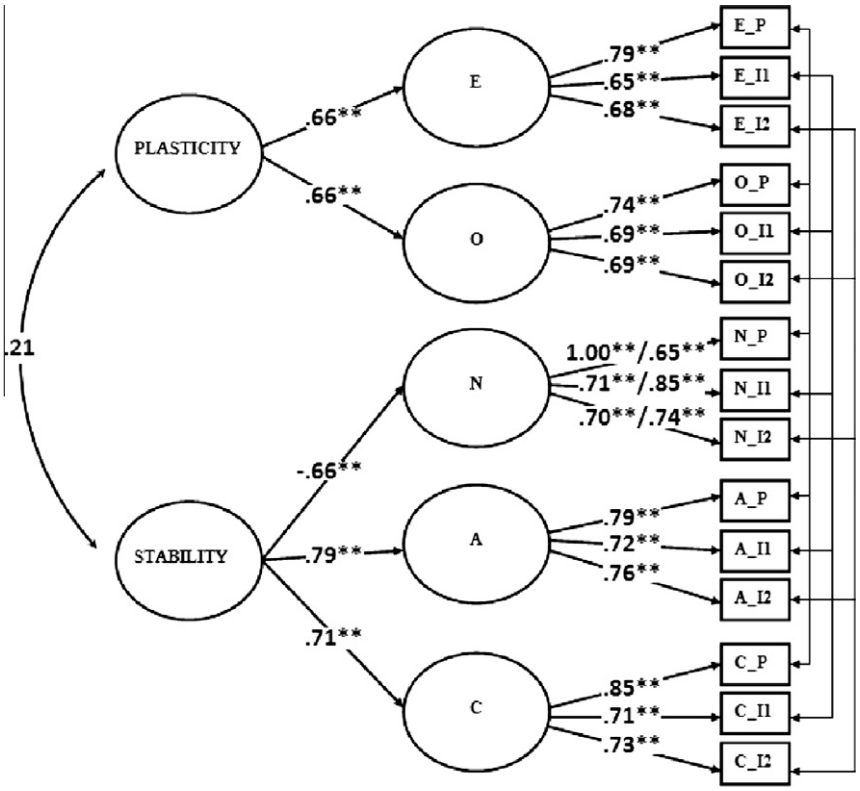

Fig. 1. Multigroup correlated uniqueness model of higher order Big Five model. *Note*. $N = 223$ (Turkish = 72, German = 151); E = Extraversion, O = Openness, N = Neuroticism, A = Agreeableness, C = Conscientiousness, all parameters were constrained to be equal across groups except for those of N; P = Participant, I1 = Informant 1, I2 = Informant 2; values before slashes indicate the values for the Turkish sample. $^{**}p < .01$.

whereas they somewhat contradict those of other researchers (e.g., Ashton et al., 2009; DeYoung et al., 2002) who used different assessment tools. Thus, it seems possible that the validity of the higher order factor structure of personality is partly dependent on the measures used as discussed by Biesanz and West (2004).

The magnitudes of the correlations and loadings on the higher order factors, however, were higher than those found in earlier research (e.g., DeYoung, 2006). Moreover, in contradiction to the findings of DeYoung (2006), Emotional Stability did not appear to be the dominant factor of *stability*. In fact, it had the weakest loading on stability although the results showed that the first-order factor loadings of Neuroticism differed across cultures. Agreeableness was shown to be the primary dimension of *stability* in both cultures. These results could be considered an important contribution to the literature given that cross-cultural equivalence was also verified by the equality of covariance structures. Although we did not mention a need for renaming the higher orders, such differences concerning the higher order factor loadings could necessitate reinterpreting the meanings of the higher order factors. Consequently, future research should include additional replication studies in a cross-cultural context by incorporating a multigroup strategy into the MTMM strategy as we did in the current study.

Regarding the relations between the metatraits, past research had obtained contrasting findings. For instance, whereas Digman's (1997) study demonstrated the orthogonality of *alpha* and *beta*, others (e.g., DeYoung et al., 2002; Hull & Beaujean, 2011; McCrae et al., 2008; Musek, 2007) found that two higher order factors were fairly strongly correlated. DeYoung (2006), on the other hand, showed that when assessed by both the BFI and the Mini-Markers, the metatraits stood as independent factors. In a recent study using self- and informant reports, Slobodskaya (2011) found that the superorder factors were slightly positively correlated in an adolescent sample. In the current study, we found a positive and small nonsignificant correlation between the two higher order factors of personality. The differences between study findings regarding the magnitude of the relation between *stability* and *plasticity* are likely to be due to methodological differences. For example, participant reports are known to be subject to biases due to social desirability and common methods (DeYoung, 2006; Schimmack, 2010). Besides, correlations between higher order factors in studies using a single informant might be artefactual (DeYoung, 2006). However, in the present study, despite the use of an MTMM analysis, the correlation between *stability* and *plasticity* was still weak and nonsignificant.

Finally, we obtained high inter-rater agreement between self- and informant ratings, which was reflected by high loadings on the first-order factors for both types of reporting. The high factor loadings indicate that the informant reports are as reliable as participant reports at least for the present data and with the inventory used. More importantly, when there is high inter-rater agreement, and when there are significant correlations among the Big Five, we can more accurately suggest that the personality dimensions are not orthogonal, and the model with two higher order factors can be said to be verified.

Some limitations of the study should be pointed out. First, the sample size was small, especially for the data from Turkey. Related to this limitation, second, the small sample size precluded using the classical MTMM strategy proposed by Campbell and Fiske (1959). A more rigorous test of the validity of the results could, thus, be obtained by using different methods of analysis with a larger sample size. Third, the generalizability of the findings should be interpreted within the characteristics of the sample, which reflected certain parts of the societies involved. In order to increase the validity of the findings, more elaborated sampling methods are needed. Finally, although we assumed the orthogonality of method effects using uniqueness models, future studies may consider using alternative methods in different research designs.

Despite its limitations, this study used highly effective statistical analyses of the data and provided support for the higher order personality factors. It also showed that the model does not seem to be culture-bound because similar findings were obtained in two countries. Considering the fact that structural relations vary across studies, mostly due to the specific measures used to assess the Big Five, further research is still needed to replicate this study by using different measurement instruments and by sampling diverse populations. Despite the convergent validity of *stability* and *plasticity* provided in the current study, it would not be accurate to assume that these factors represent the only valid personality hierarchies. Additional studies providing further tests of the structure of personality traits on the basis of powerful statistical analyses and strong methodology would definitely contribute to the literature.

Acknowledgment

We are grateful to the Alexander von Humboldt Foundation for supporting this research.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jrp.2012.04.005.

References

Alkan, N. (2006). *Reliability and validity of the Turkish version of the Big Five Inventory*. Unpublished manuscript. Ankara, Turkey: Atilim University.

Anusic, I., Schimmack, U., Pinkus, R. T., & Lockwood, P. (2009). The nature and structure of correlations among big five ratings: The halo-alpha-beta model. *Journal of Personality and Social Psychology, 97*, 1142–1156.

Ashton, M. C., Lee, K., Goldberg, L. R., & de Vries, R. E. (2009). Higher-order factors of personality: Do they exist? *Personality and Social Psychology Review, 13*, 79–91.

Becker, P. (1999). Beyond the Big Five. *Personality and Individual Differences, 26*, 511–530.

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.

Biesanz, J. C., & West, S. G. (2004). Towards understanding assessments of the Big Five: Multitrait-Multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *Journal of Personality, 72*, 845–876.

Blackburn, R., Renwick, S. J. D., Donnelly, J. P., & Logan, C. (2004). Big five or big two? Superordinate factors in the NEO five factor inventory and the antisocial personality questionnaire. *Personality and Individual Differences, 37*, 957–970.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum.

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

Chang, L., Connelly, B. S., & Geeza, A. A. (2012). Correction to Chang, Connelly, and Geeza (2011). *Journal of Personality and Social Psychology, 102*(2), 389.

Costa, P. T., & McCrae, R. R. (1985). *The NEO personality inventory manual*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1992). Multiple uses for longitudinal personality data. *European Journal of Personality, 6*, 85–102.

De Raad, B., & Peabody, D. (2005). Cross-culturally recurrent personality factors: Analyses of three factors. *European Journal of Personality, 19*(6), 451–474.

de Vries, R. E. (2011). No evidence for a general factor of personality in the HEXACO personality inventory. *Journal of Research in Personality, 45*, 229–232.

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology, 91*, 1138–1151.

DeYoung, C. G., Hasher, L., Djikic, M., Criger, B., & Peterson, J. B. (2007). Morning people are stable people: Circadian rhythm and the higher-order factors of the Big Five. *Personality and Individual Differences, 43*, 267–276.

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: Are there neuroses of health? *Personality and Individual Differences, 33*, 533–552.

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology, 73*, 1246–1256.

Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait specific method effects in multitrait-multimethod models: A multiple indicator CTC(M-1) model. *Psychological Methods, 8*, 38–60.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*, 1216–1229.

Goldberg, L. R. (1993). The structure of phenotypic personality traits: Authors' reactions to the six comments. *American Psychologist, 48*, 1303–1304.

Hull, D. M., & Beaujean, A. A. (2011). Higher order factors of personality in Jamaican young adults. *Personality and Individual Differences, 50*(6), 878–882.

Jang, K. L., Livesley, W. J., Ando, J., Yamagata, S., Suzuki, A., Angleitner, A., et al. (2006). Behavioral genetics of the higher-order factors of the Big Five. *Personality and Individual Differences, 41*, 261–272.

John, O. P., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality, 2*, 171–203.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York: Guilford.

Jöreskog, K. G., & Sörbom, D. (2006). *LISREL 8.80 for Windows [Computer Software]*. Lincolnwood, IL: Scientific Software International, Inc.

Kagitcibasi, C., Ataca, B., & Diri, A. (2010). Intergenerational relationships in the family: Ethnic, socioeconomic, and country variations in Germany, Israel, Palestine, and Turkey. *Journal of Cross-Cultural Psychology, 41*, 652–670.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112*, 165–172.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: The Guilford Press.

Koydemir, S., & Schütz, A. (2012). Emotional intelligence predicts components of subjective well-being beyond personality: A two-country study using self- and informant reports. *Journal of Positive Psychology*. http://dx.doi.org/10.1080/17439760.2011.647050.

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of correlated traitcorrelated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods, 7*, 228–244.

Lang, F. R., Lüdtke, O., & Asendorpf, J. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. [Validity and psychometric equivalence of the German version of the Big Five Inventory with young, middle-aged, and old adults]. *Diagnostica, 47*, 111–121.

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement, 15*, 47–70.

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., Jang, K. L., Ando, J., Ono, Y., Yamagata, S., Riemann, R., et al. (2008). Substance and artifact in the higher-order factors of the big five. *Journal of Personality and Social Psychology, 95*(2), 442–455.

Moskowitz, D. S. (1986). Comparison of self-reports, reports by knowledgeable informants, and behavioral observation data. *Journal of Personality, 54*, 294–317.

Mount, M. K., Barrick, M. R., Scullen, S. M., & Rounds, J. (2005). Higher order dimensions of the big five personality traits and the big six vocational interest types. *Personel Psychology, 58*, 447–478.

Musek, J. (2007). A general factor of personality: Evidence for the Big One in the five-factor model. *Journal of Research in Personality, 41*, 1213–1233.

Nezlek, J., Schütz, A., Schröder-Abé, M., & Smith, C. V. (2011). A cross-cultural study of relationships between daily social interaction and the five factor model of personality. *Journal of Personality, 79*(4), 811–840.

Olson, K. R. (2005). Engagement and self-control: Superordinate dimensions of Big Five traits. *Personality and Individual Differences, 38*, 1689–1700.

Rushton, J. P., Bons, T. A., Ando, J., Hur, Y.-M., Irwing, P., Vernon, P. A., et al. (2009). A general factor of personality from multitrait-multimethod data and cross-national twins. *Twin Research and Human Genetics, 12*, 356–365.

Rushton, J. P., & Irwing, P. (2008). A general factor of personality (GFP) from two meta-analyses of the Big Five: Digman (1997) and Mount, Barrick, Scullen, and Rounds (2005). *Personality and Individual Differences, 45*, 679–683.

Schimmack, U. (2010). What multi-method data tell us about construct validity. *European Journal of Personality, 24*, 241–257.

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

Schwartz, S. H. (2004). Mapping and interpreting cultural differences around the world. In H. Vinkin, J. Soeters, & P. Ester (Eds.), *Comparing cultures: Dimensions of culture in a comparative perspective* (pp. 43–73). Leiden, The Netherlands: Brill.

Şimşek, Ö. F., & Yalınçetin, B. (2010). I feel unique, therefore I am: The development and preliminary validation of the Personal Sense of Uniqueness (PSU) Scale. *Personality and Individual Differences, 49*, 576–581.

Slobodskaya, H. R. (2011). Two superordinate personality factors in childhood. *European Journal of Personality, 25*, 453–464.

Vaselka, L., Schermer, J. A., Petrides, K. V., & Vernon, K. A. (2009). Evidence for a heritable general factor of personality in two studies. *Twin Research and Human Genetics, 12*(3), 254–260.