

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Journal of Research in Personality

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

Only three personality factors are fully replicable across languages: Reply to Ashton and Lee

Boele De Raad^{a,*}, Dick P.H. Barelids^a, Boris Mlačić^b, A. Timothy Church^c, Marcia S. Katigbak^c, Fritz Ostendorf^d, Martina Hřebíčková^e, Lisa Di Blas^f, Zsófia Szirmák^g

^a Department of Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

^b Institute of Social Sciences Ivo Pilar, Zagreb, Croatia

^c Department of Educational Leadership and Counseling Psychology, Washington State University, United States

^d Department of Psychology, University of Bielefeld, Germany

^e Institute of Psychology, Academy of Sciences of the Czech Republic, Czech Republic

^f Department of Psychology, University of Trieste, Italy

^g Department of Psychology, Free University of Berlin, Germany

ARTICLE INFO

Article history:

Available online 10 May 2010

Keywords:

Lexical approach
Factor replicability
Cross-cultural
Three factors

ABSTRACT

We reply to Ashton and Lee's (2010) comments on our paper on the cross-cultural replicability of trait factors (De Raad et al., 2010). More specifically, we comment on the interpretation of congruence coefficients, the distinction between Agreeableness and Honesty–Humility in three languages, and the inclusion of a lexical study that has a different six-factor structure. In our view, none of the arguments given in Ashton and Lee's paper compels a change in our conclusion, based on the extensive findings of our paper, that only three factors are fully replicable across languages.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

In their comment to our study on the cross-language replicability of personality factors (De Raad et al., 2010), Ashton and Lee (2010) suggest that there are three serious problems with our study, thereby referring to (1) the assessment of factor replicability, in particular to the interpretation of congruence coefficients, (2) the identification of factors from the various lexical studies, in particular to Honesty–Humility and Agreeableness, and (3) the inclusion of the Czech lexical study, in particular to an alleged non-personality-descriptive factor in the Czech study. We reply to each of these points, showing that they are insufficient to compel a change in our conclusions.

2. Some preliminary remarks

The psycholexical approach has suffered from premature claims concerning the universality of personality trait factors (cf., De Raad & Peabody, 2005). Published lexical studies have been performed in Western languages and cultures mainly, the typical exception being Filipino. From De Raad et al., 2010 it is clear that the Filipino trait-structure is most distinct from the other trait-structures. Hardly anything is yet known about trait-structures in Asia at

large, in South America, or in Africa. Therefore, claims concerning universality should be approached with scientific skepticism.

From their commentary it is clear that a central issue for Ashton and Lee is their Honesty–Humility dimension, which they argue is the cross-cultural factor beyond the Big Five. If Ashton and Lee argue that Honesty–Humility replicates across languages, it is not sufficient support for this general conclusion to find some languages where Honesty–Humility does replicate. We found indeed some pair-wise replications for all six factors, as reported in our paper. However, many of the pair-wise comparisons did not give evidence of replicability for the six factors, which is expressed in average congruencies in Table 7 of De Raad et al. (2010), and these findings are sufficient to falsify Ashton and Lee's thesis.

On the stylistic side, the representation of our argumentation by Ashton and Lee is skewed. Ashton and Lee, for example, prefer to leave out the detail “fully” from our conclusion that three factors are *fully* replicable, while such a restriction is expressed throughout our paper. Ashton and Lee continue saying that our findings run contrary to previous suggestions that six factors can be recovered in *various* languages. In the next line, they conclude that six factors are replicable across languages, leaving out their own necessary restriction “various”. Indeed, there have been suggestions (mainly made by Ashton and Lee) that six factors can be recovered in various languages. If one would keep both restrictive wordings, there is less discrepancy than suggested by Ashton and Lee.

* Corresponding author.

E-mail address: b.de.raad@rug.nl (B. De Raad).

It should be added that “previous suggestions” of six cross-cultural factors are at best half of a fair summary of the state of affairs regarding replicability of factors. Others (e.g., Peabody & De Raad, 2002; De Raad, Perugini, Hřebicková, & Szarota, 1998) have given evidence that three or four factors are replicable across languages. Moreover, the persistent claim of six cross-cultural factors by Ashton and Lee stands in strong contrast with the more general context (sketched in De Raad et al., 2010) in which it is (usually empirically) argued that three factors make a better chance to replicate cross-culturally. That context is ignored in Ashton and Lee's comment, as well as in their other lexical publications.

Apart from a possible six-factor structure including Honesty–Humility, other structures with a sixth factor beyond the Big Five, have been suggested in the past (e.g., Andresen, 1995; Becker, 1999; Hogan, 1983). None of the additional factors in those studies are related to Honesty–Humility.

3. The assessment of factor replicability

In the paragraph, where they discuss the interpretation of congruence coefficients, Ashton and Lee also take issue with the procedure of pair-wise comparisons of the various sets of factors, arguing in favor of alternative methods to compare factors. We also discuss these other two points.

4. Interpretation of congruence coefficients

The general finding of the factor comparisons, given through the average congruencies mentioned above (De Raad et al., 2010) is that there is a sharp drop in the overall level of congruencies after three factors. These findings supported our expectation that “there is a bigger chance for a factor structure with three factors to be tenable across languages than a structure with more than three factors” (De Raad et al., 2010, p. 161). We used a bottom threshold of a congruence coefficient of 0.80. Quite a few studies have been performed in the past to arrive at an acceptable criterion below which factors should not be considered similar (e.g., Davenport, 1990; Lorenzo-Seva & Ten Berge, 2006; Nesselroade & Baltes, 1970). Using a critical level of 0.80 as a significant moment of evaluation is quite challenging, considering, for example, the recent support (Lorenzo-Seva & Ten Berge, 2006) for earlier findings by Haven and Ten Berge (Haven & Ten Berge, 1977), that one should not interpret congruence coefficients below 0.85 as indicative of factor similarity. We relaxed that criterion for good reasons, being aware of the fact that a further relaxation drastically increases the chance of false positives. In all cases, the magnitude of the congruence coefficient is an index of the similarity of the factors under comparison: the smaller the coefficient, the lesser the similarity, with a clear drop in average level after three factors being indicative of a sudden increase of disparity.

In Table 1, an illustration is given of what may happen if one would further relax the threshold of 0.80 down to 0.75 (based on Table 7 in De Raad et al. (2010)). At the level of 0.80, for example, for 7 of the 14 languages, only three factors (column “three”) were

replicated on average. At the level of 0.75, for example, for no more than four languages five factors (column “five”, last row) were replicated on average. A full set of six factors was never replicated (last column). Better chances for replicability are indeed expressed in the column “three” frequencies.

5. Comparing pair-wise versus comparing to a hypothesized structure

We pair-wise compared one-, two-, three-, four-, five-, and six-factor solutions of exploratory factor analyses in 14 taxonomies. Our study is one of a large set of psycho-lexical studies over the past few decades that have sought an optimal representation of traits acceptable in a large variety of languages or cultures. The Big Five model has repeatedly been advocated to fulfill that representational role. The psycholexical research literature is, however, replete with studies demonstrating the inadequacy of that model for such a “canonical” role. Ashton and Lee's studies are part of that still ongoing discussion (see also De Raad & Barelds, 2008). In this context, the various taxonomies all contribute, democratically so to speak, to a kind of open market. Each single taxonomy has its own peculiarities, partially due to distinct details in the procedure, and partially due to characteristics of the language, the culture, and the sample. In this ongoing exploratory process it does not make sense to declare any structure or model as prototypical, which is what Ashton and Lee try to argue. We did not test a particular Big Five model – of which there are many (cf., John, 1990), nor did we test any particular six-factor model, such as the HEXACO-model. The latter would lead to circular reasoning, where one assumes a specific configuration of a six-factor model while it has yet to be proven whether such a model makes sense in appropriately capturing the variations in trait meaning in the available taxonomies.

6. Alternative methods to compare factors

Ashton and Lee suggest three other methods of factor comparison to cope with limitations of congruence coefficients, of which two involve the use of marker-scales, and a third involves ratings of similarity of factors based on content. It is good to realize that marker-scales are restricted in coverage of trait meaning and have reduced validity (cf., De Raad & Peabody, 2005; Saucier, 1994; Widiger & Trull, 1997). Marker-scales are typically developed in such a way that they produce a simple structure of a targeted trait-taxonomy, useful mainly in research-settings where time is limited. For a trait-taxonomy in a certain language, it may be possible to develop marker-scales that reach an acceptable level of reliability but with endangered validity. In particular for research-settings such scales may have the function of measuring a well-defined target structure properly, keeping the restrictions in mind. In De Raad et al. (2010), one set of six-factor markers was constructed, from which 14 different subsets were selected based on the availability of trait terms in the pertaining taxonomies. Given these limitations, we considered our markers only useful for the purpose of identifying factors, in order to reduce subjectiveness in factor identification.

The procedure using ratings of similarity of factors is a weak bid, when solid quantitative measures are available such as congruence coefficients. Such a similarity rating procedure forms a good alternative to subjective interpretations of factors. With so many lingually and culturally different taxonomies, one should be aware of the many subjective and objective details in trait selection, semantic densities, sampling of subjects, instructions, and so forth, that characterize each single taxonomy. For a proper comparison, it is crucial that indigenous structures are compared as much as possible in their original format, without further

Table 1

Frequency of average factor replication in 14 six-factor solutions.

Criterion	Number of factors replicated						
	Zero	One	Two	Three	Four	Five	Six
.80	1	2	1	7	3	0	0
.79	1	0	1	8	4	0	0
.78	0	1	0	7	6	0	0
.77	0	0	1	5	7	1	0
.76	0	0	1	4	6	3	0
.75	0	0	1	4	5	4	0

interference. The calculation of congruencies does that. Ratings of factor similarity through raters from only one of the involved languages enhances the chances for the results of being fashioned by an etic perspective.

7. Identification of factors

As an act of leniency, we have used Ashton et al.'s (2004) findings, among other things, to arrive at marker-scales for a more objective way to identify the factors compared in our study. To arrive at a list of possible markers of the six factors, Ashton and Lee themselves have contributed to the compilation process, and their comments and suggestions were complied with, leading to the final marker-scales as given in the appendix of De Raad et al. (2010).

Generally, there were two logical alternatives for the construction of the marker-scales, one involving a broad version of Honesty–Humility including traits such as altruism and helpfulness, and another involving a narrow version (without altruism and helpfulness). The narrow version would certainly have hits in fewer taxonomies, and would thus have relatively little chance for replicability. The broader version could enhance the chance of replicability, but has as a result a narrower Agreeableness factor.

The marker-scales were agreed upon in an open process, following a simple and transparent procedure that geared towards the Ashton et al. (2004) conceptualization of the six-factor model with the broader version of Honesty–Humility. The use of those marker-scales enabled us to identify for each level of extraction the extent to which the factors capture the six-factor meanings, without relying too much on subjective interpretations.

As expected, and as confirmed in De Raad et al.'s (2010) Table 4, the use of the broader Honesty–Humility marker scale (and therefore narrower Agreeableness marker-scale) led to a substantial correlation of 0.59 between Agreeableness and Honesty–Humility marker-scales across the 14 six-factor structures. Moreover, and because of the obvious connection between the two constructs, the marker-scales seem to compete in labeling factors on the basis of their correlations. We take German as an example, but the reasoning also applies to Croatian and Filipino. The correlations (only 0.50 or higher) of the six-factor marker-scales with the six German factors are: (1) 0.91 with the Honesty–Humility marker and 0.70 with Agreeableness marker, (2) 0.83 with the Intellect marker, (3) 0.91 with the Extraversion marker, (4) 0.77 with the Conscientiousness marker, (5) 0.74 with the Emotional Stability marker, and (6) 0.54 with the Agreeableness marker. How would one label the first factor and how the sixth factor? The first 10 distinct terms loading highest positively and the 10 loading highest negatively on the first factor are *humane, warmhearted, considerate, helpful, magnanimous, kindhearted, good-natured, obliging, charitable, and veracious*, versus *greedy, pompous, bossy, domineering, pretentious, devious, possessive, ruthless, selfish, and boastful*. This listing makes the correlations with the Honesty and Agreeableness markers intelligible. Honesty–Humility is, however, clearly too narrow a label for this factor. Since the (narrow) Agreeableness marker also correlates substantially with this factor, that label is more apt because of its more inclusive connotation. At the five-factor level, the first factor is described in very much the same way with *warmhearted, humane, considerate, good-natured, magnanimous, kindhearted, helpful, obliging, charitable, and indulgent* versus *greedy, domineering, ruthless, bossy, devious, tyrannical, pompous, cold, callous, and iron-hearted*. Both these listings convey typical Agreeableness factors as defined by Big Five proponents (as opposed to the more narrow definition of Agreeableness adopted by Ashton and Lee).

The sixth German factor (in De Raad et al., 2010) is described by *meek, skillful, virtuous, artistic, and mild-tempered* versus *pigheaded,*

rigid, stubborn, hot-tempered, uncontrolled, obstinate, and choleric (there are not too many high loading terms on this factor).¹

The correlation of 0.54 with the Agreeableness marker helps to label the factor, but to label this factor Agreeableness is far from satisfactory. Neither is this factor a clear and narrow anger-related Agreeableness factor. This sixth factor correlates 0.55 with the Emotional Stability factor of the five-solution. We believe we correctly and adequately identified the first factor of the six-solution as a rather typical Agreeableness factor (a factor that includes Honesty–Humility traits), and the sixth factor as A' to express the equivocality of that factor.

8. Inclusion of the Czech lexical study

Ashton and Lee argue that the Czech study should be removed from the comparisons because the resulting trait-structure, in particular the sixth factor, was evaluated by them as being non-personality descriptive. We disagree with this approach; if there is anything to blame, the focus should be on the procedure, not on the outcome of the procedure, in particular the actual use in the Czech case of trait-variables that may convey a skills-connotation. There are two proper ways to deal with unexpected results. One is to scrutinize the procedure that led to the results, in case one does not trust them. If the procedure is found correct, the other way is to embrace the results and to make additional attempts to interpret the results as being personality descriptive.

The Czech study was performed according to the most explicit procedure among the lexical studies, namely the German psycho-lexical program. Moreover, of all the taxonomies that followed the German program, the Czech study did that most strictly. According to that program, trait terms were selected from the dictionary following well tested procedures of inclusion and exclusion. The resulting pool of personality-relevant terms was rated on frequency of use, clarity, familiarity, and prototypicality of being dispositional. The top prototypically dispositional set of terms was administered to obtain self-ratings. Because there is no sign of anomaly in this procedure, one should read the results as being personality descriptive.

The five factor Czech structure provided a rather clear-cut description of the Big Five Model. A six-factor solution added what has been called an Achievement–Ability factor that remained stable also in solutions with seven and eight factors. That is what the Czech trait-structure apparently is about. The six-factor structure

¹ In an unpublished conference presentation by Ostendorf and colleagues (2004), some preliminary findings with regard to the German lexical structure were presented. These preliminary findings were based on factor analyses that used pairwise deletion of missing data. For the De Raad et al. (2010) paper, all factor analyses were conducted using the 'replace with mean' option. This accounts for some small differences in the factor loadings. For instance, the 15 highest loading items for the second German factor as reported by Ostendorf Mlačić, Hřebíčková, and Szarota (2004) were: *pigheaded* (–.63), *obstinate* (–.57), *bull-headed* (–.54), *stubborn* (–.53), *obstructive* (–.52), *short-tempered* (–.51), *uncontrolled* (–.49), *violent-tempered* (–.47), and *hot-headed* (–.46) versus *gentle* (.51), *kind* (.47), *obedient* (.45), *soft* (.42), *artistic* (.42) and *patient* (.42). The 15 highest loading items for the corresponding German factor as used in the De Raad et al. De Raad et al. (2010) study were: *pigheaded* (–.59), *rigid* (–.54), *stubborn* (–.51), *hot-tempered* (–.48), *stubborn* (–.48), *uncontrolled* (–.45), *obstinate* (–.45), *hot-headed* (–.43), *violent-tempered* (–.41), and *unartful* (–.38) versus *artistic* (.42), *meek* (.40), *skillful* (.40), *virtuous* (.39), and *artistic* (.39). In addition, the 15 highest loading terms for the sixth factor as reported by Ostendorf (2004) were: *covetous* (–.68), *pompous* (–.65), *swanky* (–.63), *bouncing* (–.63), *avaricious* (–.62), *ostentatious* (–.62), *grabby* (–.60), *greedy for profits* (–.60), *boastful* (–.60), *lust for glory* (–.60) versus *human* (.60), *basically honest* (.56), *good-hearted* (.55), *helpful* (.55), and *honest* (.55). The 15 highest loading terms for the corresponding German factor in De Raad et al. De Raad et al. (2010) were: *greedy* (–.66), *pompous* (–.62), *bossy* (–.59), *domineering* (–.58), *pretentious* (–.58), *greedy* (–.58), *devious* (–.57) versus *humane* (.66), *warmhearted* (.63), *considerate* (.62), *helpful* (.62), *magnanimous* (.61), *kindhearted* (.61), *good-natured* (.61), and *obliging* (.59).

may deviate from expectations, but so do other indigenously derived structures more or less. Moreover, one should not forget that the many languages in the world have as yet been poorly represented in psycho-lexical studies. Ashton and Lee chose to emphasize the skills and dexterity interpretation of the sixth Czech factor. It is true that trait terms such as *agile*, *dexterous*, *handy* and *clumsy* load highest on this factor (respectively, 0.67, 0.64, 0.58, and –0.55), but the interpretation of the factor should also account for the traits *inventive*, *flexible*, *courageous*, *adventurous*, and *creative*, all loading 0.40 or higher on this factor (respectively, 0.44, 0.43, 0.43, 0.41, and 0.40). This sixth factor correlates 0.43 with the Emotional Stability factor of the five-factor solution and seems to relate to the Competence factor observed in De Raad and Barelds (De Raad & Barelds, 2008). For the interpretation, it is crucial to focus on the psychological meaning of such terms as *clumsy* and *handy*. In Dutch, for example, the term *handy* ('handig') has both the skills-connotation and the psychological connotation, as in "handy in dealing with people" [direct translation from Dutch]. Asch (Asch, 1955) has pointed out that the etymology of certain trait terms contains metaphors. 'Warm' and 'cold', for example, originally described thermal properties of things. Many other terms originated as descriptors of physical aspects of things or events. People can be deep, hard, wide, bitter, and so forth; all these terms express dual meanings.

9. A final comment

Let's give in for a moment to Ashton and Lee's wishes, skipping the Czech taxonomy, and switching the factors II and VI for Croatian, Filipino, and German. The "average" congruencies in the last row of Table 7 in De Raad et al. (2010) for a six-factor solution would then change from 0.80, 0.82, 0.79, 0.76, 0.64, and 0.71 into 0.80, 0.79, 0.79, 0.76, 0.63, and 0.75. For other factor solutions there would be no changes. The general conclusion would still be the same: after three factors, the congruencies fall below a level of acceptance of similarity across languages.

References

- Andresen, B. (1995). Risikobereitschaft (R): Der sechste Basisfaktor der Persönlichkeit: Kovergenz multivariater Studien und Konstruktionsexplikation. [Risk preference (R): The sixth basic factor of personality: Convergence of multivariate studies and construct explication]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 16, 210–236.
- Asch, S. E. (1955). On the use of metaphors in the description of persons. In H. Werner (Ed.), *On expressive language* (pp. 29–38). Worcester, Mass: Clark University Press.
- Ashton, M. C., & Lee, K. (2010). On the cross-language replicability of personality factors. *Journal of Research in Personality*, 44, 436–441.
- Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., et al. (2004). A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages. *Journal of Personality and Social Psychology*, 86, 356–366.
- Becker, P. (1999). Beyond the Big Five. *Personality and Individual Differences*, 26, 511–530.
- Davenport, E. C. (1990). Significance testing of congruence coefficients: A good idea? *Educational & Psychological Measurement*, 50, 289–296.
- De Raad, B., & Barelds, D. P. H. (2008). A new taxonomy of Dutch personality traits based on a comprehensive and unrestricted list of descriptors. *Journal of Personality and Social Psychology*, 94, 347–364.
- De Raad, B., Barelds, D. P. H., Levert, E., Ostendorf, F., Mlačić, B., Di Blas, L., et al. (2010). Only three factors of personality description are fully replicable across languages: A comparison of 14 trait taxonomies. *Journal of Personality and Social Psychology*, 98, 160–173.
- De Raad, B., & Peabody, D. (2005). Cross-culturally recurrent personality factors: Analyses of three factors. *European Journal of Personality*, 19, 451–474.
- De Raad, B., Perugini, M., Hřebíčková, M., & Szarota, P. (1998). Lingua franca of personality: Taxonomies and structures based on the psycholexical approach. *Journal of Cross-Cultural Psychology*, 29, 212–232.
- Haven, S., & Ten Berge, J. M. F. (1977). *Tucker's coefficient of congruence as a measure of factorial invariance: An empirical study*. Unpublished manuscript. University of Groningen, The Netherlands.
- Hogan, R. (1983). A socioanalytic theory of personality. In M. Page (Ed.), *1982 Nebraska symposium on motivation* (pp. 55–89). Lincoln: University of Nebraska Press.
- John, O. P. (1990). The "Big Five" factor taxonomy: Dimensions of personality in the natural language and in questionnaires. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 66–100). New York: The Guilford Press.
- Lorenzo-Seva, U., & Ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, 2, 57–64.
- Nesselroade, J. R., & Baltes, P. W. (1970). On a dilemma of comparative factor analysis: A study of factor matching based on random data. *Educational & Psychological Measurement*, 30, 935–948.
- Ostendorf, F., Mlačić, B., Hřebíčková, M., & Szarota, P. (2004). In search of a sixth big factor of personality in four European languages. Paper presented at the 12th European Conference on Personality, July 18–22, 2004, Groningen, The Netherlands.
- Peabody, D., & De Raad, B. (2002). The substantive nature of psycholexical personality factors: A comparison across languages. *Journal of Personality and Social Psychology*, 83, 983–997.
- Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of Personality Assessment*, 63, 506–516.
- Widiger, T. A., & Trull, T. J. (1997). Assessment of the Five-Factor model of personality. *Journal of Personality Assessment*, 68, 228–250.