On the cross-language replicability of personality factors

Michael C. Ashton [a,*], Kibeom Lee [b,**]

[a] *Department of Psychology, Brock University, Canada*
[b] *Department of Psychology, University of Calgary, Canada*

ARTICLE INFO

ABSTRACT

We comment on several aspects of De Raad et al. (2010), who concluded that only three factors of personality description could be recovered across languages. First, small differences in the congruence coefficients of factors were interpreted by De Raad et al. as indicating large differences in factor replicability. Also, in three languages, a factor that corresponded closely to Honesty–Humility was instead identified as Agreeableness. In addition, the analyses included a lexical study in which one factor represented motor skills rather than a dimension of personality. We suggest alternatives to the use of pairwise congruence coefficients in evaluating factor replicability, including marker scale correlations and conceptual similarity ratings. Results based on those methods indicate substantial cross-language replicability of six personality factors.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Recently, De Raad et al. (2010) examined the cross-language replicability of personality factors across 14 studies involving 12 languages, and concluded that "factor solutions with three factors are replicable across languages; solutions with more factors are not." (p. 160). This conclusion runs contrary to previous suggestions that as many as six personality factors can be recovered in studies of the indigenous personality-descriptive adjectives of various languages (e.g., Ashton et al., 2004; Lee & Ashton, 2008; Saucier, 2009). In the present article, we comment on several aspects of De Raad et al.'s study, noting three main problems that influenced the conclusions of that report. We suggest alternative methods of evaluating factor replicability, and we show that results based on those methods support the conclusion that six factors are replicable across lexical studies of personality structure.

2. A brief overview of De Raad et al. (2010)

De Raad et al. examined the similarity of factors between lexical studies by finding congruence coefficients between corresponding factors. In each of the 14 studies, they first obtained the varimax-rotated factors of solutions involving one to six factors inclusive (i.e., 1 + 2 + 3 + 4 + 5 + 6 = 21 factors). Factors obtained from each study were compared with the corresponding factors in the same factor solution obtained from each of the other 13 studies. Specifically, De Raad et al. first performed a targeted orthogonal Procrustes rotation of one factor solution aiming at the other solution, and then computed congruence coefficients between the corresponding factors. (To compute congruence coefficients between lexical factors from different languages, De Raad et al., treated variables having the same English translation as the same variable.)

Following this method, De Raad et al., computed the congruence coefficients of all lexical factors from a given study (i.e., 21 factors) with the corresponding factors of each of the other 13 studies. Finally, for each of the 21 factors of each of the 14 lexical studies, De Raad et al. averaged the 13 congruence coefficients, and then averaged those mean values to obtain the overall mean congruence coefficient for each of the 21 factors. The latter values served as indicators of the overall replicability of each factor. In interpreting those results, De Raad et al. used a value of .80 as a threshold for deciding whether or not a factor was replicable across lexical studies. Following this standard, De Raad et al. concluded that the factors of the three-factor solution were replicable across lexical studies, but that factors of solutions involving more factors were not replicable.

In this article, we bring attention to what we believe to be three serious problems with the analyses by De Raad et al. The first problem involves the interpretation of results based on pairwise congruence coefficients, especially the use of an absolute threshold for evaluating factor replicability and the corresponding interpretation of small differences in congruence as indicating large differences in replicability. In discussing this problem, we note some

---

* Correspondence to: Michael C. Ashton, Department of Psychology, Brock University, St. Catharines, ON, Canada L2S 3A1.
** Correspondence to: Kibeom Lee, Department of Psychology, University of Calgary, Calgary, AB, Canada T2N 1N4.
E-mail addresses: mashton@brocku.ca (M.C. Ashton), kibeom@ucalgary.ca (K. Lee).

alternative methods of evaluating factor replicability, including the use of marker scales and of conceptual similarity ratings. The second problem involves the identification of corresponding factors between lexical studies, and in particular the identification of the Honesty–Humility and Agreeableness factors of three lexical studies. The third problem involves the inclusion of a Czech lexical study that recovered a factor representing motor skills or manual dexterity rather than any dimension of personality variation. We suggest that when these problems are taken into account, the results indicate that six personality factors are indeed replicable across languages.

3. Interpretation of congruence coefficients

We first consider the interpretation of the congruence coefficient results by De Raad et al., particularly with regard to the level of congruence that can be taken as indicating the successful replication of a factor. Those authors adopted a mean value of .80 as an absolute threshold for replicability, whereby factors with mean congruence coefficients of .80 or above were declared as "replicable", and those with mean congruence coefficients below .80 were declared as "not replicable".

As noted by De Raad et al., congruence coefficients as low as .80, or as high as .90, have been conventionally taken as thresholds for factor replicability. These guidelines have been typically used when researchers have compared factors derived from the same set of variables (typically, scales of personality inventories or subtests of intelligence tests) but obtained in different samples. As also noted by De Raad et al., however, the analyses in their review were rather different. Those analyses involved sets of personality adjectives differing in terms of number and content; also, the congruence coefficients were computed on a subset of the total variable set in each language (i.e., the subset that is common to the two languages of a given comparison, as judged from the adjectives' English translations).

De Raad et al. (2010, p. 165) pointed out that "the congruence coefficients are likely to be attenuated by differences in taxonomic procedures, by imperfect translations of trait adjectives, and most important, by the fact that the factors from the different taxonomies are not based on the same trait variables", and hence suggested that the previous guidelines regarding congruence criteria might be overly stringent. They indicated that the lower bound of congruence coefficient thresholds suggested in the literature (i.e., .80) would be appropriate in the context of lexical personality research. This particular threshold permits De Raad et al.'s conclusion that only three lexical factors are replicated across languages. However, there is no objective basis for the choice of .80 precisely as the threshold, rather than some other particular value. Note that if De Raad et al. had adopted .85 as a threshold, then none of the lexical factors would have been judged as replicable. Alternatively, if De Raad et al. had adopted .75 as a threshold, then all but one factor of the five- and six-factor solutions would have been judged as replicable (i.e., all except the Intellect factor).[1] The use of this absolute threshold thus exaggerates the importance of small differences between congruence coefficients slightly above and below this value (see also Saucier, 2009).

This issue concerning the threshold for replicability of a factor between two languages brings us to a more general problem with

the congruence coefficients computed by De Raad et al. (2010). Specifically, De Raad et al. compared the factors of each language not with any prototypical cross-language factors, but rather with the factors of every other language in turn. This pairwise approach is less than ideal (see also Saucier, 2009), because the claim of a cross-culturally replicable structure is *not* equivalent to a claim that each observed solution will be highly similar to every other observed solution. The crucial point instead is that each observed solution will be highly similar to a hypothesized common structure (see discussion in Lee & Ashton, 2008).

Fortunately, however, there are methods that allow the obtained factor solutions to be compared directly with the hypothesized factor solutions, and not merely with other obtained factor solutions. One such method involves finding the correlations of the observed factors with a set of marker scales representing the hypothesized factors. Those marker scales might consist of the same variables from which the factors were obtained, as in the case of the adjective marker scales used by De Raad et al. in examining the lexical factors of various languages. (De Raad et al. constructed lexically-based marker scales for the proposed set of six factors, and computed the correlations of those marker scales with the lexical factors observed in the 14 lexical studies.) Alternatively, the marker scales might consist of variables external to those from which the factors were conducted, as in the case of the HEXACO-PI scales that we have previously used as markers of the hypothesized six-factor structure in examining the lexical factors of various languages (e.g., Lee & Ashton, 2008). Another method involves asking judges to examine the content of the observed factors and the hypothesized factors, and to provide ratings of the overall similarity of content between each observed factor and each hypothesized factors (e.g., Lee & Ashton, 2008).

We can use these three methods to evaluate the replicability of two variants of a factor which differed in their levels of congruence as reported by De Raad et al. (2010). Specifically, consider the Honesty–Humility factor as obtained in the Dutch lexical study and in one of the two English lexical studies (i.e., the one based on the 449-adjective variable set). As reported by De Raad et al., the mean congruence coefficients for this factor, as obtained when other languages' factors were targeted onto the Dutch and English six-factor solutions, were .68 and .84 respectively. Thus, by the .80 congruence threshold used by De Raad et al., the Dutch version of the Honesty–Humility factor has failed to replicate across languages, whereas the English version of the Honesty–Humility factor has replicated well.

But now let us consider the Dutch and English versions of Honesty–Humility in terms of the results obtained from the other methods that we have suggested above for evaluating factor replicability. First, the Dutch and English lexical Honesty–Humility factors both showed their highest correlations (.72 and .71, respectively) with the marker scale constructed by De Raad et al. to represent Honesty–Humility. Next, as we have reported in previous studies examining the Dutch and English lexical factors in relation to the scales of the HEXACO Personality Inventory (see Ashton et al., 2006; Lee & Ashton, 2008), the Dutch and English lexical Honesty–Humility factors both showed their highest correlations (.61 and .52, respectively) with the HEXACO-PI Honesty–Humility scale.[2] Finally, in our recent study of the conceptual similarity between lexical factors and HEXACO factor descriptions (Lee & Ashton, 2008), we found that the Dutch and English lexical factors were both rated as being more similar to the Honesty–Humility fac-

---

[1] As discussed later, there are two additional problems with the De Raad et al.'s method that are specific to the examination of the six-factor solutions. After correcting these problems, the corresponding congruence values for the six factors, as averaged across languages, were .80, .79, .79, .76, .63, and .75. The relatively low congruence value for the "fifth" factor (i.e., Intellect or Openness) is consistent with the well-known variation between lexical studies in the core content of this factor (see, e.g., Ashton et al., 2004).

[2] In the Dutch case, the reported correlation involved a scale score computed from raw self-ratings on the adjectives. However, the same correlation (.61) is obtained with the Honesty–Humility factor obtained when six varimax-rotated factors are extracted from ipsatized self-ratings on the adjectives.