

**On the effects of residual structure on factor congruence following Procrustes rotation**

Peter Hästö and Janne Lönnqvist\*

Abstract

In this note we simulate Big Five type data and derive critical values for congruence coefficients of Procrustes rotated factor pattern matrices. We consider residual structure, arising from extraneous factors like acquiescence, semantic similarity, etc., and find that such structure causes high congruence coefficients. Based on these results we propose a very simple procedure which may be used to estimate the effects of residual structure. The procedure is used to re-evaluate data from Borkenau & Ostendorf (1998) [The Big Five: how useful is the five-factor model in describing intraindividual variations over time? *JRP*, 34: 202–221.]. The reanalysis indicates that the apparent Big Five structure found by Borkenau & Ostendorf can be explained by semantic similarity and affect.

Keywords: Big Five; Procrustes rotation; Congruence Coefficient; Residual Structure; Factor Structure Comparison

Ever since paper and pencil personality questionnaires were first used in personality measurement, researchers have been aware of several sources of unwanted but systematic variance that may affect the outcomes of the measurement procedure. For instance, Cronbach (1946) labeled the tendency to give affirmative answers to test items independent of item content acquiescent response set. Edwards (1953) noted that the more desirable a trait was judged, the more people ascribed that trait to themselves, suggesting that people may be purposefully dishonest in their self-descriptions. These are two examples of response bias which constitutes a potentially considerable source of extraneous variance in personality measurement. More generally, we define extraneous variance as systematic variance in the outcome of the measurement procedure that is not caused by the attribute that one intends to measure (see Borsboom, Mellebergh, & van Heerden, 2004). We show that failure to recognize such extraneous variance may be particularly pernicious in the context of factor (component)<sup>1</sup> analyses aimed at generalizing the applicability of a model from one domain to another.

Currently the most popular model of personality structure is the Big Five (Goldberg, 1990) or the Five-Factor Model (McCrae & Costa, 1996). This model of personality structure has been generalized to describe not only inter-individual variation in personality, but also culture-level differences (McCrae & Terracciano, 2005), inter-individual variation in other species

---

\* This research was supported in part by the Academy of Finland Research Grants 108384 and 12106181.

<sup>1</sup> In the spirit of personality psychology research, we refer also to constructs identified through component analysis, e.g. PCA, as factors.

(Gosling, Kwan, & John, 2003), and intra-individual personality variation over time (Borkenau & Ostendorf, 1998). Principal Component Analyses (PCA) using Varimax or Procrustes rotation is typically used to assess whether the factor structure found in a novel domain (e.g., the intra-individual structure) matches a theoretical or previously found structure (typically the inter-individual structure). The match of the factor solutions is evaluated with the congruence coefficient. Typically, the congruence coefficient has to exceed some critical value for the factors to be considered matching; 0.90 has been regarded as a rule of thumb (e.g., Barrett, 1986).

To our knowledge, the influence of partial or residual structure on factor congruence coefficients has not previously been explored; at least it has not been taken into account in personality psychology research. The null hypothesis in the evaluation of congruence is typically a completely random structure. Even in this setting Mulaik (1972; see also McCrae, Zonderman, Costa, Bond & Paunonen 1996, Paunonen, 1997) found high values of congruence coefficients. This null hypothesis, however, is not reasonable: several well documented effects, including response acquiescence (Cronbach, 1946), the effect of the social desirability of items (Edwards, 1953), item overlap (Nicholls, Licht, Pearls, 1982), and semantic similarity between items (Peabody & Goldberg, 1989), imply that there will always be some residual structure in the data, even if the true effect is absent. One might think that this is just a secondary influence which is subsumed by allowance made for errors.

This is shown not to be the case. We use Monte Carlo simulations to generate data that allows for the estimation of factor congruence with a null hypothesis of a weak structure attributable to extraneous effects. It is found that very weak, but systematic variations may lead to congruence values exceeding e.g. the 0.90 threshold.

As a case example, and as a model of how to use the simulation data we present, we re-evaluate the data reported by Borkenau & Ostendorf (1998) on the applicability of the Big Five factor structure in describing intra-individual variation in personality over time. Interestingly, this study has recently been cited as evidence both for and against the usefulness of the Big Five factor structure in describing intra-individual variation over time (in Schutte, Malouff, Segrera, Wolf, & Rodgers, 2003, and Cervone, 2005, respectively). Borkenau & Ostendorf (1998) themselves also collected data on the strength of the covariate, semantic similarity. However, they stated that they do not know of a method which allows them to evaluate its contributions to the observed factors. In our re-evaluation, we show how extraneous variance due to semantic similarity and affect can account for the findings of the study without the presence of true latent variables.

#### Method

To measure partial structure we recall some concepts. If our data consists of a signal with variance  $A^2$  and an error with variance  $B^2$ , then the signal-to-noise (s2n) ratio is  $A^2/B^2$ . The percentage of systematic variation (reliability) is given by the ratio of the signal variance to the total variance, i.e.  $A^2/(A^2 + B^2)$  if the error and signal are independent.

#### *Data*

The setup for our simulation is the following. Data is generated for 90 subjects on 6 test items for each of the five factors. We think of an ideal simple factor structure, where the first 6

items are supposed to load only on the first factor, with no secondary loadings, and similarly for the four other groups of 6 items. Thus the first six responses ( $x_{i,j}$ ) of each simulated subject  $i$  were generated according to the model

$$x_{i,j} = \pm a f_{1,i} + e_{i,j},$$

where  $f_{1,i}$  is the factor score of subject  $i$  on factor I,  $e_{i,j}$  is the error, and  $a \geq 0$  is a parameter. The "+"-sign was used in three of the responses, and the "-"-sign in the other half. Similarly, there were six responses influenced only by  $f_{2,i}$ , the second factor score, etc.<sup>2</sup>

Each factor score  $f_{k,i}$  and error score  $e_{i,j}$  was drawn from an independent normal distributions with mean 0 and variance 1. Thus the signal-to-noise ratio of each response  $x_{i,j}$  is  $a^2$  and the ratio of systematic to total variance (expected item reliability) is  $a^2/(1+a^2)$ . Since each factor is defined by 6 items, the expected factor reliability is  $6a^2/(1+6a^2)$ .

### Procedure

The simulations were run on a PC using MATLAB version 7.1. The statistical package of MATLAB includes procedures for the generation of pseudorandom numbers with normal distribution, principal component analysis and Procrustes rotation, which were used in our program.

In a PCA analysis 5 components were extracted irrespective of eigenvalues. The components were rotated by an (orthogonal) Procrustes rotation towards a target matrix, which consisted of 1's, 0's and -1's in appropriate places.

The simulations were repeated 10000 times, and the factor congruences were pooled over the five components, to give 50000 congruence ratings in each condition. To evaluate the effects of the number of participants ( $N$ ) and the number of items ( $K$ ) per factor, we ran 3 more simulations. They were identical to the one presented above, except that  $N$  was either 90 or 200, and the number of items per factor,  $K$ , was either 6 or 12.

### Results

**Table 1.** Median, and 95% and 99% confidence intervals of factor congruence coefficients as a function of the signal-to-noise ratio ( $K = 6$ ,  $N = 90$ ).

s2n	Reliability		$K =$		Median	95%		99%	
	Item	Factor	6	30					
0.00	0.00	0.00	33.2%	8.1%	0.347	0.120	0.598	0.074	0.666
0.02	0.02	0.11	33.3%	8.3%	0.407	0.143	0.668	0.084	0.735
0.04	0.04	0.19	33.5%	9.2%	0.465	0.171	0.723	0.104	0.783
0.06	0.06	0.26	33.8%	10.6%	0.520	0.198	0.768	0.121	0.818

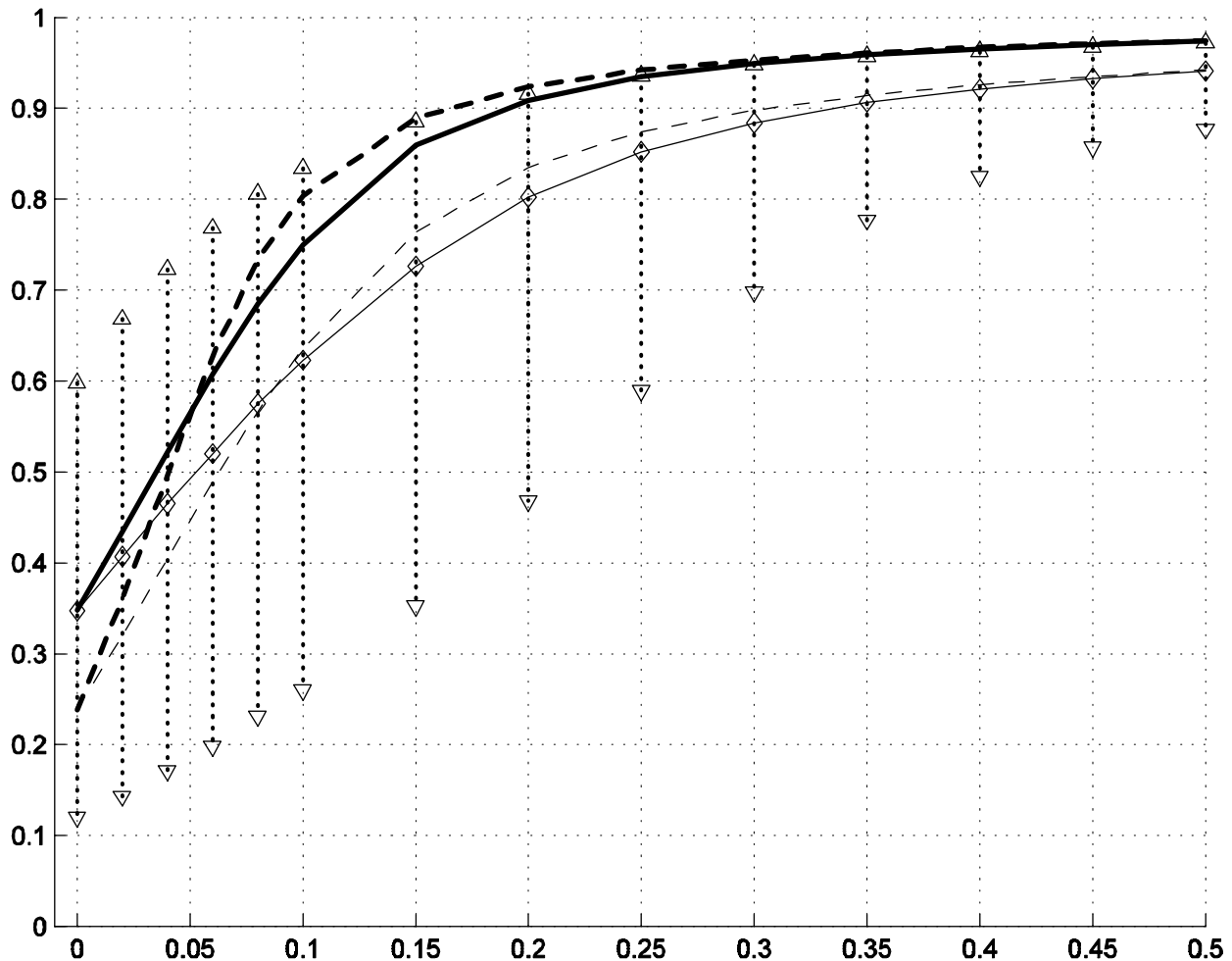
<sup>2</sup> Some authors, e.g. Paunonen (1997), advocate fitting random permutations of data instead of generating new data. The motivation is that secondary loadings will be correctly taken into account. Unfortunately, this method is not applicable in the context of residual structure. Furthermore, as will be shown, even weak residual structure has much greater effect than secondary loadings.

### Residual Structure and Factor Congruence 4

0.08	0.07	0.32	34.2%	12.1%	0.575	0.231	0.806	0.138	0.850
0.10	0.09	0.38	34.6%	13.6%	0.623	0.260	0.834	0.163	0.873
0.15	0.13	0.47	36.1%	17.2%	0.726	0.352	0.885	0.225	0.912
0.20	0.17	0.55	37.8%	20.7%	0.802	0.468	0.916	0.307	0.936
0.25	0.20	0.60	39.7%	23.9%	0.852	0.589	0.936	0.413	0.950
0.30	0.23	0.64	41.5%	26.7%	0.884	0.698	0.948	0.544	0.960
0.35	0.26	0.68	43.5%	29.4%	0.907	0.777	0.957	0.671	0.967
0.40	0.29	0.71	45.2%	31.8%	0.921	0.825	0.963	0.753	0.972
0.45	0.31	0.73	46.9%	34.2%	0.933	0.857	0.968	0.814	0.975
0.50	0.33	0.75	48.6%	36.4%	0.941	0.877	0.972	0.844	0.978
0.55	0.35	0.77	50.1%	38.5%	0.947	0.892	0.975	0.860	0.980
0.60	0.38	0.78	51.6%	40.4%	0.952	0.904	0.977	0.879	0.983
0.65	0.39	0.80	53.0%	42.1%	0.957	0.913	0.979	0.891	0.984
0.70	0.41	0.81	54.4%	43.8%	0.960	0.922	0.981	0.903	0.985
0.75	0.43	0.82	55.6%	45.5%	0.963	0.928	0.982	0.910	0.986
0.80	0.44	0.83	56.8%	47.0%	0.966	0.933	0.984	0.917	0.987
0.85	0.46	0.84	57.9%	48.4%	0.968	0.938	0.985	0.925	0.988
0.90	0.47	0.84	59.0%	49.6%	0.970	0.943	0.986	0.929	0.989
0.95	0.49	0.85	60.0%	51.0%	0.972	0.946	0.986	0.935	0.989
1.00	0.50	0.86	61.0%	52.4%	0.974	0.949	0.987	0.938	0.990

*Note.* The first column gives the signal-to-noise ratio. The second and third columns give the expected item and factor reliabilities calculated from the s2n ratio. Note that the low reliability values follow from the fact that we have only six items per factor. The fourth column indicates how many percent of the variance in the simulation was explained by the first five components. These quite large values are, however, artifacts of the short inventory (a total of 30 items) simulated. Thus the fifth column gives the percentage of variance explained by the first five components in data with the same signal-to-noise ratio, but with  $N = 400$  and  $K = 30$  (i.e., a total of 150 items). As expected, the variance explained in this condition equals the item reliability, up to a small margin. Columns 6–10 give the main results of the simulations: the median, and 95% and 99% confidence intervals of factor congruence values.

Table 1 shows the results of the first simulation. We observe a very rapid increase in the congruences even with very little structure. The two-tailed 95% confidence interval includes the 0.90 criterion already for a signal-to-noise ratio of around 0.18, and the 99% confidence interval includes it already for signal-to-noise ratio around 0.12.



**Figure 1.** The figure displays the congruences obtained from simulated data of 90 (thin lines) or 200 (heavy) subjects and 6 (full) or 12 (dashed) items per factors as a function of signal-to-noise ratio. Also shown are 95% confidence intervals for the case  $N = 90$ ,  $K = 6$ .

Figure 1 shows the results of all four simulations. We see that  $N$  had no effect for  $s2n = 0$ , but became significant when a signal was present. This is probably a consequence of the fact that the signal, when present, could be more accurately estimated when more data was available. The number of items had an effect mostly when  $s2n = 0$ , when the congruence decreased with  $K$ . The effect of  $K$ , although quite small, was consistently reversed for  $s2n > 0$  as compared to  $s2n = 0$ . This follows since increasing  $K$  will also increase the reliability of the factors. For  $s2n = 0$ , the reliability is not affected by  $K$ , but the dimensionality of the problem grows, which accounts for the lower congruences with higher  $K$  in this case. The pattern was confirmed by simulations for  $K = 8$  and  $10$ , but these graphs are not shown.

*A recipe for using the table*

The next section contains a concrete example of the procedure which will be described next, and should help illustrate its use. When you apply this procedure note also the caveats presented in the discussion, below.

The following steps detail the procedure for using the table:

0. Choose a table where the  $K$  and  $N$  match your experiment. These numbers plays a critical role in the outcome. A table for  $K = 6$  and  $N = 90$  is provided above; tables for other values can easily be generated by the program provided<sup>3</sup>.
1. Estimate the original signal-to-noise ratio from the explained variance in a large sample. That is, by looking at the explained variance in column 5 identify an appropriate row. The original signal-to-noise ratio is given in column 1 of that row.
2. Estimate the relative strength of residual structure, e.g. from other experimental data or a theoretical model.
3. Multiply the s2n ratio from step 2 with the percentage from step 3. This is an estimate of the s2n of the residual structure.
4. Use the s2n value from step 4 to find the appropriate row in Table 1. The last four columns of this row give the .95 and .99 confidence intervals for congruence coefficient values.

*An application*

Borkenau & Ostendorf (1998) had 22 subjects answer a 30 item Big Five personality inventory on 90 consecutive days. Participants were instructed to indicate on a seven point scale how well the adjective described how they had been feeling that day.

The 90×30 ratings were subjected to a principal component analysis for each subject, and the resulting structure was rotated by a Procrustes rotation towards the normative structure of the same test applied in an interpersonal setting. The medians of the factor congruences were 0.84, 0.85, 0.73, 0.77 and 0.74, for Neuroticism, Extroversion, Openness, Agreeableness and Conscientiousness, respectively. Obviously, these congruences do not satisfy the 0.90 rule-of-thumb, but the authors nevertheless interpreted them as indicative of the existence a five factor structure in the intrapersonal data.

Borkenau & Ostendorf (1998) also included a simulation study which showed that their fit was well above what would be expected for random data. Their simulation showed that a median of 0.34 should be expected for random permutations of their data, which agrees well with our previous simulations for  $s2n = 0$ .

In this section we consider whether the following stronger null hypothesis can also be rejected:

H0: The data of Borkenau & Ostendorf (1998) can be explained by Positive and Negative affect (the Extraversion and Neuroticism factors, respectively) and by semantic similarity (the Openness, Agreeableness, and Conscientiousness factors).

---

<sup>3</sup> The program can be downloaded from <http://cc.oulu.fi/~phasto/residualstr/>

Note that we could, alternatively, have hypothesized that semantic similarity between items would account for all five factors. However, Konstabel & Virkus (2006) showed, in a somewhat different setting, that Extroversion and Neuroticism are not well explained by semantic similarity. Measures of the two major affective dimensions, Positive and Negative Affect, have been found to be strongly and systematically related to Extraversion and Neuroticism, respectively (e.g., Zuckerman, Kuhlman, Joireman, & Teta, 1993). Furthermore, the subjective experiences of Positive and Negative Affect have been established to fluctuate as a function of daily life events (e.g., Clark & Watson, 1988; Watson, 2000). This means that fluctuations in mood could explain the observed fluctuations in Big Five Extraversion and Neuroticism markers.

To investigate the weaker null hypothesis we follow the steps given above:

0. Six items per factor were used, so Table 1 is appropriate.
1. Since the five factors typically account for at least about 40% of the variance in large data sets (e.g. Peabody & Goldberg, 1989, Table 1), we estimate a signal-to-noise ratio of at least 0.6 from Table 1, row 16.
2. The most direct evidence for a relation between Positive Affect and Extroversion comes from a study by Fleeson, Malanos & Achille (2002), who compared to two in an intrapersonal setting. They found a correlation of 0.69 between them. Thus 48% of the variance was in common. The survey of relations between Big Five factors and PANAS-X factors by Watson (2000, pp. 179–181) indicates that a similar relation holds between Negative Affect and Neuroticism, whereas the other Big Five factors show more modest correlations.

The other three Big Five factors will be explained by semantic similarity of the items. Based on data collected by Borkeu & Ostendorf (1998) in an intra-individual setting, the medial correlation between averaged semantic similarity data and averaged factors is 0.63, so 40% of the variance is in common.<sup>4</sup>

3. By multiplication we find s2n ratios of 0.29 (0.60 x 0.48) for E and N, and 0.24 (0.60 x 0.40) for O, A and C.
4. According to Table 1, this gives an expected median congruence of 0.88 for N and E, and 0.84 for O, A and C.

Recall that the data of Borkeu & Ostendorf (1998) gave congruences of 0.84, 0.85, 0.73, 0.77 and 0.74, for N, E, O, A and C, respectively. In other words, the null hypothesis predicts the data rather well—in fact, better than the conclusion that the usual Big Five structure can be found in the data: the signal-to-noise ratio typically found in Big Five data is around .60, which would imply expected congruences around 0.95. Although we were forced to estimate

---

<sup>4</sup> Note that it would be incorrect to use both affect and semantic similarity as explaining factors, for E and N, since we do not know how much of the explanation provided by these causes is in common.

quite many parameters, there does not seem to be a case for rejecting the null hypothesis, even at the  $p < .05$  level.

There is a slight flaw in the previously described application of the simulation results: in the simulation we assumed that each of the five components had the same signal-to-noise ratio, which was not the case in the application. We performed a new simulation where two factors had s2n ratio 0.29, and the other three had s2n ratio 0.24; this yielded the same critical congruence values as were derived above (which is not surprising, since the factors are orthogonal and hence relatively independent).

### Discussion

There are some discrepancies between our simulated data and the data from Borkenau & Ostendorf (1998). First, we note that the 5 first components in their study accounted for 46–66% of the variance depending on the individual, whereas in the simulation this number was 40%. A second difference is that when Borkenau & Ostendorf found rather high reliability ( $> 0.9$ ) values for their factors; in the simulation, the number was 0.6. These differences probably arise because we included only part of the residual structure which correlates with the Big Five. There is a remaining, systematic part of these factors, which could be captured by PCA but would not contribute to the congruence, since it is not aligned with the Big Five structure.

We have proposed to explain dynamic observations of responses to a personality questionnaire by the static structure of semantic similarity. Of course, the semantic similarity of items does not change over time, yet the experimental data shows that something does change. Therefore, semantic similarity is not a complete explanation. However, when something does change, the semantic similarity of specific items will cause the change to have similar effects across these items. Therefore, the static structure might give rise to the illusion of a dynamic latent variable with the characteristics of latent variable identified in the interpersonal setting.

### Conclusions

In this note we introduced residual structure into simulated data and found that quite low signal strengths give rise to quite high congruence values. In this setting it seems that the 0.90 rule is not especially conservative; in fact, one would hope for much higher congruence values in order to exclude explanations based on weak residual structure.

An apparent weakness in our analysis is the parameters needed to be estimated in order to draw conclusions about the data. In particular, the analysis requires values of the correlations between the factors and the supposed covariates. Further, the relation between this correlation and the implied signal strength is not entirely straightforward. Another short-coming concerns the simplicity of the model: we assumed perfect simple structure in our data. Although these are valid issues, one must compare the method with the alternative: the conventional approach is based simply on setting all of these parameters to zero, and using random data. This surely is a worse estimate.

## Acknowledgment

We would like to thank M. Verkasalo, J. Lipsanen and K. Vehkalahti for comments on earlier versions of this note.

## References

- Barrett, P. (1986). Factor comparison: an examination of three methods. *Personality and Individual Differences*, 7, 327-340.
- Borkenau, P., & Ostendorf, F. (1998). The Big Five: how useful is the five-factor model in describing intraindividual variations over time? *Journal of Research in Personality*, 34, 202-221.
- Borsboom, D., Mellebergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111, 1061-1071.
- Cervone, D. (2005). Personality architecture: within-person structures and processes. *Annual Review of Psychology*, 56, 423-452.
- Clark, L. A., & Watson, D. (1988). Mood and the mundane: relations between daily life events and self-reported mood. *Journal of Personality and Social Psychology*, 54, 296-308.
- Cronbach, L. J. (1946). Response set and test validity. *Educational and Psychological Measurement*, 6, 475-494.
- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 37, 90-93.
- Fleeson, W., Malanos, A. B., & Achille, N. M. (2002). An intraindividual process approach to the relationship between extraversion and positive affect: is acting extraverted as "good" as being extraverted? *Journal of Personality and Social Psychology*, 83, 1409-1422.
- Goldberg, L. R. (1990). An alternative 'description of personality': the Big-Five factor structure. *Journal of Personality and Social Psychology*, 59, 1216-1229.
- Gosling, S. D., Kwan, V. S. Y., & John, O. P. (2003). A dog's got personality: a cross-species comparative approach to personality judgments in dogs and humans. *Journal of Personality and Social Psychology*, 85, 1161-1169.
- Konstabel, K., & Virkus, A. (2006). How similar are the conceptual and empirical structures of personality traits? *European Journal of Psychology*, 20, 337-353.
- McCrae, R. R., & Costa, P. T., Jr. (1996). Toward a new generation of personality theories: theoretical contexts for the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality: Theoretical perspectives* (pp. 51-87). New York, NY: Guilford.
- McCrae, R. R., & Terracciano, A. (2005). Personality profiles of cultures: aggregate personality traits. *Journal of Personality and Social Psychology*, 89, 407-425.
- McCrae, R. R., Zonderman, A., Costa, Jr., P. T., Bond, M., & Paunonen, S. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, 70, 552-566.
- Mulaik, S. A. (1972). *The Foundations of Factor Analysis*. New York, NY: McGraw-Hill.
- Nicholls, J. G., Licht, B. G., & Pearls, R. A. (1982). Some dangers of using personality

- questionnaires to study personality. *Psychological Bulletin*, 92, 572–580.
- Paunonen, S. (1997). On chance and factor congruence following orthogonal Procrustes rotation. *Educational and Psychological Measurement*, 57, 33–59.
- Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology*, 57, 552–567.
- Schutte, N. S., Malouff, J. M., Segre, E., Wolf, A. & Rodgers, L. (2003). States reflecting the Big Five dimensions. *Personality and Individual Differences*, 34: 591–603.
- Watson, D. (2000). *Mood and Temperament*. New York: Guilford Press.
- Zuckerman, M., Kuhlman, D. M., Joireman, J., & Teta, P. (1993). A comparison of three structural models for personality: the Big Three, the Big Five, and the Alternative Five. *Journal of Personality and Social Psychology*, 65, 757–768.