Psychology Science Quarterly, Volume 51, 2009 (1), pp. 65 - 75

A Reanalysis of Toomela (2003): Spurious measurement error as cause for common variance between personality factors

MATTHIAS ZIEGLER[1], AARO TOOMELA[2], & MARKUS BÜHNER[3]

Abstract

The present article reanalyzed data collected by Toomela (2003). The data contain personality self ratings and cognitive ability test results from n = 912 men with military background. In his original article Toomela showed that in the group with the highest cognitive ability, Big-Five-Neuroticism and -Conscientiousness were substantially correlated and could no longer be clearly separated using exploratory factor analysis. The present reanalysis was based on the hypothesis that a spurious measurement error caused by situational demand was responsible. This means, people distorted their answers. Furthermore it was hypothesized that this situational demand was felt due to a person's military rank but not due to his intelligence. Using a multigroup structural equation model our hypothesis could be confirmed. Moreover, the results indicate that an uncorrelated trait model might represent personalities better when situational demand is partialized. Practical and theoretical implications are discussed.

Key words: Personality, Faking, Situational demand, Spurious measurement error

[1] Correspondence should be addressed to Prof. Matthias Ziegler, PhD, Humboldt-Universität Berlin, Rudower Chaussee 18, 12489 Berlin, Germany; email: zieglema@hu-berlin.de

[2] University of Tartu, Estonia

[3] Ludwig-Maximilians-Universität München, Germany

The BIG 5 personality model is the most frequently used psychological construct for personality. The five factor model (FFM) as we know it today, consisting of five higher order personality traits (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness), was proposed by Costa and McCrae (1992). Meanwhile extensive research has been conducted and the Big 5 have been replicated in different cultures and with different instruments. The BIG 5 are often described as five independent dimensions which describe the pattern of covariations between individual personality trait descriptions with sufficient accuracy. Therefore, the most common method to find the BIG 5 is an explorative factor analysis with a Varimax rotation (e.g. Toomela, 2003 and Costa & McCrae). However, there exists a lot of empirical research indicating that the BIG 5 are not as uncorrelated as the application of Varimax rotation would suggest. In a meta-analysis by Mount, Barrick, Scullen, and Rounds (2005) correlations up to .50 between the BIG 5 were reported. Several solutions have been offered to explain these correlations. On the one hand people have argued that this overlap is due to higher order factors (Digman, 1997; Musek, 2007). On the other hand it was argued that social desirability caused this overlap (Bäckström, 2007; Ziegler & Buehner, in press). The present paper reanalyzes data by Toomela (2003) using the hypothesis that a spurious measurement error caused by situational demand is responsible for the overlap between personality traits. This idea was originally proposed by Ziegler and Buehner (in press).

The study by Toomela

In the study conducted by Toomela n = 912 native-born Estonian men aged 17 to 68 years participated and filled out the Estonian version of NEO-PI-R (Kallasmaa, Allik, Realo, & McCrae, 2000), a structure of word meaning test, and the Estonian Cognitive Ability Scale (Pulver, 1999). The analyzed sample, which is also the basis for this paper, contains n = 870 people. Among other things, five cognitive ability groups were established based on scores from an intelligence test. These cognitive ability groups did not only differ in the mean level of the Big Five scores but also in the number of personality factors within an exploratory factor analysis. The unanticipated result was that within the group with the highest cognitive ability, only 3 or 4 factors of personality emerged, even though the groups had comparable variance in their personality scores. Based on these results Toomela suggested that a cultural factor (individual word meaning structure) offers important tools (that is words) for the development of personality, and that intelligence can be understood as the ability to master these tools.

Meanwhile the data have been reanalyzed by Allik and McCrae (2004). They followed recommendations by McCrae, Zonderman, Costa, Bond, and Paunonen (1996) to use an approach of confirmatory analysis. Using a form of Procrustes rotation (Barrett, 1986) Allik and McCrae found that Toomela's data could be fitted quite well in accordance to the American standardization sample. Therefore, one could argue that the problem was solved. However, we want to approach the topic from a different angle. Costa and McCrae (1992, cf. Appendix F, pp. 100 - 101) reported a correlation coefficient of r = -.53 between neuroticism and conscientiousness in the standardization sample which was used as the target in the form of Procrustes rotation. Thus, even though the data obtained by Toomela showed convergence

to another factor solution, the question remains where the overlap between the factors neuroticism and conscientiousness stems from.

Within the following paragraphs three other explanations will be discussed: (1) differences in intelligence as proposed by Toomela, (2) the existence of some higher order personality factors, (3) a spurious measurement error caused by situational demand.

Personality and intelligence. The relationship between personality and intelligence has received a lot of attention in the last decade (e.g., Ackerman, 1996; Furnham, Moutafi, & Chamorro Premuzic, 2005; Toomela, 2003; Ziegler, Knogler, & Bühner, in press). Even though the results repeatedly show significant correlation coefficents between different personality traits and intelligence (Harris, Vernon, & Jang, 2005; Moutafi, Furnham, & Crump, 2003; Moutafi, Furnham, & Paltiel, 2004; Zeidner & Matthews, 2000) these correlation coefficents are rather low to moderate. Moreover, drawing on the differentiation hypothesis from intelligence research (Deary & Pagliari, 1991; Spearman, 1927; Lienert & Crott, 1964), some researchers have assumed that the variability in personality traits increases parallel to the level of intelligence (Brand, Egan, & Deary, 1993; Harris et al., 2005). This indicates that the BIG 5 should actually emerge easier in groups of highly able people. This is also in line with findings by Möttus, Allik, and Pullmann (2006) who also used an Estonian sample. These authors reported only small differences in loadings for a group of cognitive highly able and a group of cognitive less able participants in their study. However, the number of factors was the same. They also reported that intercorrelations between personality dimensions were somewhat smaller within the more able group. However, in the study by Toomela the opposite results could be observed. Thus, it seems unlikely, that individual differences in intelligence were really responsible for the problems in Toomela's data.

Higher order personality factors. Digman (1997) proposed a hierarchical personality model in which agreeableness, conscientiousness, and neuroticism establish a higher order factor, labeled alpha. Extraversion and openness load on a further higher order factor, labeled β. This proposal was based on confirmatory factor analyses (CFA) in several samples. As Digman points out (pp. 1250 – 1253), these factors' content is similar to constructs by other scientists such as Freud, Adler, McAdams, and Hogan. Transferring these results to Toomela's data raises the question whether the increased correlation coefficient between neuroticism and conscientiousness (r=-.71) might be due to a higher order factor. The fact that agreeableness, which loaded higher on alpha in Digman's results than conscientiousness, could be clearly separated from the other two factors of alpha contradicts this hypothesis. Moreover, the correlations between agreeableness and both other factors were below $r = |.32|$ leaving little common variance. Finally, the problem of indistinguishable factors only occurred in the group of highly able men, not in the other groups. A common higher order factor should however impact results in all groups and especially in the lower ability groups according to the differentiation by intelligence hypothesis. Nevertheless, a higher order factor model will be tested in the present reanalysis.

Spurious measurement error. Several authors have argued that the overlap between the Big 5 is caused by a bias (Bäckström, 2007; Bäckström, Björklund, & Larsson, in press; Biesanz & West, 2004). Bäckström reported one higher order factor above the Big 5, which he linked to social desirability. In a follow-up study he showed that the common variance greatly declines using items with low social desirability ratings. Ziegler and Bühner (in press) proposed that high situational demand results in an interaction between person and situation. That means people do fake. However, they do not fake every questionnaire in any

situation. Rather than that they fake depending on the situational circumstances. This means there is an interaction between the person (and how he or she perceives the situation) and the situational characteristics. Such an interaction is called spurious measurement error (Schmidt, Le, & Ilies, 2003). Ziegler and Bühner proposed a structural equation model combined with an experimental design to model such errors. However, the present data do not allow the application of such a model. However, a spurious measurement error is a systematic variance source affecting all traits which are faked. In other words, in a situation in which it is important to appear fearless, stress-resistant, and emotionally stable, neuroticism is likely to be faked (Bradley & Hauenstein, 2006; Khorramdel & Kubinger, 2006; Marcus B., 2006). Consequently, the scores in a questionnaire do not only reflect individual trait differences but also the spurious measurement error. This means that such an error could be modeled as a common method variance. Podsakoff, MacKenzie, Lee, and Podsakoff (2003) suggested to model such variance as a latent variable with loadings from all affected variables. Such an approach has also been suggested by other authors (e.g. Billiet & McKee, 2000; Williams & Anderson, 1994). To confirm that this variable contains faking variance Podsakoff et al. suggested to correlate this latent variable with scores from a social desirability questionnaire. However, such questionnaires contain a substantial amount of personality variance (Paulhus, 1991, 2002). Thus, using this approach would not help to rule out the possibility of a higher order personality factor. Therefore we used a different variable. All participants in the Toomela study were in the Estonian military. However, some were just doing their service time (recruits) while others pursued a professional career (commissioned and non-commissioned officers). Even though data collection was done anonymously faking cannot be ruled out. Empirical evidence shows that even under such conditions large proportions of people tend to fake (Zickar, Gibby, & Robie, 2004). It is plausible to assume that situational demand should be felt differently by the different military groups. Officers who have chosen the military as a career should be more interested to portray themselves in a positive light than recruits doing their military service.

To sum up, there are at least three different explanations for the high intercorrelation between neuroticism and conscientiousness in Toomela's data: (1) differences in intelligence as proposed by Toomela, (2) higher order personality factors, (3) a spurious measurement error. The first possibility seems unlikely considering the results by Möttus, Allik, and Pullmann (2006) as well as the differentiation by intelligence hypothesis (Brand et al., 1993). Consequently, the second and third possibilities will be tested in the following reanalyses.

Method

Statistical analyses. Confirmatory factor analyses (maximum likelihood) were conducted using AMOS 17.0. One important assumption for such an analysis is a multivariate normal distribution (Byrne, 2001; Bühner, 2006). We tested this assumption with the Mardia Test. The result of the Mardia Test (multivariate kurtosis = 22.476, *c.r.* = 7.383, $p < .001$) shows a violation of the assumption. In such a case the χ^2-test is too liberal and the p-value should be corrected. Therefore, we conducted a Bollen-Stine bootstrap with $N = 200$ samples to correct the p-value for the χ^2-test. All results with $p<.05$ are considered aignificant.

The assessment of the global-goodness-of-fit was based on the Standardized Root Mean Square Residual (*SRMR*) and the Root Mean Squared Error of Approximation (*RMSEA*) as

recommended by Hu and Bentler (1999). The authors also give some advice regarding possible cutoffs for the indices. Thus, the *SRMR* should be lower or equal to .11 and the *RMSEA* should be less than .06 for $n > 250$ and less than .08 for $n < 250$. Additionally, we looked at the Comparative Fit Index (*CFI*) as advised by Beauducel and Wittmann (2005). According to Hu and Bentler the *CFI* should have a value of approximately .95.

Sample description. A detailed sample description as well as a description of the used measures can be found in Toomela (2003). Table 1 displays means and standard deviations for the military groups in both relevant personality dimensions. Besides this, Table 1 also includes comparisons between the three military groups: recruits, non-commissioned officers, and officers. It can be seen that both commissioned and non-commissioned officers described themselves as emotionally more stable and more conscientious.

A description of the analyses steps will be reported in the next paragraphs.

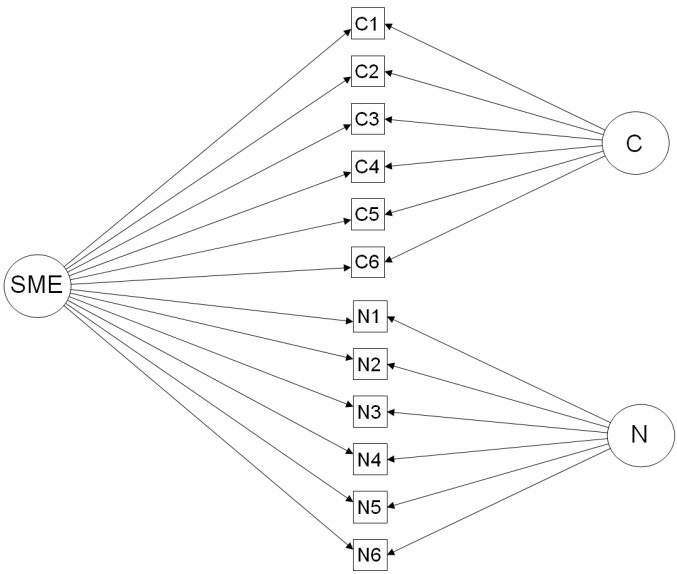
Table 1:
Descriptive statistics for the 3 military groups

	Military groups						*Comparisons*		
	(1) recruits + students[1]		(2) non – commissioned officers[2]		(3) officers[3]		(1) vs (2)	(1) vs (3)	(2) vs (3)
	M	*SD*	*M*	SD	*M*	*SD*	*g*	*g*	*g*
N	85.1	24.8	75.8	27.6	74.2	25.8	**-0.35**	**-0.41**	-0.06
C	115.8	21.7	120.7	24.2	122.5	21.6	**0.22**	**0.30**	0.08
CA	16.9	7.4	21.8	6.6	25.3	5.9	**-0.63**	**-1.09**	**-0.47**

Notations. N = Neuroticism; C = Conscientiousness; CA = Cognitive Ability. [1] $N = 421$; [2] $N = 201$; [3] $N = 248$. The comparisons depict Hedge's g using the pooled standard deviation as denominator. All differences in bold are significant Bonferroni adjusted at $p < .05$. The maximum number of points in the cognitive ability test was 36.

Higher order personality factor. In order to find out whether a higher order personality factor was responsible for the correlation between neuroticism and conscientiousness within the group of cognitive highly able men, a CFA model including neuroticism and conscientiousness facets as manifest variables, neuroticism and conscientiousness as well as α as a higher order factor (hierarchical model) will be compared with a model excluding α (non-hierarchical model). Model comparisons will be conducted using the guidelines suggested by Meade, Johnson, and Braddy (2008). They suggested that a change in CFI as small as .002 indicates a worse model fit.

Spurious measurement error. The spurious measurement error (SME) will be extracted in a bimodal CFA model (SME model) including the same manifest variables as before. However, instead of a second order α factor, a common method factor was included in the model with loadings from all facets (see Figure 1).



Notations. N = Neuroticism, C = Conscientiousness, SME = Spurious measurement error. Facet names of personality domains: N1: anxiety; N2: angry hostility; N3: depression; N4: self-consciousness; N5: impulsiveness; N6: vulnerability; C1: competence; C2: order; C3: dutifulness; C4: achievement striving; C5: self-discipline; C6: deliberation. Error variances have been omitted.

Figure 1:
Common Method Bias CFA Model

This approach also allows a direct comparison of the hierarchical model with the SME model. In a further analysis, a multigroup CFA will be conducted with the military rank as grouping variable. As mentioned Toomela's data contain three different military ranks: recruits[4], non-commissioned officers, and commissioned officers[5]. It will be tested whether there are meaningful differences between the three ranks in any of the model parameters.

In order to conduct the multigroup analysis with latent mean comparisons the intercept for each manifest variable had to be set equal in the three groups (see Byrne, 2001). Moreover, all latent means were fixed at zero in the group containing recruits. Thus, the means in the other groups represent the increase in the latent variable compared to the recruit group.

Results

The results will be reported in the same order as they were outlined above, starting with the analysis of a higher order personality factor.

Higher order personality factor. Model fits for both tested models can be found in Table 2.

[4] The recruits and the students were merged into one group.

[5] The members of the EDL and the commissioned officers were merged into one group.

Table 2:
Model fit for all CFA models.

	χ^2 (df)	*p* - value	*SRMR*	*RMSEA (90% CI)*	*CFI*
Non-hierarchical	1328.40 (54)	<.05	.336	.165 (.157 - .173)	.831
Hierarchical[1]	1328.40 (53)	<.05	. 336	.166 (.159 - .174)	.831
SME model	340.28 (42)	<.05	.039	.09 (.082 - .099)	.961
Multigroup model	483.30 (144)	<.05	.049	.052 (.047 - .057)	.945

Notations: [1] Both loadings on alpha were set equal.

The results show that neither model achieved an acceptable fit. However, the model without the higher order factor had a slightly better RMSEA and is the more parsimonious model. Yet, all in all, neither model can be accepted.

Spurious measurement error. Model fit for the model with a latent variable containing spurious measurement variance was good. Most notably, this model performed better than both other models (ΔCFI = .13). It has to be noted here that no correlation between the Big 5 dimensions was assumed. Consequently, the results provide evidence for the existence of a spurious measurement error. To further explore the character of SME, we regressed it on cognitive ability and military rank (χ^2 (42)=116.78, p <.05, CFI=.93, SRMR=.052, RMSEA=.11). The regression weight for cognitive ability was not significant (a = .12). However, a significant regression weight was found for military rank (a = .41, p <.05).

To further support the idea of a spurious measurement a multigroup analysis was conducted with military rank as grouping variable. If the notion of a spurious measurement error is correct, professional soldiers (non-commissioned and commissioned officers) should achieve larger latent means in the spurious measurement error variable. On the other hand differences in the Big 5 means should decline. This model fitted the data well (see Table 2). Loadings on the personality traits and the SME variable were mostly significant in all three groups. Table 3 displays the latent means for the three groups.

As can be seen, both officer groups differ moderately and significantly from the recruits in SME. Differences in both traits remain but are less pronounced. Thus, these results partly confirm the idea of a spurious measurement error.

Table 3:
Latent means and significance of comparison with recruits

	recruits+ students			non-commissioned officers			commissioned officers		
	M	*S²*	*H*	*M*	*S²*	*H*	*M*	*S²*	*H*
N	0	11.09	-	-1.56	13.03	**-.13**	-2.50	8.87	**-.24**
C	0	4.55	-	1.10	3.28	**.27**	.76	8.21	.13
SME	0	10.04	-	3.35	12.78	**.31**	3.70	11.30	**.35**

Notations. C = conscientiousness, N = neuroticism, SME = spurious measurement error. H = effect size according to Hancock (2001), bold means significant at p <.05. The mean in the groups of recruits and students was fixed at 0.

Discussion

The goal of this paper was to reanalyze the data reported by Toomela controlling for situational influence. Toomela could not distinguish the factors conscientiousness and neuroticism in a group of highly intelligent men. In the course of this reanalysis it was hypothesized that a spurious measurement error was the reason for this problem. The hypothesis was, that not intelligence but military rank was the attribute responsible for the influence of the situational demand. Results show that such a model indeed fits the data better than either a model assuming a higher order factor or a model with simply two latent personality dimensions. The SME was not related to intelligence but related to military rank. It could be shown that the mean for the latent variable SME increases with military rank.

Spurious measurement error. Ziegler and Bühner suggested that faking can be understood as a spurious measurement error. This error variance contains the results from an interaction between person and situation. In this paper we modeled such a variable using an approach proposed by Podsakoff et al. The results provided here confirm the idea by Ziegler and Bühner. More clearly, in the data by Toomela it looks as if men with a higher military rank felt a greater need to obey the situational demand and thus appear as more prototypical military. The result was that they presented themselves as less anxious, more stable, hard-working and conscientious. Thus, a crucial amount of variance in the personality questionnaires was due to the situational demand. It has to be noted, that military rank and intelligence are confounded. However, the analysis also shows that SME is not related to cognitive ability. Hence, it is unlikely that differences in intelligence are behind the latent variable SME.

Another result is that SME has two sides. First it is felt differently by different groups. Recruits probably had the least ambition to appear as a military person. However, with increasing rank the demand was felt stronger. Secondly, the effect of the situational demand also is different in the groups. One could assume the more the situational demand was felt, the less personality variance should be left. However, this is not the case. The personality dimensions are not equally affected. This is in line with results by Pauls and Crost (2005) who demonstrated the selective effects of faking. While neuroticism keeps a big amount of variance, conscientiousness has a lot less variance left after modeling SME. Thus, conscientiousness seemed to have higher importance for the noncommissioned officers and officers in order to appear as a military person than neuroticism. At this point a note of caution is appropriate. The amount of variance due to situational demand is probably overestimated. After all, Deinzer et al. (1995) could show that under anonymous conditions the amount of variance explained by situational influences hardly exceeds 20 %. Only experimental designs can bring more information into this matter. Summing up, there is evidence that a specific "faking" took place which caused the two personality dimensions neuroticism and conscientiousness to correlate highly. By splitting the variance into trait and SME variance the personality dimensions can be better distinguished.

BIG 5 models. The present results replicate findings by Bäckström as well as by Ziegler and Bühner that higher order factors of the Big 5 are most likely linked to social desirability. The fact that they appear even under anonymous conditions should not be surprising. As mentioned before, Zickar et al. could show that even under these circumstances many people tend to portray themselves more positively. Based on the present results as well as on the results by Ziegler and Bühner it seems unlikely to assume that there are substantial factors above the Big 5. This is also supported by findings which attribute such factors a small

amount of heritability (Jang et al., 2006). In the present data the original mean differences in neuroticism and conscientiousness between the groups did not disappeare but only decreased after controlling for SME. However, it can be assumed that the three groups differed in age. Consequently, it is very unlikely that no mean differences at all should be observed (Roberts, Walton, & Viechtbauer, 2006).

Limitations and further research. One major shortcoming of this paper is the fact that this is a reanalysis and not an experimental study. We can only hypothesize that situational demand led higher military ranks to portray stronger faking behavior. One could argue that there are numerous other possible explanations for the differences in the situational demand factor. For example, age or amount of life experience. At least for intelligence the analyses conducted here show that it has little or no influence on the situational demand variable. That is, only with the help of an experimental design would it be possible to find clearer evidence for this hypothesis. Such a design was used by Ziegler and Buehner and their results support the present findings.

Summing up, there is evidence to assume that situational demand was responsible for the large, negative correlation between conscientiousness and neuroticism. Thus, situational demand has an influence on the structure of the BIG 5 and should be regarded with more care as hitherto because by doing this the proposed uncorrelated personality structure is more likely to appear. All in all, regarding the current results an uncorrelated trait model seems very plausible if situational influence is controlled for.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22*, 227-257.

Allik, J., & McCrae, R. R. (2004). Escapable Conclusions: Toomela (2003) and the Universality of Trait Structure. *Journal of Personality and Social Psychology, 87*, 261-265.

Bäckström, M. (2007). Higher-Order Factors in a Five-Factor Personality Inventory and its Relation to Social Desirability. *European Journal of Psychological Assessment, 23*, 63-70.

Bäckström, M., Björklund, F., & Larsson, M. R. (in press). Five-factor Inventories Have a Major General Factor Related to Social Desirability Which Can Be Reduced by Framing Items Neutrally. *Journal of Research in Personality, In Press, Accepted Manuscript*.

Barrett, P. (1986). Factor comparison: An examination of three methods.*Personality and Individual Differences,* 7, 327-340.

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling, 12*, 41-75.

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.

Billiet, J. B., & McKee, J. M. (2000). Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items. *Structural Equation Modeling, 7*, 608-628.

Bradley, K. M., & Hauenstein, N. M. A. (2006). The moderating effects of sample type as evidence of the effects of faking on personality scale correlations and factor structure. *Psychology Science, 48*, 313-335.

Brand, C., Egan, V., & Deary, I. J. (1993). Personality and general intelligence. In Bonaiuto, P., Van Heck, G. L. (Eds), *Personality psychology in Europe*, Vol. 4. (pp. 203-228). Tilburg, Netherlands: Tilburg University Press. xiv, 416 pp.

Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion [Introduction to test- and questionnaire construction]*. München: Pearson.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO Five-Factor inventory (NEO-FFI): Professional Manual.*: Odessa: Psychological Assessment Resources.

Deary, I. J., & Pagliari, C. (1991). The strength of g at different levels of ability: Have Detterman and Daniel rediscovered Spearman's "law of diminishing returns"? *Intelligence, 15*, 247-250.

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

Furnham, A., Moutafi, J., & Chamorro Premuzic, T. (2005). Personality and Intelligence: Gender, the Big Five, Self-Estimated and Psychometric Intelligence. *International Journal of Selection and Assessment, 13*, 11-24.

Goldberg, L. R. (1981). Language and individual differences: the search for universals in personality lexicons. *Review of Personality and Social Psychology, 2*, 141-165.

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and mimic approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika, 66*, 373-388.

Harris, J. A., Vernon, P. A., & Jang, K. L. (2005). Testing the differentiation of personality by intelligence hypothesis. *Personality and Individual Differences, 38*, 277-286.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

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. Vol, 41*, 261-272.

Kallasmaa, T., Allik, J., Realo, A., & McCrae, R. R. (2000). The Estonian version of the NEO-PI-R: An examination of universal and culture-specific aspects of the five-factor model. *European Journal of Personality, 14*, 265-278.

Khorramdel, L., & Kubinger, K. D. (2006). The effect of speediness on personality questionnaires: an experiment on applicants within a job recruiting procedure. *Psychology Science*, 48, 378-397.

Lienert, G. A., & Crott, H. W. (1964). Studies on the factor structure of intelligence in children, adolescents and adults. Vita Humana, 7, 147–163.

Marcus, B. (2006). Relationships between faking, validity, and decision criteria in personnel selection. *Psychology Science, 48*, 226-246.

McCrae, R. R., Zonderman, A. B., Bond, M. H., Costa, P. T., & Paunonen, S. V. (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.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and Sensitivity of Alternative Fit Indices in Tests of Measurement Invariance. *Journal of Applied Psychology, 93*, 568-592.

Möttus, R., Allik, J., & Pullmann, H. (2006). Does personality vary across ability levels? A study using self and other ratings. *Journal of Research in Personality*, (in press).

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. *Personnel Psychology, 58*, 447-478.

Moutafi, J., Furnham, A., & Crump, J. (2003). Demographic and personality predictors of intelligence: A study using the Neo Personality Inventory and the Myers-Briggs Type Indicator. *European Journal of Personality, 17*, 79-94.

Moutafi, J., Furnham, A., & Paltiel, L. (2004). Why is conscientiousness negatively correlated with intelligence? *Personality and Individual Differences, 37*, 1013-1022.

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.

Paulhus, D. L. (1991). Measurement and control of response bias. In P. R. S. In J.P. Robinson, & L.S. Wrightman (Eds.) (Ed.), *Measures of Personality and Social Psychological Attitudes* (pp. 17-59): San Diego, CA: Academic Press.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement.* (pp. 49-69): Mahwah, NJ, US: Lawrence Erlbaum Associates.

Pauls, C. A., & Crost, N. W. (2005). Effects of different instructional sets on the construct validity of the NEO-PI-R. *Personality and Individual Differences, 39*, 297-308.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.

Pulver, A. (1999). Vaimsete vo˜imete skaala VVS2. Unpublished manual. Tartu, Estonia: University of Tartu, Department of Psychology.

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of Mean-Level Change in Personality Traits Across the Life Course: A Meta-Analysis of Longitudinal Studies. *Psychological Bulletin, 132*, 1-25.

Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual-differences constructs. *Psychological Methods, 8*, 206-224.

Spearman, C. (1927). *The abilities of man*: Oxford, England: Macmillan.

Toomela, A. (2003). Relationships between personality structure, structure of word meaning, and cognitive ability: A study of cultural mechanisms of personality. *Journal of Personality and Social Psychology, 85*, 723-735.

Williams, L. J., & Anderson, S. E. (1994). An Alternative Approach to Method Effects by Using Latent-Variable Models: Applications in Organizational Behavior Research. *Journal of Applied Psychology, 79*, 323-331.

Zeidner, M., & Matthews, G. (2000). Intelligence and personality. In Sternberg, R. J. (Ed.), *Handbook of intelligence*. (pp.581 610). New York, NY, US: Cambridge University Press. xiii, 677 pp.

Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods, 7*, 168-190.

Ziegler, M., & Buehner, M. (in press). Modeling Socially Desirable Responding and Its Effects. *Educational and Psychological Measurement*.

Ziegler, M., Knogler, M., & Buehner, M. (in press). Conscientiousness, Achievement Striving, and Intelligence as Performance Predictors in a Sample of German Psychology Students: Always a Linear Relationship? *Learning and Individual Differences*.