University of Tartu

Department of Psychology

René Mõttus

THE RELATIONSHIP OF COGNITIVE ABILITY TO THE STRUCTURE, VARIABILITY AND MEASUREMENT OF PERSONALITY

Master Thesis

Supervisors: Jüri Allik, *PhD*
Helle Pullmann, *MSc*

Keywords: Personality structure; interindividual variability; readable personality inventory; cross-observer-agreement

Tartu 2005

Introduction

The relationship between personality and intelligence, two principal domains of interindividual differences, has traditionally been a matter of ample debate. By now the research has gone beyond investigating linear relationships between psychometrically measured cognitive ability and different personality dispositions. Instead, one line of research has concentrated on the relationship of cognitive ability to the structure and interindividual variability of personality. For example, it has been proposed that individuals at higher level of cognitive ability have more differentiated personality structure (e.g. they might need a greater number of dimensions for a comprehensive description of their personalities or they have trait more orthogonal dimensions than lower-ability individuals; Brand, Egan, & Deary, 1994; Austin, Deary, & Gibson, 1997; Austin, Deary, Whiteman, et al, 2002). Shure and Rogers (1963) and Toomela (2003) noticed that the factor structure of personality questionnaires varied as a function of cognitive ability. In addition, independently of the changes in personality structure, several authors have suggested that interindividual variability of personality test scores is greater in case of high-ability individuals, meaning that they are more dissimilar to each other than lower-ability individuals (Brand et al., 1994; Austin et al., 1997; Austin, Hofer, Deary, & Eber, 2000; Harris, Vernon, & Jang, 2005).

However, all these studies have used self-report ratings, which are prone to measurement errors. For example, it is possible that lower-ability individuals might have difficulties with giving reliable ratings, which in turn might bring along “anomalies” in personality structure or variability of scores. The difficulties might occur either in the level of analyzing item relevant information and making judgments on the basis of this information or at the level of properly understanding and answering questionnaire items. Austin and her colleagues (1997) noticed that internal reliabilities of scales were lower in the group of individuals with lower cognitive ability. They concluded that it was in fact impossible to distinguish measurement confoundings from real differences in variability of personality test scores. In a recent study McCrae and Terracciano (2005) demonstrated that data quality (as a cross-cultural variable) has a considerable relationship with interindividual variability of scores. Allik and McCrae (2004) argued that the less coherent factor structure of the NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) found in individuals with low cognitive ability (Toomela, 2003) might probably be

rather an measurement artifact than real “chaos” in the personality structure itself. Unfortunately Toomela (2003) did not report about data quality (e.g. internal reliability) in different ability groups. In summary, some evidence for the effect of ability on personality features has been reported. Yet, due to the methodological constraints the results of previous studies have been ambiguous.

However, a simple way to overcome limitations of self-report ratings is to obtain additional ratings by well-informed judges. Aggregated ratings are more reliable (Paunonen, 1984). It has also been demonstrated that aggregated observer ratings are more valid source of personality data than self-report ratings (Kolar, Funder, & Colvin, 1996). Thus, the principal aim of the present research was to test whether the relationship of cognitive ability to personality structure and interindividual variability of personality test scores can be replicated in judge rating data. Another extension in comparison with the previous studies was the use of a specially created more readable personality questionnaire, in order to diminish the possible difficulties with understanding personality questionnaire items. If significant and systematic differences between ability groups were observed this would be powerful convergent evidence for previous studies reporting tentative support for the effect of ability on personality structure and interindividual variability. If no such difference appeared, it would rather support the alternative view (Allik & McCrae, 2004) that the variations in personality in different levels of cognitive ability are artifacts caused by measurement errors.

In the first study (Mõttus, Pullmann, & Allik, submitted) we developed a more readable personality measure on the basis of International Personality Item Pool (IPIP; Goldberg, 1999). In fact, it was a study in its own right since besides the development a comprehensive multiscale personality measure we sought to answer the question whether in relatively highly educated populations (probably most convenient samples) the linguistically minimalist personality measure can demonstrate reliability and validity comparable to standard (more sophisticated) measures (e.g. the NEO-PI-R). For example, it is possible that short and minimalist items can be ambiguous, lacking necessary shades of meaning, and therefore bring along loss of exactness in measurement. However, fortunately the results of this study indicated that the linguistically simpler measure performed at least as well as its more sophisticated analogue. More specifically, the more readable instrument Estonian Personality Item Pool NEO (EPIP-NEO; Mõttus et al., submitted) is a parallel of the widely used NEO-PI-R, measuring five personality

dimensions that are described by the 30 facet scales. As even the number of items is identical, the only clear difference between the instruments is the linguistic complexity of the items. The items of the EPIP-NEO are on average 3 words, 7 syllables, and 18 characters shorter than the NEO-PI-R items. The average number of commas is also nearly five times lower in the EPIP-NEO indicating the relative syntactic simplicity. We administered the two instruments in parallel (n = 297), which made it possible to directly compare them. The facet scales of the EPIP-NEO demonstrated on average higher internal reliability (α = .79) than the facet scales of the NEO-PI-R (α = .79). Both factor structures were highly similar to the "normative" North-American structure of the NEO-PI-R (Costa & McCrae, 1992). The relationships with relevant demographic and self-reported behavioral criterions were similar or nearly identical in case of these two instruments. The correlations between corresponding domain scales ranged from r = .83 to .90. At the level of facet scales the correlations ranged from r = .45 to .84 with an average correlation of r = .73. Thus, given that in a convenient sample the measures performed equally it is reasonable to hypothesize that in the case of lower-ability samples the EPIP-NEO would be superior due to its linguistic simplicity.

In the second study (Mõttus, Allik, & Pullmann, unpublished manuscript) we tested the hypothesis about the effect of cognitive ability on the personality structure and interindividual variability in the sample of individuals (n = 154) with various age and educational background. We used the EPIP-NEO for obtaining self and judge ratings on personality traits. On the basis of the used Cognitive Ability Test (CAT) scores we divided the sample in two ability groups (n = 78 and 76, respectively for high and low-ability groups; difference between the CAT score means of two ability groups was 1.60 standard deviations, expressed in IQ points this equals 24 IQ points). Cross-observer agreement on personality traits was generally good in both ability groups, indicating that even low-ability individuals can make valid personality judgments and they can be agreed upon. However, the internal reliability of scales tended to be slightly lower in the group of individuals with lower ability, in both self and judge ratings, indicating possibly lower quality of data in this group. When the structure and interindividual variability of the scores of the EPIP-NEO was examined separately in two ability groups some small but significant differences appeared. The EPIP-NEO domains tended to be less orthogonal and the intercorrelations of the 30 facet scales tended to be higher in the low-ability group. The differences were slightly more pronounced in case of averaged ratings of

judges. Concerning the judge ratings, the factor structure of the EPIP-NEO was also slightly but significantly less similar to the normative structure of the NEO-PI-R (Costa & McCrae, 1992) in low-ability group. Interestingly, this effect was not observed in self-ratings. Nevertheless, it must be noted that the relationship of cognitive ability to the structure of personality was relatively weak and non-systematic. In general, on the basis of present results it can be concluded that ability has some effect on the observed[1] personality structure but it is definitely not large enough to support the hypothesis that personality structure is substantially different at different levels of cognitive ability. The hypothesis of a larger interindividual variability of personality test scores in the group of individuals with a higher cognitive ability found minimal support. In more than a half of the scales standard deviations were indeed higher in the high-ability group but the differences in variance were significant only in a few cases. In addition, these results were ambiguous since the lower-ability group also demonstrated lower internal reliabilities of facet scales, in both self and judge ratings. Internal reliability, in turn, is related to the variance of the scores (for example, see the formula of Cronbach's alpha). To obtain more extreme scores one must respond consistently in a given direction. Inconsistent responding to items leads to both lowered internal reliability and decreased variance of the scores. Thus, it is difficult to establish whether the differences in variance were caused by the higher interindividual differences *per se* or by differences in internal reliability of data. This problem has also been discussed by Austin and her colleagues (Austin et al., 1997; 2000). Similarly, several cross-cultural studies have reported that interindividual variability of personality test scores (including differences between men and women) is higher in Westernisized cultures and lower in economically less developed countries (Costa, Terracciano, & McCrae, 2001; McCrae, 2002; McCrae & Terracciano, 2005). In these data it has also been difficult to find out the exact source of differences, although variance in data quality has been considered a plausible candidate (McCrae & Terracciano, 2005). Unfortunately, the design of the current study did not provide conclusive evidence, since in judge ratings the reliability was also depressed in the group of low-ability targets.

[1] Although personality was rated by multiple observers we cannot draw conclusions about *intrinsic* personality structure without caution since no matter how many raters we have they can only make judgments on the basis of observable attributes of a person, in terms of Five-Factor Theory (FFT; McCrae & Costa, 1999) *characteristic adaptions*. For example, it is possible that there are indeed differences in the range of typical behaviors of high and low-ability individuals, which in turn might influence judgments about *intrinsic* personality. No matter how accurate, observer ratings are only approximations to the real features of an individual's personality.

In general, the present project has three major implications for personality research. First, on the basis of present results it is obvious that linguistically simpler personality measure (the EPIP-NEO) performs at least as well as its more sophisticated counterpart in convenient samples with generally above average level of education. As it is also plausible that the simpler measures can have advantages in populations with lower cognitive ability who might have problems with properly understanding linguistically complex items[2], it is reasonable to propose that more readable personality measures should generally be preferred. Second, since the cross-observer agreement, along with predictive validity, is considered the best indicator that traits are measured accurately (Funder, 1995), adequate 'self-judge' and 'judge-judge' agreement found in both ability groups suggests that making valid judgments about one's personality and reporting these judgments does not require a remarkable amount of cognitive ability. On the contrary, this task was also suitable for individuals with lower levels of ability, given that a readable personality measure was used. Incidentally, numerous reports about investigating cross-observer agreement in different cultures and populations are available (e.g. McCrae, Costa, Martin, et al., 2004; Ready & Clark, 2002; Ready, Clark, Watson, & Westerhouse, 2000) but there are few studies about the effect of ability on the agreement. In this sense, the present results have an important message for personality research, showing that cross-observer agreement is in large part independent of cognitive ability. The final implication based on the results of this study is related to the conclusion that cognitive ability does not have a substantial effect on the personality structure. More specifically, this conclusion means that there is no obvious need for developing different personality models or taxonomies for individuals with different levels of cognitive ability. This is definitely good news for personality research. If our results had demonstrated a remarkable effect, it would have meant that personality research might be even more complicated than it has so far been considered. Not least important is the conclusion that cognitively less able individuals probably do not have more "primitive" and uniform personalities or, as it has been put by Harris and her colleagues (2005), "less" personality.

[2] Evidently, to provide firm conclusion, this hypothesis should be tested.

Acknowledgements[3]

I am most grateful to my supervisors Jüri Allik and Helle Pullmann with whom collaboration has been very smooth and most pleasant. In addition, my special thanks belong to Kenn Konstabel who has been very patient reading my papers and given plentiful helpful advices. Also I would like to thank Airis Erme, Liina Haring and Anu Realo for their participation and comments during the development of the EPIP-NEO and Maie Kreegipuu, Kristo Burk and numerous other enthusiastic persons for their efforts with organizing and carrying out testing.

[3] This project was supported by the Estonian Science Foundation Grant 5677 and an Estonian Ministry of Education and Science Grant 0180543.

References

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.

Austin, E.J., Deary, I.J., & Gibson, G.J. (1997). Relationships Between Ability and Personality: Three Hypotheses Tested. *Intelligence, 5,* 49-70.

Austin, E.J., Deary, I.J., Whiteman, M.C., Fowkes, F.G.R., Pedersen, N.L., Rabbitt, P., Bent, N., & Mcinnes, L. (2002). Relationships Between Ability and Personality: Does Intelligence Contribute Positively to Personal and Social Adjustment? *Personality and Individual Differences, 32,* 1391-1411.

Austin, E.J., Hofer, S.M., Deary, I.J., & Eber, H.W. (2000). Interactions Between Intelligence and Personality: Results From Two Large Samples. *Personality and Individual Differences, 29,* 405-427.

Brand, C., Egan, V., & Deary, I.J. (1994). Intelligence, Personality, and Society: "Constructivist" versus "Essentialist" Possibilities. In D.K. Detterman (Ed.), *Current Topics in Human Intelligence.* Volume 4. (pp. 29-42). Norwood, New Jersey: Ablex.

Costa, P.T.Jr., & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEOFFI) – Professional Manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Costa, P.T., Terracciano, A., & McCrae, R.R. (2001). Gender Differences in Personality Traits Across Cultures: Robust and Surprising Findings. *Journal of Personality and Social Psychology, 81,* 322-331.

Funder, D.C. (1995). On the accuracy of personality judgment: a realistic approach. *Psychological Review, 102,* 652-70.

Goldberg, L.R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In .I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe.* (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

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.

Kolar, D.W., Funder, D.C., & Colvin, C.R. (1996). Comparing the accuracy of personality judgements by the self and knowledgeable others. *Journal of Personality, 64,* 311-37.

McCrae, R.R., Costa, P.T., Martin, T.A., Oryol, V.E., Rukavishnikov, A.A., Senin, I.G., Hrebickova, M., & Urbanek, T. (2004). Consensual Validation of Personality Traits Across Cultures. *Journal of Research in Personality, 38,* 179-201.

McCrae, R.R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: data from 50 cultures. *Journal of Personality and Social Psychology, 88,* 547-61.

McCrae, R.R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparisons. In R.R. McCrae & J. Allik (Eds.), *Five-Factor model of personality across cultures.* (pp. 105-125). Kluwer Academic/Plenum Publishers.

Mõttus, R., Pullmann, H., & Allik, J. (submitted for publication). *Towards More Readable Five-Factor Personality Inventories*. European Journal of Psychological Assessment.

Mõttus, R., Allik, J., & Pullmann, H. (unpublished manuscript). *How Does Cognitive Ability Relate to the Structure, Variability and Measurement of Personality: A Study Using Self and Judge Ratings*. Department of Psychology, University of Tartu.

Paunonen, S. V. (1984). Optimizing the validity of personality assessments: The importance of aggregation and item content. *Journal of Research in Personality, 18,* 411-431.

Ready, R.E., & Clark, L.A. (2002). Correspondence of Psychiatric Patient and Informant Ratings of Personality Traits, Temperament, and Interpersonal Problems. *Psychological Assessment, 14,* 39-49.

Ready, R.E., Clark, L.A., Watson, D., & Westerhouse, K. (2000). Self- and Peer-Reported Personality: Agreement, Trait Ratability, and the "Self-Based Heuristic". *Journal of Research in Personality, 34,* 208-224.

Shure, G.H., & Rogers, M.S. (1963). Personality factor stability for three ability levels.: Personality factor stability for three ability levels. *Journal of Psychology: Interdisciplinary & Applied, 55,* 445-456.

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,* 723-35.

Kokkuvõte eesti keeles

Magistritöö põhiliseks eesmärgiks oli uurida psühhomeetriliselt mõõdetud vaimse võimekuse seoseid isiksuseomaduste struktuuri ja testiskooride variatiivsuses peegelduvate inimestevaheliste erinevustega. Varasemad uurimused on viidanud võimalusele, et vaimselt vähemvõimekate inimeste isiksuseomaduste struktuur võib olla erinev võimekamate inimeste omast. Samuti on oletatud ja tagasihoidlikul määral ka uurimistulemustega kinnitatud, et isiksusetesti skooride variatiivsus on vähemvõimekate inimeste puhul väiksem, mis tähendab, et nad võivad olla üksteisele sarnasemad kui kõrgema võimekusega inimesed. Samas on need uurimused põhinenud enesekohastel küsimustikel, mistõttu on olnud keerukas eristada tegelikke ja metodoloogilistest probleemidest tingitud gruppidevahelisi erinevusi. Nimelt on võimalik, et „anomaalne" isiksusestruktuur või skooride väiksem varieeruvus vähemvõimekate inimeste puhul on tingitud nende kehvemast võimest teha enda kohta järeldusi või mõista sageli üsna keerulisi testiväiteid. Eesmärgiga ületada eelnevate tööde kitsaskohti, kontrolliti käesolevas uurimuses gruppidevahelisi erinevusi, kasutades lisaks inimese enda hinnangutele ka kahe teda hästitundva inimese hinnanguid tema isiksuseomadustele. Lisaks kasutati keeleliselt võimalikult lihtsat viie-faktorilist isiksuseküsimustikku.

Magistritöö koosneb kahest uurimusest. Esimeses uurimuses adapteerisime küsimustikule NEO-PI-R analoogse, ent keeleliselt lihtsama küsimustiku, mille eestikeelne versioon kannab nime EPIP-NEO ja uurisime, kuidas keeleline lihtsus mõjutab testi omadusi. Loodud küsimustiku psühhomeetrilised näitajad olid üldiselt vähemalt sama head kui NEO-PI-R-il. Lisaks märkasime, et skaalade sisereliaablused korreleerusid negatiivselt nende keelelise keerukusega näitajatega, seda mõlema küsimustiku puhul. Tulemused lubasid järeldada, et keeleliselt lihtsamatel küsimustikel on olulisi eeliseid keerukamate ees. Teises uurimuses täitsid inimesed, kellega viidi läbi vaimse võimekuse test, enda kohta EPIP-NEO ja palusid selle nende kohta täita ka kahel neid hästi tundval inimesel. Jagasime valimi vaimse võimekuse alusel võimekate ja vähemvõimekate grupiks ning võrdlesime gruppe isiksusestruktuuri ja inimestevaheliste erinevuste osas, seda nii enesekohtaste kui teiste hinnangute andmetel. Ilmnes, et kuulumisel võimekate või vähemvõimekate gruppi oli isiksusestruktuurile üpris väike mõju, ehkki kohati olid erinevused suuremad juhuslikest. Sarnaselt varasematele uurimustele ilmnes, et võimekamate inimeste grupis kaldus testiskooride varieeruvus

olema suurem. Samas olid erinevused valdavalt statistiliselt ebaolulised. Samuti ei luba skaalade madalam sisereliaablus vähemvõimekate grupis sedastada, kas väiksem varieeruvus selles grupis näitab tõepoolest, et inimesed on üksteisele sarnasemad, või on tegemist madalamast andmete kvaliteedist tingitud kunstlikult vähenenud variatiivsusega.

Kokkuvõtvalt näitasid tulemused, et vaimsel võimekusel puudub isiksuseomaduste struktuurile ja varieeruvusele märkimisväärne mõju. Lisaks, ehkki skaalade sisereliaablused olid vähemvõimekamate inimeste grupis mõnevõrra madalamad, näitab võrdlemisi hea hindajatevaheline kokkulangevus, et ka vähemvõimekad inimesed suudavad enda isiksuseomaduste kohta üsna valiidseid hinnanguid anda.

Original manuscripts

1. Mõttus, R., Pullmann, H., & Allik, J. (submitted for publication). Towards More Readable Five-Factor Personality Inventories. *European Journal of Psychological Assessment*.
2. Mõttus, R., Allik, J., & Pullmann, H. (unpublished manuscript). *How Does Cognitive Ability Relate to the Structure, Variability and Measurement of Personality: A Study Using Self and Judge Ratings*. Department of Psychology, University of Tartu.

Running head: THE IPIP-NEO PERSONALITY INVENTORY

Towards More Readable Big Five Personality Inventories

René Mõttus, Helle Pullmann, and Jüri Allik

Department of Psychology, University of Tartu, Estonia

The Estonian Centre of Behavioural and Health Sciences

Submission date: March 29, 2005

Keywords: International Personality Item Pool; Revised NEO Personality Inventory; readable personality inventories; internal reliability; comparative validity

Corresponding Author:

René Mõttus
Department of Psychology
University of Tartu
Tiigi 78, Tartu 50410
Estonia
Phone: +372 375 902
Fax: +372 376 152
E-mail: Rene.Mottus@ut.ee

Abstract

The Estonian version of the International Personality Item Pool NEO (IPIP-NEO; (Goldberg. 1999) was administered to 297 participants in parallel with the Estonian version of the NEO-PI-R (Kallasmaa, Allik, Realo, & McCrae, 2000). On average, the EPIP-NEO items were 3 words, 7 syllables, and 18 characters shorter than the NEO-PI-R items. By all relevant psychometrical properties the EPIP-NEO was comparable to the NEO-PI-R. The mean convergent correlation between the facet scales was .73. The scales with shorter and grammatically simpler items tended to have higher internal reliability. It is concluded that the EPIP-NEO, as a more readable personality inventory compared to the NEO-PI-R, is suitable for a wider range of samples with different levels of abilities and educational background.

Introduction

The Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) is the most comprehensive and widely used instrument for measurement of the Big Five personality dimensions - Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A), and Conscientiousness (C). Although originally designed for adult populations, it was later applied to college students as well who required separate norms (Sherry, Henson, & Lewis. 2003). Recent studies have shown that even 12-year-old children are able to understand and respond properly to items from the NEO-PI-R or its shorter version the NEO-FFI personality questionnaires (Allik, Laidra, Realo, & Pullmann, 2004; De Fruyt, Mervielde, Hoekstra, & Rolland, 2000; Markey, Markey, Tinsley, & Ericksen, 2002; McCrae, Costa, Terracciano, et al., 2002; Parker & Stumpf, 1998). However, some studies have indicated that the application of the NEO questionnaires may have some limitations. For example, McCrae et al. (McCrae, Costa, Terracciano, et al, 2002) administered the NEO-PI-R to high school students with instructions to leave blank any item they did not understand and found that there were 30 "difficult" items out of 240 containing either obscure terms (such as *fastidious* and *lackadaisical*) or being difficult to comprehend. Further, it has been suggested that some items of the NEO-PI-R should be modified since they are either out of date ("I believe that the 'new morality' of permissiveness is no morality at all") or too difficult for general population (Costa & McCrae, 1997; McCrae & Costa, 2004). Allik and his colleagues (Allik, Laidra, Realo, & Pullmann, 2004) found that the adult factor structure of the Estonian NEO-FFI was less clearly replicated in 12-year-olds than in older adolescents and adult samples. Because similar slight deviations were observed in less educated groups (Allik & McCrae, 2004; Austin, Deary, & Gibson, 1997; Toomela. 2003), it was suggested that a certain minimal amount of abilities is required for observing one's own personality dispositions and for giving reliable self-reports on the basis of these observations (Allik et al., 2004). Clinicists have also observed that comprehension deficits may be responsible for a significant change in validity of some scales (Krakauer, Archer, & Gordon, 1993; Paolo, Ryan, & Smith, 1991). Thus, although the NEO-PI-R has widely proven its validity, a more readable measure of personality would be desirable for a wider range of application including young adolescent and less educated samples.

One promising candidate for a more universal personality measure is the IPIP-NEO questionnaire that has been developed on the basis of the International Personality Item Pool (IPIP; Goldberg, 1999; Johnson, 2005). The IPIP-NEO was designed to match the original NEO-PI-R structure in which five personality dimensions are described by the 30 facets. The IPIP-NEO facet scales consist of ten items each, whereas facet scales of the NEO-PI-R include eight items. According to Goldberg (1999) the average of the coefficient alpha values of the IPIP-NEO scales was a little higher than that of the NEO-PI-R (.80 and .75, respectively), which is almost exactly what could be predicted from the Spearman-Brown prophecy formula. The average cross-instrument agreement between corresponding scales of the NEO-PI-R and the IPIP-NEO were .73 (Goldberg, 1999). Thus, the psychometric quality of the IPIP-NEO scales looks promising. However, more studies, including translations into different languages, are needed for generalizability of this preliminary observation.

This study

Considering the necessity for a more readable personality measure that would be suitable for a wider range of samples with different levels of abilities and educational background, the IPIP-NEO is a promising instrument. The IPIP-NEO items are generally short, using mainly familiar words, and with a simple grammar. It is possible, however, that shorter, less sophisticated, and linguistically minimalist personality items are not able to convey the original meaning. For example, it is claimed that openness to experience is not well represented in natural language (McCrae, 1990) and it may be also problematic to represent it with unsophisticated statements. It is without doubt that simpler and less specific statements become more ambiguous items and can bring in undesirable shades of meaning. In addition, it is possible that easier items are more vulnerable to socially desirable responding that can also compromise the validity (cf. Stricker, 1963).

The main goal of this study is to evaluate psychometrical properties of the Estonian version of the IPIP-NEO, further named as the Estonian Personality Item Pool NEO (EPIP-NEO), which was administered in parallel with the Estonian version of the NEO-PI-R (Kallasmaa, Allik, Realo, & McCrae, 2000). The aim was to establish to what extent the new and more readable personality inventory retained the intended five-factor structure, showed acceptable internal consistency and the external validity compared to the original NEO-PI-R.

Method

Participants

The sample of this study consisted of 297 participants (32% men and 68% women, mean age 31.34 [*SD* = 13.73] years ranging from 16 to 86, for one subject the age was unknown) who completed the test package in 2004. The subjects were instructed to fill out the questionnaires in two subsequent days. In order to avoid practice effects roughly half of the subjects filled the EPIP-NEO on the first day and other half on the second day. The subjects had various educational and professional backgrounds.

Measures

EPIP-NEO. The original IPIP-NEO items were translated by the first author of this article retaining their short and simple style. Five experts, including the co-authors, revised the translated items and made their suggestions how to make the items as unsophisticated and readable as possible. Where it was necessary grammar was simplified and all uncommon words or words with foreign origin were replaced with more common terms. It is suggested to use balanced scales in order to diminish the confounding effects of acquiescent responding (McCrae, Herbst, et al. 2001). In the original IPIP-NEO not all facet scales are balanced in terms of the number of inverted and non-inverted items. In order to control the acquiescence bias, few additional items were generated for unbalanced facet scales. The final item pool with 397 items was back-translated by an English teacher who was unfamiliar with the inventory. The back-translated items were reviewed by John A. Johnson and according to his suggestions some revisions were made. In order to make the EPIP-NEO similar to the NEO-PI-R, eight items were selected for each facet scale. The selection was based on convergent and discriminant relations between items: selected items correlated most with the other items of the intended scale and least with other scales. All facet scales except two (O6: Liberalism and A4: Cooperation) were balanced containing equal number of positively and negatively keyed items.

NEO-PI-R. The main psychometric properties of the Estonian version of the NEO-PI-R are fully described elsewhere (Kallasmaa, Allik, Realo, & McCrae, 2000).

Validity criterions. In order to validate both self-report personality questionnaires, a subset of participants (*n* = 239, 31% men and 69% women with mean age 33.6, *SD* =

13.73) also rated on a 10-point scale (1 – *never*; ... 10 – *several times a day*) the frequency of several concrete behavioral acts which are typical indicators of either neuroticism or extraversion. First, respondents were asked to indicate how frequently they have taken medication against depression. Based on their ratings the subjects were assigned to two groups: those who had taken antidepressants at least once a year (28%) and those who did not report taking the medication (72%). Second, an indicator of social activity was compiled from three items (“I am at a party or crowded event”, “I meet my friends”, and “I ask my friends whether they have plans for party”) and had an internal consistency of $\alpha = .78$.

Results

Linguistic and psychometric properties of the EPIP-NEO and the NEO-PI-R domains are presented in Table 1. Generally, in the EPIP-NEO the items were in average 3 words, 7 syllables, and 18 characters shorter than the NEO-PI-R items. In Estonian, the number of commas can also be used as a fairly good indicator of grammatical complexity. The EPIP-NEO had 0.10 whereas the NEO-PI-R included 0.48 commas per an item.

On average, the Cronbach alphas (see Table 1) of the EPIP-NEO facet scales were slightly higher than those of the NEO-PI-R facet scales (mean values were .79 and .76, respectively). The internal reliabilities were related to the linguistic properties of items: the shorter items tended to be internally more consistent. The correlation between the average number of letters and the Cronbach alphas was r = -.56 (p = .001) and r = -.33 (p = .08) for the EPIP-NEO and the NEO-PI-R, respectively. Assuming that shorter items are more readable, this observation was consistent with previous reports that more comprehensible items lead to more stable and valid scales (Angleitner, John, & Löhr, 1986).

The mean values and standard deviations of the EPIP-NEO and the NEO-PI-R facet scales were rather similar. The mean profiles of the EPIP-NEO and the NEO-PI-R were similar and highly correlated, r = .86 (p < .001). The convergent correlations between the corresponding scales (last column in Table 1) ranged from r = .45 to .84 with an average correlation of r = .73. The correlation of r = .52 between O6 (Liberalism/Openness to Values) scales can easily be interpreted as a result of lower internal consistencies of this facet in both questionnaires. When corrected for attenuation the correlation increased to r = .75. The modest correlation of r = .45 between A3 (Altruism) scales can be considered as the result of a slightly different angle of two inventories towards the construct. Specifically, the items of the A3 in the NEO-PI-R more generally tap being kind, friendly and likable, whereas items of the corresponding scale in the EPIP-NEO more specifically asked whether one was concerned about others wishes and also willingness to help others ignoring his or her own needs.

Table 1. Linguistic and psychometric properties of the EPIP-NEO and the NEO-PI-R scales.

	EPIP-NEO							**NEO-PI-R**							
	Words	*Syllables*	*Letters*	*Commas*	*α*	*Mean*	*SD*	*Words*	*Syllables*	*Letters*	*Commas*	*α*	*Mean*	*SD*	*r*
Neuroticism	4.94	10.21	24.73	0.06	.95	75.82	28.08	7.54	16.63	40.46	0.38	.94	85.84	26.09	.89
Extraversion	4.69	10.27	25.10	0.02	.93	111.05	25.38	7.56	17.67	42.79	0.27	.94	110.82	27.05	.90
Openness	5.98	13.56	32.79	0.13	.89	128.16	19.93	9.77	22.94	56.69	0.67	.89	117.31	21.50	.86
Agreeableness	5.65	12.06	28.96	0.23	.90	126.26	20.59	8.33	18.67	45.69	0.61	.87	118.01	18.28	.83
Conscientiousness	5.54	11.29	27.42	0.08	.93	120.89	23.79	8.21	18.40	44.46	0.48	.91	112.36	22.14	.87
N1 Anxiety	4.75	10.00	24.00	0.00	.86	15.97	6.70	6.38	13.63	32.63	0.25	.85	15.98	6.31	.82
N2 Anger	3.88	8.13	19.63	0.00	.91	11.98	7.26	6.88	15.75	38.13	0.38	.79	12.73	5.43	.78
N3 Depression	4.88	9.63	23.00	0.13	.89	11.32	6.72	7.00	15.25	37.25	0.25	.86	14.19	6.59	.84
N4 Self-Consciousness	5.25	10.63	26.75	0.00	.78	12.40	5.59	10.13	21.38	52.75	0.63	.75	15.30	5.20	.73
N5 Immoderation	5.50	11.38	26.63	0.25	.73	14.60	5.18	6.25	14.13	34.00	0.25	.70	17.45	4.94	.66
N6 Vulnerability	5.38	11.50	28.38	0.00	.84	9.55	5.15	8.63	19.63	48.00	0.50	.85	10.19	5.04	.80
E1 Friendliness	5.50	12.38	30.25	0.00	.83	21.82	5.43	7.38	18.00	42.88	0.25	.76	21.73	5.15	.77
E2 Gregariousness	4.25	10.00	25.38	0.00	.83	17.69	6.33	9.00	19.50	50.88	0.75	.85	16.56	6.53	.84
E3 Assertiveness	4.88	10.63	25.50	0.13	.75	17.24	5.19	8.38	19.38	46.00	0.25	.84	16.64	5.92	.80
E4 Activity Level	4.25	9.50	23.13	0.00	.74	16.99	5.21	6.88	15.25	33.88	0.13	.83	17.92	6.26	.69
E5 Excitement-Seeking	5.13	11.00	26.88	0.00	.78	16.59	6.21	6.88	17.63	43.50	0.13	.73	17.35	5.66	.79
E6 Cheerfulness	4.13	8.13	19.50	0.00	.86	20.72	5.94	6.88	16.25	39.50	0.13	.86	20.63	6.25	.77
O1 Imagination	4.88	11.50	26.13	0.00	.86	23.15	5.77	8.63	21.25	50.25	0.50	.83	20.91	5.89	.78
O2 Artistic Interests	5.38	11.13	29.38	0.13	.79	24.61	5.24	8.75	18.25	47.13	0.50	.80	20.91	5.90	.76
O3 Emotionality	6.00	13.25	31.88	0.13	.83	22.75	5.24	9.25	22.25	52.63	0.38	.80	22.18	4.91	.78

(Table Continues)

Table 1 (continued): *Linguistic and psychometric properties of the EPIP-NEO and the NEO-PI-R scales*.

	EPIP-NEO							**NEO-PI-R**							
	Words	*Syllables*	*Letters*	*Commas*	*α*	*Mean*	*SD*	*Words*	*Syllables*	*Letters*	*Commas*	*α*	*Mean*	*SD*	*r*
O4 Adventurousness	5.75	12.25	29.38	0.13	.69	18.77	4.40	9.50	21.75	56.13	0.63	.65	15.16	4.78	.63
O5 Intellect	5.88	14.88	35.88	0.00	.79	22.26	5.36	8.75	23.13	56.38	0.38	.86	18.89	6.63	.74
O6 Liberalism	8.00	18.38	44.13	0.38	.58	16.62	4.54	13.75	31.00	76.75	1.63	.61	19.26	4.42	.52
A1 Trust	5.75	12.63	28.25	0.38	.84	20.78	5.34	7.75	18.25	43.13	0.88	.79	20.71	5.07	.78
A2 Morality	6.63	12.38	31.25	0.50	.79	22.79	5.54	9.13	20.38	47.88	0.88	.76	19.26	5.46	.72
A3 Altruism	5.38	11.00	28.50	0.13	.79	22.52	4.39	7.88	18.50	46.25	0.38	.63	21.31	3.76	.45
A4 Cooperation	4.50	9.75	23.75	0.00	.69	19.64	4.86	8.50	17.50	43.50	0.63	.62	15.81	4.38	.63
A5 Modesty	5.38	12.75	28.75	0.00	.79	17.24	5.71	7.38	15.00	37.50	0.50	.8	18.91	5.63	.80
A6 Sympathy	6.25	13.88	33.25	0.38	.77	23.29	4.87	9.38	22.38	55.63	0.38	.55	22.02	3.91	.66
C1 Self-Efficacy	5.38	10.38	24.88	0.25	.74	20.18	4.23	7.13	17.13	42.13	0.38	.66	19.09	4.36	.68
C2 Orderliness	5.38	10.88	26.13	0.00	.84	19.86	6.18	8.75	18.50	43.75	0.50	.70	17.90	5.00	.71
C3 Dutifulness	4.88	9.75	24.25	0.13	.80	22.62	4.93	7.25	17.13	41.00	0.25	.67	22.08	4.47	.69
C4 Achievement Striving	6.38	12.38	30.25	0.00	.75	20.95	5.04	8.50	18.38	46.38	0.50	.75	17.72	5.40	.70
C5 Self-Discipline	5.50	10.75	26.00	0.00	.87	17.68	6.36	11.13	23.50	55.00	1.00	.77	18.22	5.08	.82
C6 Cautiousness	5.75	13.63	33.00	0.13	.77	19.61	5.06	6.50	15.75	38.63	0.25	.73	17.36	4.91	.73

NOTE: n = 297; *Words* = average number of words per item; *Syllables* = average number of syllables per item; *Letters* = average number of letters per item (without spaces and punctuation marks); *Commas* = average number of commas per item; α = Cronbach's alpha; *Mean* = mean scores; *SD* = standard deviation of scores; r = correlation between corresponding scales of the EPIP-NEO and the NEO-PI-R (all correlations are significant at $p < .01$). Only the facet names of the EPIP-NEO are provided.

Principal component analysis followed by Varimax rotation applied to 30 facet scales of the EPIP-NEO revealed a typical five-factor structure (see Table 2). The first seven eigenvalues were 6.53, 5.06, 2.98, 2.49, 1.67, .99 and .93. Parallel analysis (Zwick & Velicer, 1986) and other criteria (e.g. scree test) suggested to retain five factors which accounted for 62.4% of the total variance (in case of the NEO-PI-R the respective figure was 61.90). The factor congruence coefficients between Varimax-rotated factor structures of the EPIP-NEO and the NEO-PI-R were .98, .97, .95, .95 and .97 for the N, E, O, A and C factors, respectively. Generally, the typical five-factor structure of the NEO-PI-R was replicated by both inventories. After Procrustes rotation targeted at the North-American normative structure of the NEO-PI-R (Costa & McCrae, 1992), the factor congruence coefficients between North-American factors and respective factors obtained from the current data ranged from .95 to .96 and from .95 to .98 for the EPIP-NEO and the NEO-PI-R, respectively. However, some unintended primary loadings were observed for both measures. First, the N5 (Impulsiveness) of the EPIP-NEO had its primary loading on the C factor and similar tendency appeared in the structure of the NEO-PI-R as the N5 had equal loadings on the intended N factor and C factor. Incidentally, the "wrong" loading of the N5 on the C factor was not specific to this sample as the same has been seen previously (Kallasmaa, Allik, Realo, & McCrae, 2000). Analogously for both inventories A5 (Modesty) had its primary loading on the E factor. Again, the similar deviation has been reported previously (Kallasmaa, Allik, Realo, & McCrae, 2000). More specific to the EPIP-NEO was the unintended primary loading of the O4 (Adventurousness) on the E factor. In the Varimax-rotated structures of the NEO-PI-R data of the current sample and Estonian normative data (Kallasmaa, Allik, Realo, & McCrae, 2000) O4 (Actions) also had relatively strong secondary loadings on the E factor (.23 and .28, respectively) but in the structure of the EPIP-NEO this tendency was much more pronounced. The facet scale A2 (Straightforwardness) of the NEO-PI-R had "wrong" primary loading on the E factor. Interestingly, the corresponding facet A2 (Morality) of the EPIP-NEO had strong secondary loading on the C factor. The Procrustes-rotated structure of the EPIP-NEO, targeted at the normative North-American Varimax-rotated structure of the NEO-PI-R (Costa & McCrae, 1992), slightly more resembles the simple structure since the unintended primary loadings of the N5 and O4 diminished remarkably and the primary loading of A5 resettled on the intended A factor.

Table 2. *Varimax rotated factor structures and congruence coefficients of the EPIP-NEO and the NEO-PI-R.*

	EPIP-NEO						**NEO-PI-R**					
	N	*E*	*O*	*A*	*C*	*Cong*[1]	*N*	*E*	*O*	*A*	*C*	*Cong*[2]
N1 Anxiety	**-.85**	-.08	.01	.09	.01	.98	**-.85**	-.03	.05	-.05	-.01	.98
N2 Anger	**-.75**	.08	-.04	-.31	.02	.96	**-.81**	.09	-.01	-.27	-.09	.98
N3 Depression	**-.80**	-.16	.02	-.14	-.14	.96	**-.84**	-.24	.03	-.03	-.17	.99
N4 Self-Consciousness	**-.61**	**-.43**	-.21	.10	-.29	.95	**-.68**	-.36	-.17	.03	-.16	.98
N5 Immoderation	**-.52**	.20	.13	-.02	**-.60**	.95	**-.47**	.28	.10	.05	**-.47**	.97
N6 Vulnerability	**-.70**	-.32	-.15	-.01	-.34	.99	**-.67**	-.32	-.16	.10	-.32	.98
E1 Friendliness	.29	**.72**	.09	.29	.12	.90	.18	**.71**	.21	.38	.09	.93
E2 Gregariousness	.04	**.78**	-.20	.21	-.07	.97	.10	**.75**	-.03	.11	-.01	.95
E3 Assertiveness	.05	**.76**	.17	-.27	.12	.94	.13	**.65**	.25	-.28	.21	.96
E4 Activity Level	.01	**.62**	.12	-.07	.27	.98	.07	**.68**	.09	.13	**.43**	.95
E5 Excitement-Seeking	.04	**.70**	.16	-.10	-.23	.96	.04	**.72**	.21	-.12	-.05	.98
E6 Cheerfulness	.18	**.69**	.08	.33	-.01	.97	.26	**.70**	.17	.29	.02	.95
O1 Imagination	-.09	.34	**.63**	.04	-.19	.94	-.22	.26	**.64**	-.02	-.19	.97
O2 Artistic Interests	.01	.08	**.61**	**.41**	.07	.96	-.06	.01	**.59**	**.43**	.11	.94
O3 Emotionality	-.30	.33	**.61**	.28	.14	.96	-.24	.36	**.65**	.26	.16	.97
O4 Adventurousness	.07	**.51**	.30	.08	-.14	.84	.18	.23	**.46**	.02	-.22	.94
O5 Intellect	.18	.10	**.77**	-.01	.04	.99	.10	.05	**.80**	-.09	.03	.98
O6 Liberalism	.20	-.05	**.51**	-.09	-.30	.92	.35	.03	**.58**	-.04	-.21	.96
A1 Trust	.33	.12	.14	**.52**	-.08	.97	.31	.07	.16	**.67**	-.02	.98
A2 Morality	.06	-.22	.00	**.60**	**.51**	.95	.11	**-.50**	-.04	**.49**	.22	.97
A3 Altruism	-.01	.09	.06	**.75**	.28	.91	.06	.16	.06	**.66**	.24	.97
A4 Cooperation	.38	-.27	.05	**.67**	.09	.99	.23	**-.44**	-.07	**.58**	-.02	.98

(Table Continues)

Table 2 (continued): *Varimax rotated factor structures and congruence coefficients of the EPIP-NEO and the NEO-PI-R*.

	EPIP-NEO						**NEO-PI-R**					
	N	*E*	*O*	*A*	*C*	*Cong*[1]	*N*	*E*	*O*	*A*	*C*	*Cong*[2]
A5 Modesty	-.03	**-.57**	-.31	**.42**	.16	.90	-.12	**-.57**	-.13	**.46**	.03	.94
A6 Sympathy	-.17	.25	.10	**.76**	.15	.93	-.24	.11	.02	**.70**	.07	.91
C1 Self-Efficacy	**.42**	.33	.18	-.06	**.60**	.99	**.45**	.22	.04	-.04	**.67**	.99
C2 Orderliness	-.10	-.06	-.17	.11	**.72**	.95	-.03	.09	-.09	.01	**.77**	.99
C3 Dutifulness	.04	-.08	-.11	.33	**.74**	.95	.06	-.15	-.11	.33	**.70**	.96
C4 Achievement Striving	.01	.21	.17	.06	**.73**	.97	-.01	.25	-.03	.02	**.80**	.96
C5 Self-Discipline	.19	.13	-.14	.11	**.78**	.98	.22	.12	.05	.12	**.78**	.98
C6 Cautiousness	.17	-.24	.00	.07	**.64**	.99	.19	-.34	-.15	.11	**.61**	.98
	Factor congruence coefficients						Factor congruence coefficients					
	.96	.95	.95	.95	.96		.96	.98	.95	.95	.98	

NOTE: n = 297; Cong1 = congruence coefficients of the EPIP-NEO with American normative structure of the NEO-PI-R (Costa & McCrae, 1992) after Procrustes rotation; Cong2 = congruence coefficients of the NEO-PI-R with American normative structure of the NEO-PI-R (Costa & McCrae, 1992) after Procrustes rotation. Factor loadings above |.40| are in bold. Only the facet names of the EPIP-NEO are provided.

In order to evaluate the contribution of both inventories to the measurement of the underlying five-factor structure, a Multitrait-Multimethod (MTMM) confirmatory factor analysis was performed. Maximum likelihood Correlated Traits Correlated Methods (CTCM) analysis at the domain level gave reasonable estimations of goodness of fit [$\chi^2(14, 297) = 29.80$, $p < .01$; *GFI* = .98; *AGFI* = .92]. Correlations between measured personality domains and latent traits ranged from .71 to .98 with slightly higher mean values for the NEO-PI-R than for the EPIP-NEO (.93 and .85, respectively). The correlation between two methods (i.e. two instruments) was $r = .87$ ($p < .001$) indicating that the two instruments are measuring very similar constructs.

Analysis of validity

Age differences. The changes in the mean levels of the five personality factors across the life span are well known and replicated in different cultures (McCrae, Costa, De Lima, et al., 1999). Similarly to previous results, the N, E, and O domain scales correlated negatively (r = -.15, -.38, and -.19, respectively, $p < .01$) and the A and C scales positively (r = .28 and .16, respectively, $p < .01$) with age. Similar pattern of correlations appeared between age and the NEO-PI-R domains.

Gender differences. Correspondingly to previous results (Costa, Terracciano, & McCrae, 2001), women scored significantly higher on the N domain , $t(295) = 3.18$, $p < .01$, and the A domain of the EPIP-NEO, $t(295) = 4.50$, $p < .001$. With the NEO-PI-R these differences were rather similar. Neither of the measures reported significant gender differences on E, though in both cases women scored significantly higher in E1 (Warmth). Women had higher mean score on O for both, the EPIP-NEO, $t(295) = 3.23$, $p < .01$, and the NEO-PI-R, $t(295) = 2.44$, $p < .05$ with the gender differences most salient in O2 (Artistic Interests/Aesthetics) and O3 (Emotionality/Feelings). The higher mean score of men in O5 (Intellect/Ideas) did not reach statistical significance in neither of the measures. Curiously, women had significantly higher mean value on the EPIP-NEO C domain, $t(295) = 3.01$, $p < .01$. Though the gender difference on C did not reach the level of significance in case of the NEO-PI-R, there also appeared significant differences in the second and the third facet scale of C domain. Thus, except for C domain the gender differences in the EPIP-NEO scores generally correspond to the previous (Costa, Terracciano, & McCrae, 2001) and present findings with the NEO-PI-R.

Behavioral criteria. The comparison of the mean values of the EPIP-NEO domains and behavioral criterion of depression indicated that subjects who reported

taking medication against depression had significantly higher scores on the N scale, $t(237) = 4.40$, $p < .001$, and lower scores on the E scale, $t(237) = -2.70$, $p < .01$ compared to those who did not confirm it. Other three domains were not significantly related to taking medication. This result is perfectly in line with previous findings that clinical depression is primarily related to Negative Affectivity or Neuroticism and Positive Affectivity or Extraversion (see Clark, Watson, & Mineka, 1994; Terracciano, McCrae, & Costa, 2003). Nearly identical relationships were found for the N and E domain scales of the NEO-PI-R. Further, extraverted people are usually considered to be socially active, seeing frequently their friends and going to parties or events. According to this description it is obvious that individuals with higher values on the Extraversion scale should report the relevant behavior taking place more frequently. As expected, the indicator of social activity was significantly related to the E scale of the EPIP-NEO ($r = .51$, $p < .001$). There also appeared a moderate relationship with the O scale ($r = .31$, $p < .001$), which is predictable considering the relationship between personality domains E and O (in the current case $r = .50$, $p < .001$). Social activity was not significantly related to the other three EPIP-NEO domains. The relationships between social activity indicator and original NEO-PI-R domains were rather similar.

Discussion

Although it might be argued that shorter items of personality measures might be less specific and therefore bring along the loss of necessary exactness in measurement of clearly defined traits, the current study provides a clear demonstration that scales with short and simple items can be reliable and valid.

The results of this study confirmed that the EPIP-NEO as a linguistically minimalist Big Five personality inventory had psychometrical properties comparable to those of the well established and linguistically more sophisticated NEO-PI-R. In this respect, the EPIP-NEO can be compared with a short and more readable version of the MMPI, which also demonstrated comparable psychometrical properties with its original version (Ward & Selby, 1980). Concerning internal consistency, the EPIP-NEO was even superior to the original NEO-PI-R. An important result was that scales with shorter items (i.e. containing smaller number of letter, syllables, words and commas), tended to have higher internal reliability. This is consistent with previous conclusions that items which were rated as more comprehensible were in fact more reliable (Angleitner, John, & Löhr, 1986). The structure of the EPIP-NEO replicated the original five-factor structure of the

NEO-PI-R (Costa & McCrae, 1992) with only minimal differences and generally not worse than NEO-PI-R itself. The cross-instrument agreement between the EPIP-NEO and the NEO-PI-R was excellent at the domain level and with few exceptions at the level of facet scales. Finally, the analysis of external validity lent support to the convergent and discriminant validity of the linguistically simple EPIP-NEO as the instrument generally replicated the established theoretical and empirical relationships with different external variables. Furthermore, the differential validity of the EPIP-NEO was comparable to that of the NEO-PI-R.

Acknowledgements

Authors are indebted to Liina Haring and Airis Erme for their efforts with data collection and to Kenn Konstabel for his extremely helpful comments throughout the study. Authors also thank Anu Realo, Pirjo Saadla and John. A. Johnson for their help with item translation. This project was supported by the Estonian Science Foundation Grant 5677 and an Estonian Ministry of Education and Science Grant 0180543 provided for the third author.

References

Allik, J., Laidra, K., Realo, A., & Pullmann, H. (2004). Personality development from 12 to 18 years of age: changes in mean levels and structure of traits. *European Journal of Personality, 18,* 445-462.

Allik, J., & McCrae, R.R. (2004). Escapable conclusions: Toomela (2003) and the universality of trait structure. *Journal of Personality and Social Psychology, 87,* 261-5.

Angleitner, A., John, O.P., & Löhr, F.-J. (1986). It's *what* you ask and *how* you ask it: an itemmetric analysis of personality questionnaires. In A. Angleitner & J. S. Wiggins (Eds.), *Personality Assessment via Questionnaires.* (pp. 61-108). Berlin: Springer-Verlag.

Austin, E.J., Deary, I.J., & Gibson, G.J. (1997). Relationships between ability and personality: three hypotheses tested. *Intelligence, 25,* 49-70.

Clark, L.A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103,* 103-116.

Costa, P.T.Jr., & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEOFFI) – Professional Manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Costa, P.T., & McCrae, R.R. (1997). Stability and change in personality assessment: the Revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment, 68,* 86-94.

Costa, P.T., Terracciano, A., & McCrae, R.R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of Personality and Social Psychology, 81,* 322-331.

De Fruyt, F., Mervielde, I., Hoekstra, H.A., & Rolland, J.P. (2000). Assessing adolescents' personality with the NEO-PI-R. *Assessment, 7,* 329-345.

Goldberg, L.R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe.* (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Johnson, J.A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality, 39,* 103-129.

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.

Krakauer, S.Y., Archer, R.P., & Gordon, R.A. (1993). The development of the Items-Easy (Ie) and Items-Difficult (Id) subscales for the MMPI-a. *Journal of Personality Assessment, 60,* 561-571.

Markey, P.M., Markey, C.N., Tinsley, B.J., & Ericksen, A.J. (2002). A preliminary validation of preadolescents' self-reports using the Five-Factor Model of personality. *Journal of Research in Personality, 36,* 173-181.

McCrae, R.R. (1990). Traits and trait names - how well is openness represented in natural languages. *European Journal of Personality, 4,* 119-129.

McCrae, R.R., & Costa, P.T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences, 36,* 587-596.

McCrae, R.R., Costa, P.T., De Lima, M.P., Simoes, A., Ostendorf, F., Angleitner, A., Marusic, I., Bratko, D., Caprara, G.V., Barbaranelli, C., Chae, J.H., & Piedmont, R.L. (1999). Age differences in personality across the adult life span: parallels in five cultures. *Developmental Psychology, 35,* 466-477.

McCrae, R.R., Costa, P.T., Terracciano, A., Parker, W.D., Mills, C.J., De Fruyt, F., & Mervielde, I. (2002). Personality trait development from age 12 to age 18: longitudinal, cross-sectional, and cross-cultural analyses. *Journal of Personality and Social Psychology, 83,* 1456-1468.

McCrae, R.R., Herbst, J.H. & Costa, P.T. Jr. (2001). Effects of acquiescence on personality factor structures. In R. Riemann, F.M. Spinath, & F. Ostendorf (Eds.), *Personality and temperament: Genetics, evolution, and structure* (pp.217-231). Berlin: Pabst SciencePublishers.

Paolo, A.M., Ryan, J.J., & Smith, A.J. (1991). Reading difficulty of MMPI-2 subscales. *Journal of Clinical Psychology, 47,* 529-532.

Parker, W.D., & Stumpf, H. (1998). A validation of the Five-Factor Model of personality in academically talented youth across observers and instruments. *Personality and Individual Differences, 25,* 1005-1025.

Sherry, A., Henson, R.K., & Lewis, J.G. (2003). Evaluating the appropriateness of college-age norms for use with adolescents on the neo personality inventory-revised. *Assessment, 10,* 71-78.

Stricker, L.J. (1963). Acquiescence and social desirability response styles, item characteristics, and conformity. *Psychological Reports, 12,* 319-341.

Terracciano, A., McCrae, R.R., & Costa, P.T. (2003). Factorial and construct validity of the Italian Positive and Negative Affect Schedule (PANAS). *European Journal of Psychological Assessment, 19,* 131-141.

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,* 723-35.

Ward, L.C., & Selby, R.B. (1980). Abbreviation of the MMPI with increased comprehensibility and readability. *Journal of Clinical Psychology, 36,* 180-186.

Zwick, W.R., & Velicer, W.F. (1986). Comparison of 5 rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432-442.

Running head: PERSONALITY AND COGNITIVE ABILITY

How Does Cognitive Ability Relate to the Structure, Variability and Measurement of Personality: A Study Using Self and Judge Ratings

René Mõttus, Jüri Allik, and Helle Pullmann

Department of Psychology, University of Tartu, Estonia

The Estonian Centre of Behavioural and Health Sciences

Keywords: personality differentiation; cross-observer agreement; cognitive ability and personality; readable personality inventory

Corresponding Author:

René Mõttus
Department of Psychology
University of Tartu
Tiigi 78, Tartu 50410
Estonia
Phone: +372 375 902
Fax: +372 376 152
E-mail: Rene.Mottus@ut.ee

Abstract

To test the hypothesis that personality structure or interindividual variability of test scores differs across levels of cognitive ability, 154 participants with various age and educational background were divided into two groups on the basis of their ability test scores and their personality was rated by themselves and two well-informed judges using the Estonian Personality Item Pool NEO (EPIP-NEO; Mõttus, Allik, & Pullmann, submitted). Relatively high cross-observer agreement showed that traits were rated accurately in both ability groups. Although in high-ability group traits were slightly less orthogonal and factor structures were somewhat more similar to the normative American self-report structure, it was not possible to conclude that personality structure would substantially differ across ability groups. Variability of scale scores tended to be higher in high-ability group but this might have been related to the differences in data quality.

Introduction

The relationship between personality and cognitive ability has been conceived in two opposite ways. Some researchers (e.g. Cattell, 1957) hold that intelligence is so closely intertwined with personality that it must be considered an inseparable part of personality dispositions. Indeed, it seems intuitively reasonable to expect individuals higher on Openness to receive more information and those higher on Conscientiousness to be more consistent in their studies, which in turn could result in higher scores in intelligence tests. However, the findings are contradictory, since most studies report weak and frequently non-significant correlation between ability and personality constructs (for review see Ackerman & Heggestad, 1997). Within the Five-Factor Model of Personality (FFM), Openness to Experience is the only dimension that systematically tends to correlate with cognitive ability, yet these two are considered to form separate dimensions (McCrae & Costa, 1985). On the basis of low and nonsystematic association, Eysenck (1994) has promoted the conclusion that personality dispositions and intelligence are mutually independent dimensions of individual differences.

Nevertheless, the possible interaction of personality and cognitive ability or other aspects of cognition has gained continuous interest. It is suggested, for example, that individuals with different levels of ability might use their intellectual resources differently to express their individuality (Allik & Realo, 1997; Allik, Laidra, Realo, & Pullmann, 2004). One line of research for studying more complex interactions of personality and ability has concentrated on the possible relation of ability to the structure of personality. An interesting idea came from Brand and his colleagues who proposed the personality differentiation hypothesis, which holds that individuals with higher cognitive ability have more differentiated personalities (Brand, Egan & Deary, 1994). According to the hypothesis, people with higher ability have more choices or freedom in development and this can lead to a more differentiated structure of personality (Austin, Deary, & Gibson, 1997; Austin, Hofer, Deary, & Eber, 2000). As a result, the more intellectually talented might need a greater number of dimensions to describe their personalities or their responses might be more variable. However, the support for the differentiation hypothesis has been modest at best. Contrary to expectations, Austin and her colleagues (1997) did not observe higher intercorrelations between personality traits in the group of lower ability. Nonetheless, in a more recent study Austin, Deary, Whiteman with their colleagues (2002) found that the correlation between Eysenck's Psychoticism and

Neuroticism scales decreased with the higher levels of ability. A similar tendency for the decrease of correlation between personality dimensions with increase of cognitive abilities or age has been noticed by other researchers (Allik et al., 2004).

In addition to intercorrelations between traits, researchers have noticed that different ability groups differ in the interindividual variation. Austin and her colleagues (1997) found that more able individuals had higher standard deviations in the NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992) scores, with significant differences on Neuroticism and Openness. Similarly, Harris, Vernon and Jang (2005) observed higher variance in 15 of the 20 scales (three of 15 differences were significant) of the Personality Research Form (PRF; Jackson, 1989) in the group of individuals with higher mental ability. In other words, there are modest indications that individuals with lower intelligence are more similar to one another compared to individuals with higher intelligence. However, it is important to notice that the increase of interindividual variation does not necessarily lead to the decrease of correlation between personality dimensions. These two problems – differentiation of personality dimensions and interindividual variation – are logically two separate problems.

There have been several other attempts to demonstrate that self-reported personality structure depends on the cognitive ability of the evaluator. For example, Shure and Rogers (1963) found that structure of personality traits was slightly different at different levels of ability. More recently, Toomela (2003) claimed that the structure of personality is related to the Vygotskian developmental stages in the word meaning. His assertion was based on the idea that people using primarily more simply organized everyday concepts encode information about the self and world in a different way than people with primarily hierarchical scientific concepts. Toomela (2003) did not make any explicit assumptions about the direction of the effect of word meaning structure for personality structure (e.g. the number and nature of the factors needed for its description or their variance). Rather, he claimed that the robust five-factor structure is not present in “everyday thinkers”. He proposed that two opposite tendencies might be involved. First, individuals with primarily scientific concepts possess mental tools that allow organizing their perception more efficiently and therefore fewer components would be necessary for the description of their personalities. On the other hand, “scientific thinkers” have more differentiated understanding of the world and this might lead more complex personalities with higher number of dimensions. As a result of these opposite tendencies, the content, but not necessarily the number of personality dimensions may vary with different levels

of cognitive ability. In his study, Toomela (2003) demonstrated that typical five-factor structure of the Revised NEO Personality Inventory (NEO-PI-R; Costa & McCrae, 1992) was poorly replicated in the group of subjects using the most everyday concepts.

Importantly, the structure of word meaning is theoretically related to cognitive ability. Word meaning structure provides a set of mental tools for cognitive operations and cognitive ability "can be understood as one's level of mastery of those tools" (Toomela, 2003: p. 725). The correlation between the structure of word meaning and cognitive ability measured by conventional intelligence test was reported as high as r = .54, p < .0001 (Toomela, 2003). Consistently, the typical five-factor structure of the NEO-PI-R was less clear in groups of lower cognitive ability.

However, Allik and McCrae (2004) reanalyzed Toomela's data and reported that when using more appropriate targeted rotation, the initially non-typical factor matrixes of the NEO-PI-R resembled much more the "normative" structure (Costa & McCrae, 1992). Nevertheless, the fit was not perfect in the groups with the most extreme everyday concept meaning structure and with the lowest cognitive ability. Allik and McCrae (2004) suggested that the deviations might be due to measurement error in the self-reports, instead of real variations in the structure of personality, which can be estimated among other ways by ratings of external observers.

The self-report data are inconclusive because they cannot tell the difference between enduring dispositions to think, feel, and behave in a particular consistent manner and the individual's ability to estimate and report about these dispositions. It is possible, for example, that individuals with insufficient cognitive ability have difficulties with understanding items of personality questionnaire or they might not be able to make reliable inferences about their own personality dispositions (Austin et al., 1997; Allik & McCrae, 2004). This, however, does not prevent external observers, who know the target well, making accurate judgments about his or her personality. Although the description of one's own personality is not a very complicated task, it still requires a minimal amount of ability (Allik et al., 2004). For example, Austin and her colleagues (1997) noticed that in the groups with lower ability the personality scales had lower internal reliabilities. Such problems can obviously bring along the loss of validity in the personality judgments and, in turn, lead to changes in the reported personality structure. The confounding effects of poor comprehension or judgments do not necessarily have to hold for all dispositions, since in Toomela's study (2003) the Neuroticism factor was clearly replicated in all ability groups.

These methodological problems can be overcome using alternative sources of personality data. A promising solution is to use aggregated ratings of knowledgeable other people (further named *judges*). It has been demonstrated that the consensus between well-informed judges is the most valid source of personality ratings (Kolar, Funder, & Colvin, 1996). Therefore, if the differences in personality structure at different levels of cognitive ability indeed exist, they would likely be reflected in the ratings of knowledgeable judges. In addition, cross-observer agreement provides a reasonable way to establish whether personality traits at all exist and are valid characteristics of individuals (McCrae, 1982; McCrae, Costa, Martin et al., 2004; Woodruffe, 1985). If ratings on traits obtained from different sources converge, it is more likely that the traits are not incidental entities. Of course, the criterion of cross-observer agreement is not sufficient as many people can perfectly agree on things that are demonstrably wrong (Funder, 1995). However, the agreement is one of the few important and available criterions of accuracy of the judgments on personality (Funder, 1995; Funder & Colvin, 1988) and therefore it is a necessary precondition of establishing the existence of traits. Consequently, the consensus between personality ratings of self and judges or between at least two judges can be used to study whether or not certain personality dispositions are judged accurately – and therefore probably exist – in different populations, including groups with various level of cognitive ability. In other words, if a personality trait is not characteristic of a person at certain level of ability, the judges are less likely to agree on that.

This study

The principal aim of this study was to establish whether the ability-related differences in personality structure and interindividual variability of personality test scores can be replicated and if these differences exist in self-reports whether they will also be observable in the ratings of external observers. So far only self-reported personality has been studied (Austin et al., 1997, 2000, 2002; Toomela, 2003; Harris et al., 2005) and it is possible that the observed differences in different ability groups characterize not the relationship between personality traits as such but the individual's limited ability to describe these traits. Moreover, in order to diminish the possibility that the changes in personality might be caused by difficulties with understanding the items of questionnaire, a more readable version of the Big Five personality measure was employed. In brief, alternative sources of ratings and simplified measure served the

purpose of separating personality traits from possible errors in the description and reporting these traits.

Another aim of this study was to examine how the quality of personality data is related to cognitive ability of the rater or the person being rated. Mõttus, Allik, and Pullmann (submitted) proposed that internal reliability might be related to linguistic complexity of measures (e.g. average number of words, syllables, letters or commas per items). It is reasonable to hypothesize that this relationship is stronger in individuals with lower level of ability since they are probably more likely to encounter difficulties with understanding sophisticated items. Therefore it is possible that linguistic complexity of personality measures moderates the relationship of ability and quality of ratings.

Method

Sample

The sample consisted of 154 participants (53 men and 101 women) with the mean age 43.89 years (*SD* = 17.88) years ranging from 16 to 83. About 8% of subjects had age below 21 years, 40% had age between 21 and 40 years, 29% had 41 to 60 years of age and 23% had age over 60. Participants had various educational backgrounds as approximately 3% had lower than elementary, 15% elementary, 33% vocational, 26 % secondary and 17 % higher education. Six percent did not report their education. The sample was intended to be as closely representative of Estonian population as possible (in relation to age, sex and educational level), although finally the proportion of women was much higher. The participants were volunteers who were reached through personal contacts of collaborators.

In addition to self-reported personality ratings, each participant (target) was rated by two well-informed judges. The sample of judges (*n* = 308) included 203 women and 67 men, 38 participants did not report their gender. The mean age of judges was 38.15 (*SD* = 15.94) years ranging from 16 to 81. On average, the judges were about five years younger and more highly educated than the targets. In general, about 52% of judges were close relatives or partners, 25% were friends, 12% colleagues. For 11% of judges the exact relation to the targets fell into the category "Other" (which included more distant relatives, neighbors etc.). About 93% of judges knew their targets more than two years, with 61% knowing him or her more than 10 years and only 1% less than a year.

Measures

Personality. The EPIP-NEO (Mõttus et al., submitted) measures Five personality domains – Neuroticism (N), Extraversion (E), Openness (O), Agreeableness (A) and Conscientiousness (C) – described by six facet scales each. The structure of the EPIP-NEO is analogous to the NEO-PI-R (Costa & McCrae, 1992) but the EPIP-NEO is designed to be linguistically simpler, containing shorter and grammatically less complex items. The scales of the EPIP-NEO demonstrated on average higher internal reliabilities than parallel scales of the NEO-PI-R (Mõttus et al., submitted). The average convergent correlation between the corresponding facet scales of the EPIP-NEO and Estonian version of the NEO-PI-R (Kallasmaa, Allik, Realo, & McCrae, 2000) was *r* = .73. The EPIP-NEO

replicates typical five-factor structure of its analogue the NEO-PI-R well since after targeted rotation to the North-American structure of the NEO-PI-R (Costa & McCrae, 1992) the average congruence coefficient was .95.

Cognitive ability. The Cognitive Ability Test (CAT) that consisted of three subscales was constructed to measure general mental ability (*g*). All three subscales – Vocabulary, Similarities and General Information – are considered good measures of *g* (Kaufman & Lichtenberger, 1999). The items in subscales were selected on the basis of a pilot study. First, the Vocabulary subscale included 58 words which respondents were asked to define (possible scores were 1, 2 and 3). The Vocabulary subscale had internal consistency of Cronbach α = .94 and inter-item correlation of r_{ii} = .23. Second, the Similarities subscale included 34 pairs of words (possible scores were 1, 2 and 3). Respondents had to indicate in what way two words were similar. The subscale had internal consistency of α = .87 and inter-item correlation of r_{ii} = .23. Third, the General Information subscale consisted of 47 questions tapping person's level of knowledge in various domains (possible scores were 1 and 2). The subscale had internal consistency of α = .93 and inter-item correlation of r_{ii} = .23. The correlations between subscales were r = .79 (p < .001), r = .77 (p < .001) and r = .74 (p < .001), respectively for Vocabulary–Similarities, Vocabulary–General Information and Similarities–General Information. The CAT was administered individually under a supervision of specially trained psychologists.

Ability groups

The observed total scores on the CAT were distributed normally (Kolmogorov−Smirnov d = .06, p > .20), ranging from 31 to 203 with the mean of 133.48 (*SD* = 35.94). Based on the participant scores on the CAT the sample was divided according to the median split into two ability groups:

(1) The high-ability group consisted of 78 subjects, 24 men and 54 women [mean age 41.10 (*SD* = 15.96), years ranging from 16 – 79]. About 8% of subjects had age below 21 years, 44% had age between 21 and 40 years, 33% had 41 to 60 years of age and 15% had age over 60. Approximately 9% had elementary, 28% vocational, 28 % secondary and 32 % higher education. About 3% did not report their education. The mean score of the CAT was 161.88 (*SD* = 20.62).

(2) The low-ability group consisted of 76 subjects, 29 men and 47 women [mean age 46.75 (*SD* = 19.35), years ranging from 17 – 83]. About 8% of subjects had age below

21 years, 37% had age between 21 and 40 years, 24% had 41 to 60 years of age and 32% had age over 60. Approximately 5% had lower than elementary, 21% elementary, 39% vocational, 24 % secondary and 1 % higher education. About 9% did not report their education. The mean score was 104.33 (*SD* = 22.31).

The difference between the mean scores of two ability groups was 1.60 standard deviations, which is reasonably large to consider the two groups different with respect to their ability level. An analysis of variance (ANOVA) revealed significant main effect due to level of education, $F(4, 140) = 19.50$, $p < .0001$ that can be taken as evidence of the external validity of the CAT. Importantly, cognitive ability was related to age ($r = -.30$, $p < .001$) which points to the necessity to be henceforth aware of the possible confounding effects of age.

For both ability groups the judges were younger and more highly educated.

Results

Ability and the quality of personality judgments

It would be expected that the low-ability group is less consistent in their self-description than the high-ability group. Although there were no remarkable differences in the internal consistency values for both ability groups along the domains (see Table 1), the group of higher ability demonstrated in nearly all cases higher alpha values (average value α = .80) than the low-ability group (average value α = .73) across the facets. On average, the judge ratings had slightly higher internal consistencies than ratings of the targets. Interestingly, the relationship between ability and internal consistency was noticed in judge ratings as well, since the groups of others who rated the less able targets similarly showed lower internal consistencies (average α = .84 and .77, for high and low-ability group, respectively).

The relationship between internal consistencies of the 30 facet scales and indicators of linguistic complexity of the items were systematically and significantly related only in the group of lower ability. For example, concerning self-report data of low-ability targets the average number of letters per scale correlated significantly with the Cronbach alphas as high as Spearman R = -.59, p < .001. In the high-ability group the correlation was insignificant (R = -.24, p = .20). The difference in coefficient alpha values between ability groups (alphas of high-ability group minus alphas of low-ability group) was related to the number of letters R = .62, p < .001, indicating that the linguistically more complex scales tended to have larger drop in internal consistencies. There was no systematic relationship between coefficient alphas and indicators of linguistic complexity in judge ratings.

Another indication that cognitive ability may influence responding to personality questionnaire items was a negative correlation between ability and acquiescent responding – the sum of all items before inversion of negatively keyed items was performed – in self-reports (r = -.47, p < .001).

Table 1. *Cronbach alphas of the scales of the EPIP-NEO in the groups of higher and lower cognitive ability.*

	High-ability group[1]		Low-ability group[2]	
	Self-report	*Judge ratings*	*Self-report*	*Judge ratings*
Neuroticism	.95	.95	.94	.93
Extraversion	.92	.94	.92	.90
Openness	.90	.90	.86	.87
Agreeableness	.89	.94	.88	.94
Conscientiousness	.91	.95	.93	.96
N1 Anxiety	.87	.88	.80	.74
N2 Anger	.91	.92	.89	.89
N3 Depression	.87	.87	.84	.85
N4 Self-Consciousness	.77	.78	.71	.70
N5 Immoderation	.79	.82	.75	.78
N6 Vulnerability	.84	.80	.80	.82
E1 Friendliness	.82	.87	.78	.86
E2 Gregariousness	.84	.85	.81	.77
E3 Assertiveness	.77	.80	.76	.70
E4 Activity Level	.72	.82	.66	.65
E5 Excitement-Seeking	.81	.87	.72	.73
E6 Cheerfulness	.88	.91	.85	.87
O1 Imagination	.86	.84	.80	.76
O2 Artistic Interests	.86	.85	.68	.82
O3 Emotionality	.86	.81	.69	.77
O4 Adventurousness	.79	.77	.68	.68
O5 Intellect	.84	.80	.66	.75
O6 Liberalism	.64	.70	.38	.26
A1 Trust	.85	.87	.79	.82
A2 Morality	.80	.86	.65	.82
A3 Altruism	.79	.88	.65	.83
A4 Cooperation	.53	.77	.66	.73
A5 Modesty	.81	.84	.72	.76
A6 Sympathy	.77	.81	.68	.82
C1 Self-Efficacy	.79	.79	.70	.79
C2 Orderliness	.89	.89	.81	.84
C3 Dutifulness	.74	.88	.82	.89
C4 Achievement Striving	.75	.79	.69	.82
C5 Self-Discipline	.83	.88	.84	.86
C6 Cautiousness	.77	.84	.76	.83

NOTE: [1]n = 78, [2]n = 76.

Cross-observer agreement

To examine how well different raters agreed on personality traits, the correlations were calculated between the corresponding scores of self-reports and judge ratings (see Table 2). Generally, the cross-observer agreement was relatively high in both ability groups, being comparable to findings reported by McCrae and his colleagues (2004). At the domain level there were no remarkable differences in agreement between ability groups, neither in ‘self-judge’ or ‘judge-judge’ agreement. At the facet level the average ‘self-judge’ correlations were r_{μ} = .44 and .43, respectively for high and low-ability groups (in 18 facet scales the correlations were higher in high-ability group). Average 'judge-judge' correlations were r_{μ} = .41 and .40, respectively for the groups of higher and lower ability (in 17 facet scales the correlations were higher in high-ability group). Thus, on average the agreement was nearly equal in both ability groups. A closer examination of the correlations revealed that the most significant decrease of self-other agreement in low-ability group was in the N4 (*Self-Consciousness*) and the O6 (*Liberalism*). Concerning the O6, the difference was partially related to the drop of internal reliability in low-ability group but this was not the case in the N4. Thus, the N4 was the only scale of the EPIP-NEO in which ‘self-judge’ agreement was significantly lower in low-ability group than in high-ability group. For this facet scale 'judge-judge' agreement was also significantly lower in case of lower-ability targets. On the other hand, it is worth pointing out that in two facet scales – the N3 (*Depression*) and the C3 (*Dutifulness*) – 'self-other' agreement was significantly higher in the low-ability group.

An alternative way to study whether the 'self-judge' or 'judge-judge' agreement is related to target’s level of cognitive ability, was to compare self and judge reported profiles of the scores of the 30 facet scales. The correlation (Spearman *R*) between these profiles was taken as an index of the agreement. Average 'self-judge' profile agreement was R = .65 (values ranged from -.23 to .98) and average 'judge-judge' profile agreement was R = .60 (values ranged from -.20 to .97). When the agreement between profiles of self and averaged ratings of two judges were compared with ability scores the correlation was insignificant (r = .14, p = .80). Neither was the agreement of the profiles of two judges significantly correlated with ability (r = .01, p = .94). Hence, in general these results showed that cross-observer agreement was relatively independent of target’s cognitive ability.

Table 2. *Cross-observer agreement in the groups of higher and lower cognitive ability*.

	Self-Judge Correlations		Judge-Judge Correlations	
	High-Ability[1]	*Low-Ability*[2]	*High-Ability*[1]	*Low-Ability*[2]
Neuroticism	.46***	.49***	.46***	.43***
Extraversion	.62***	.64***	.59***	.60***
Openness	.55***	.50***	.50***	.47***
Agreeableness	.44***	.42***	.39***	.38***
Conscientiousness	.37***	.41***	.48***	.51***
N1 Anxiety	.53***	.42***	.45***	.43***
N2 Anger	.37***	.46***	.40***	.30***
N3 Depression	.44***	.61***[c]	.49***	.56***
N4 Self-Consciousness	.45***	.24**[c]	.52***	.12[a]
N5 Immoderation	.35***	.44***	.39***	.47***
N6 Vulnerability	.42***	.40***	.36***	.37***
E1 Friendliness	.45***	.54***	.38***	.39***
E2 Gregariousness	.49***	.43***	.49***	.47***
E3 Assertiveness	.52***	.50***	.45***	.47***
E4 Activity Level	.45***	.37***	.44***	.39***
E5 Excitement-Seeking	.54***	.57***	.44***	.41***
E6 Cheerfulness	.58***	.56***	.56***	.54***
O1 Imagination	.45***	.34***	.32***	.28***
O2 Artistic Interests	.47***	.55***	.54***	.51***
O3 Emotionality	.42***	.37***	.35***	.28***
O4 Adventurousness	.48***	.39***	.43***	.45***
O5 Intellect	.48***	.34***	.37***	.33***
O6 Liberalism	.53***	.32***[c]	.57***	.32***[b]
A1 Trust	.32***	.38***	.26**	.43***
A2 Morality	.41***	.31***	.47***	.42***
A3 Altruism	.37***	.32***	.34***	.32***
A4 Cooperation	.34***	.40***	.20*	.30***
A5 Modesty	.54***	.36***	.57***	.42***
A6 Sympathy	.29***	.37***	.33***	.34***
C1 Self-Efficacy	.30***	.32***	.19*	.23**
C2 Orderliness	.52***	.35***	.59***	.39***[c]
C3 Dutifulness	.34***	.55***[c]	.48***	.56***
C4 Achievement Striving	.44***	.32***	.44***	.44***
C5 Self-Discipline	.42***	.41***	.53***	.49***
C6 Cautiousness	.35***	.49***	.45***	.50***

NOTE: [1]n = 78, [2]n = 76.

*** $p < .001$, ** $p < .01$, * $p < .05$.

[a] difference between ability groups significant at $p < .001$

[b] difference between ability groups significant at $p < .01$

[c] difference between ability groups significant at $p < .05$

Correlations between ability and personality domains

To explore whether personality domains are related to cognitive ability and whether these relationships are in turn affected by general ability level, the correlations between ability and personality domains were examined in the whole sample and separately in the two ability groups. In the whole sample E, O, A and C domains were significantly related to cognitive ability in case of the self-reports [r = .24 (p < .01), .42 (p < .001), -.20 (p < .05) and -.16 (p < .05), respectively]. In the averaged scores of two judges N, E and O domains were significantly related to ability [r = -.22 (p < .01), .24 (p < .01) and .47 (p < .001), respectively]. However, when age was taken into account, only self-reported O correlated significantly with ability (r = .36, p < .001) but in averaged scores of two judges the correlation were less affected by target's age [r = -.22 (p < .01), .18 (p < .05) and .41 (p < .001), respectively for N, E and O]. In the high-ability group, ability correlated significantly with E and O domains in self-reports (r = .33 and .37, p < .01, respectively) and with O in judge ratings (r = .36, p < .01). When the age was taken into account, only the correlations with O remained statistically significant [were r = .30 (p < .01) and .25 (p < .05), for self and judge ratings, respectively]. In the low-ability group, ability correlated significantly with O domain in both self and judge ratings (r = .26 and .29, p < .05, respectively). When the age was taken into account, only in judge ratings the correlation remained significant (r = .25, p < .05). Thus, in the whole sample and both ability groups only O domain was somewhat more systematically related to cognitive ability, whereas relationship of ability with other domains was minor. This conclusion is in agreement with previous results (e.g. Costa & McCrae, 1992). Consequently, except for Openness, individual's score on personality dimensions was relatively independent of his or her level of cognitive ability.

The relationship between ability and the structure of personality

A simple way to study the effect of ability on personality structure is to calculate the ten possible correlations between five personality domains and compare them across ability groups (see Table 3). It appears that generally the intercorrelations were higher in the low-ability group with the tendency more remarkable in the averaged ratings of judges. Yet most of the differences between ability groups were not significant. In self-reports the correlations of N domain with A and C domains were significantly higher in low-ability group. In averaged ratings of judges, additionally, the correlations of O

domain with A and C domains were significantly higher for lower-ability participants. Interestingly, the depressing effect of maturation and gains in cognitive ability on the correlation between A and C domains noticed by Allik and his colleagues (2004) was not clearly confirmed by the current results. Neither did the results support the proposal of Brand and his colleagues (1994) that the relationship between E and C domains decreases with level of cognitive ability. There was no good reason to expect any mediating effects of age on the intercorrelations and, indeed, taking age into account changed the correlations only very slightly. Also, the correlations did not change when acquiescent responding was taken into account. Thus, generally the results lend modest support to the hypothesis that among less able individuals the Big Five dimensions are less orthogonal.

Table 3. *Intercorrelations of five personality domains.*

	N		E		O		A	
	HA	*LA*	*HA*	*LA*	*HA*	*LA*	*HA*	*LA*
E	-.28*	-.34**						
	-.44***	-.37**						
O	.07	-.12	.45***	.55***				
	-.06	-.29*	.45***	.47***				
A	-.03	-.31**[c]	-.24*	-.05	.09	.06		
	-.12	-.42***[b]	-.16	.04	.09	.38**[c]		
C	-.36**	-.65***[b]	.03	.17	-.08	.11	.39***	.47***
	-.28*	-.63***[a]	-.09	.04	-.18	.30**[a]	.47***	.64***

NOTE: *HA* = high-ability group (*n* = 78); *LA* = low-ability group (*n* = 76); *N* = Neuroticism; *E* = Extraversion; *O* = Openness; *A* = Agreeableness; *C* = Conscientiousness. The correlations in self-reports are given in upper rows. The correlations in averaged ratings of judges are given in bottom rows.
*** $p < .001$, ** $p < .01$, * $p < .05$.
a = difference between correlations in high and low-ability groups significant at $p < .01$ (two-tailed)
b = difference between correlations in high and low-ability groups significant at $p < .05$ (two-tailed)
c = difference between correlations in high and lo-ability groups significant at $p < .05$ (one-tailed)

A reasonable way to estimate the effect of ability on the interrelations of the lower level personality traits (i.e. the facet scales of the EPIP-NEO) was to compare the eigenvalues of the correlation matrixes (expressed as the variance accounted for by the first unrotated principal component) across ability groups. In self-reports the first unrotated principal component (1PC) in the Principal Component Analysis (PCA) accounted for 22.0% and 26.7% of the total variance, respectively in the groups of higher and lower ability. In the averaged ratings of two judges the percentages of accounted variance were 24.7% and 33.3% for high and low ability groups, respectively. In order to test whether the percentages of accounted variance differed significantly across ability groups a series of PCA was performed so that the sample was 1000 times randomly divided into the two subgroups of the size of ability groups. The percentages of explained variance by the 1PC in two subgroups were then compared. In self-reports the difference at least 4.73% (i.e. 26.71 – 21.98%) was observed in 113 cases. In averaged ratings of judges the difference at least 8.66% (i.e. 33.33 – 24.67%) appeared in only one case. Expressed in probabilities of obtaining the observed difference by chance, in self-report ratings the difference in explained variance was not significant (p = .11) but in the averaged ratings of two judges the difference it was significant (p = .001). When age was taken into account the percentages of explained variance changed only very slightly. Evidently it must be noted that so far both within and between domains intercorrelations of facet scales were indistinguishably involved in analysis and therefore the larger percent of explained variance by 1PC in low-ability group might have occurred due to domains with more tightly interrelated facets. However, the intercorrelations of facet scales within domains showed that only C domain systematically demonstrated more tightly interrelated facets in low-ability group (the mean inter-facet correlations were r_{if} = .37 to .56 in self-reports and r_{if} = .53 to .71 in averaged judge ratings, respectively in high and low ability groups). In other domains the differences were of less remarkable magnitude and, moreover, in some cases the inter-facet correlations were higher in individuals with higher ability. Hence, it is not likely that the somewhat higher intercorrelations of facet scales in low-ability group were only due to internally more consistent domains. Consequently, there are indications that lower level traits are to some extent more tightly intercorrelated in individuals with lower ability.

To test Toomela's (2003) claim that the typical five-factor structure is not present in individuals with lower cognitive ability, factor structures were compared across ability groups. Of course due to its small size the current sample cannot be divided into five

ability groups as Toomela did. In self-reports the first five factors explained 66.5% and 66.9% of the total variance in high and low-ability groups, respectively. In case of the averaged ratings of two judges the respective percentages were 74.4 and 72.1. In self-reports the Varimax-rotated factor structures were far from simple structure in both ability groups and even in the whole sample. The factor congruence coefficients with the North-American "normative" structure of the NEO-PI-R[4] (Costa & McCrae, 1992; further named as normative structure) ranged from .64 to .92, from .77 to .91 and from .77 to .93, in the groups of high and low ability and the whole sample, respectively. In the averaged ratings of two judges the Varimax-structure slightly more resembled simple structure. The congruence coefficients with the normative structure ranged from .85 to .93, from .75 to .88 and from .85 to .93, in the groups of high and low ability and whole sample, respectively. Considering these results, a typical five-factor structure was definitely not clearly replicated. However, it is well known that the NEO-PI-R (and consequently its analogue the EPIP-NEO) does not show simple structure (Church & Burke, 1994). As a result, the Varimax rotation might not be the most appropriate rotation method, especially in small samples were the random perturbation in data can distort the structure (Allik & McCrae, 2004). Therefore McCrae, Zonderman, Costa, Bond, and Paunonen (1996) suggested that exploratory factor analysis should be followed by confirmatory analysis, which uses targeted rotation. According to this method the factor structure under study is matched to the "target" structure so that the congruence is maximal. This strategy is considered more appropriate when assessing the replication of a certain factor structure.

Table 4 shows the factor congruence coefficients of the Procrustes-rotated factor structures of the EPIP-NEO with the normative structure. In self-report ratings the five factors extracted from the data of the whole sample replicated well those of the normative structure since all coefficients exceeded .90. In both ability groups one factor failed to break the .90 level (O factor in high-ability group and E factor in low-ability group). The total congruence in high-ability group was even lower (.89) than in low-ability group (.91). To test the significance of differences in congruence coefficients between ability groups, PCA followed by targeted rotation was repeated 1000 times dividing the sample randomly into subgroups of the size of original ability groups. For all factors and total congruence, the difference of at least the same absolute value as difference between the

[4] The North-American structure of the NEO-PI-R was here considered the most appropriate criterion when estimating the degree of replication of the five-factor structure since this criterion has become widely used and the so far „normative" structure of the EPIP-NEO (Mõttus et al., submitted) was obtained from too small sample to allow solid conclusions. However, when the factor stuctures obtained in this study were still targeted at the structure of Mõttus and his colleagues (submitted) similar trends were observed.

original ability groups appeared in more than half of the analysis (hence $p > .50$). Thus, the congruence with normative structure was not significantly different across ability groups. The factor structures obtained from averaged ratings of two judges replicated well the normative structure in the data of the whole sample and high-ability group, since all congruence coefficients exceeded .90. In the low-ability group the factor congruence coefficients were to some extent lower, ranging from .86 to .89. The difference in congruence was significant in case of O and A domains (dividing the whole sample randomly in subgroups of the same size as ability groups and performing PCA 1000 times gave the difference of at least same magnitude less than 50 times, hence $p < .05$). Total congruence in high-ability group was significantly higher (.92) than in low-ability group (.87). In summary, it can be concluded that both ability groups showed at least satisfactory fit to the normative five-factor structure. Even so, the more able demonstrated somewhat better congruence with the intended structure, although this appeared only in one source of ratings. To rule out a possible confounding effect of age the analysis was re-performed on the data in which the age was taken under control. Nevertheless the similar trend appeared.

Table 4. *Factor congruence coefficients of the EPIP-NEO after targeted rotation at the normative structure of NEO-PI-R* (Costa & McCrae, 1992).

	Self-reports						Averaged ratings of judges					
	N	*E*	*O*	*A*	*C*	*Total*	*N*	*E*	*O**	*A**	*C*	*Total***
High-ability group[1]	.90	.90	.83	.90	.93	.89	.93	.93	.94	.92	.91	.92
Low-ability group[2]	.91	.86	.90	.94	.92	.91	.88	.88	.87	.86	.89	.87
Whole sample[3]	.91	.91	.91	.93	.94	.92	.93	.93	.95	.91	.92	.92

NOTE: [1]n = 78, [2]n = 76, [3]n = 154; *N* = Neuroticism; *E* = Extraversion; *O* = Openness; *A* = Agreeableness; *C* = Conscientiousness.

* Dividing the whole sample randomly in subgroups of the size of ability groups and performing PCA 1000 times gave the difference of at least same magnitude as was the difference across ability groups less than 5% of cases

** Dividing the whole sample randomly in subgroups of the size of ability groups and performing PCA 1000 times gave the difference of at least same magnitude as was the difference across ability groups less than 1% of cases

Interindividual differences in ability groups

In order to test whether individuals with lower ability are more similar to each other than high-ability individuals, variability of the scores of personality scales were

compared across ability groups. However, first a possible problem has to be discussed. It is possible that the smaller standard deviations obtained by individuals with lower ability might be the result of lower reliabilities of scales (Austin et al., 1997). In other words, to obtain large standard deviations one must consistently give more extreme ratings but the low reliabilities may mean that this is not the case. In addition, it was noticed that acquiescent responding was significantly related to ability (r = -.47, p < .001). It is possible that in some cases acquiescent responding might reduce the variance of scores. Thus, the possible confounding effect of lower quality of personality ratings must be kept in mind when comparing ability groups.

In four out of the five personality domains the standard deviations were larger in the group of higher ability (see Table 5). On the fifth dimension, Conscientiousness, the interindividual differences were larger in the low-ability group. This was true in both self-reports and the averaged ratings of knowledgeable others. Interestingly, C also had higher internal reliability and inter-facet correlation in the low-ability group. However, the only significant difference in variances between high and low-ability groups (as revealed by Levene test of homogeneity of variances) was in Extraversion (p < .05) and even this was to true only in case of averaged ratings of two judges. At the facet level in 21 and 24 scales the standard deviations were higher in high-ability (respectively in self-reports and judge ratings; in both cases the difference was significant in six facet scales)[5]. Taking the acquiescent responding into account did alter the standard deviations of domains only very slightly.

Costa, Terracciano, McCrae with their colleagues (2001) noticed that men and women were less similar in relation to their scores on several personality scales in economically more developed countries were individuals (especially women) have better opportunities for education. Whatever is the exact cause of their findings, it seems that gender differentiation is a dimension on which different samples differ. In order to study whether cognitive ability has an effect to the degree men and women are similar in relation to their personality profile mean values of men and women in 30 facet scales of the EPIP-NEO were compared. In high-ability group the correlations between the profiles of mean values of men and women were Spearman R = .79 (p < .001) and .91 (p < .001), respectively in self-reports and averaged ratings of judges. In the group of lower ability the correlations were R = .91 (p < .001) and .95 (p < .001). Hence, according to self-

[5] After Bonferroni correction for multiple-comparisons none of the differences was significant, neither between domains nor facet scales.

reports in the group of lower ability men and women were only marginally more similar in relation to their personality profiles but in case of the averaged ratings of others the difference was negligible. At the domain level no systematic trend appeared when the differences of mean scores of men and women were compared across ability groups. Consequently, the results indicate that ability does not have substantial effect on the similarity of men's and women's personalities.

Table 5. *Means and standard deviations of five domains of the EPIP-NEO in the whole sample and ability groups*.

	Whole sample [1]		High-ability group[2]		Low-ability group[3]			
	μ	σ	μ	σ	μ	σ	*F*	*p*
N	77.95	29.33	74.55	30.48	81.43	27.87	0.33	.57
	73.69	23.69	69.67	25.13	77.82	21.50	2.29	.13
E	104.38	25.33	108.26	25.38	100.41	24.81	0.16	.69
	111.58	24.11	116.35	26.26	106.68	20.74	5.97	.02*
O	115.68	21.50	121.96	22.28	109.24	18.73	1.51	.22
	108.65	19.06	115.47	19.10	101.66	16.38	1.59	.21
A	132.70	19.67	129.22	19.87	136.28	18.93	0.39	.54
	127.20	21.94	125.68	22.51	128.77	21.38	0.01	.92
C	127.99	23.36	125.09	21.94	130.96	24.51	0.68	.41
	133.64	25.09	132.87	24.32	134.44	26.00	0	.95

NOTE: [1]n = 154, [2]n = 78, [3]n = 76; *N* = Neuroticism; *E* = Extraversion; *O* = Openness; *A* = Agreeableness; *C* = Conscientiousness; *F* = *F*-test for differences in variance (Levene Test). The means and standard deviations in self-reports are given in upper rows. The means and standard deviations in averaged ratings of judges are given in bottom rows.
* $p < .05$.

Discussion

All attempts to test the hypotheses that personality structure or interindividual variability of the scores of personality measures is related to cognitive ability have been inconclusive since self-report data have possible limitations. For example, it is possible that individuals with lower cognitive ability have difficulties with properly understanding items of personality measures or they are not able to make reliable judgments about their own personality dispositions (Austin et al. 1997; Allik & McCrae, 2004). As a result, the found differences in personality structure might be caused by measurement errors rather than real variation in the covariation of personality traits. The same might be true for interindividual variability. The present study aimed to overcome these limitations by using ratings of well-informed judges in addition to self-report data and by employing a more readable personality measure.

Generally the present study lent only minimal support to the hypothesis that ability has an effect on the structure of personality. Although there were some significant differences across ability groups they were non-systematic and in general little above chance. According to the present results there is little support for the idea that personality structures are substantially different at different levels of cognitive ability. Sometimes the possible effect of ability on personality structure has also been associated with the issue of interindividual variability on personality test scores (e.g. Austin et al., 1997; Harris et al., 2005). However, it must be kept in mind that these two properties of personality are not necessarily related. Different groups can differ in relation to their variability on personality dimensions but this does not mean that the structure (covariation) of traits has to be different. Consistently with previous result of Austin and her colleagues (1997, 2000) and Harris and her colleagues (2005), the present results showed that intellectually more gifted individuals had larger interindividual variation on more than half of the used personality scales. But as it was argued by Austin and her colleagues (1997, 2000), the difference in standard deviations might not be due to greater differentiation of higher-ability individuals but the difference in variability might as well be caused by poorer reliability of the personality ratings of less gifted. Thus, the internal reliability and variability of scores are confounded. Unfortunately this possibility was not discussed by Harris and her colleagues (2005). In this study it was difficult to give a conclusive answer to the problem of variability even when aggregated ratings of others were used since in this data also the ratings on lower-ability targets were less reliable. Thus, the question

about of the effect of ability on interindividual variability in personality is still not conclusively answered. Here a parallel from cross-cultural studies can be drawn. McCrae (2002) reported that there appeared to be a systematic cross-cultural trend in the variation of personality test scores. Cultures that had higher variation for some scales tended to have consistently higher variation for other scales as well. The variability of scores tended to be lower in Asian and African cultures and higher in Western countries. McCrae (2002) gave several possible explanations (among others he discussed lower internal reliabilities in low-variance cultures) but neither of them was clearly preferable. In a recent study McCrae and Terracciano (2005) discussed the effect of data quality on gender and age differences in personality and concluded that the "variations in quality contribute to, but not fully account for, the observed age effects, and the same is likely true of sex effects" (pp. 558). In summary, the problem of separating measurement error from the real variability of personality test scores has so far proven to be too difficult to be completely resolved.

The present study has two major implications for personality research. First, although cognitive ability was to some extent related to the quality of personality ratings, a relatively high cross-observer agreement demonstrated that traits were assessed rather accurately in both levels of cognitive ability. Of course, it is difficult to establish which correlations between ratings of different observers can be considered adequately high (McCrae, 1982). However, since there are sufficiently studies that have examined the agreement between observes in different cultures the most reasonable criterion is to compare present findings with previous results. The comparison is favorable for the present results since in most cases the cross-observer agreement on five personality domains was at least as high or higher than the "international median" (for five domains r = .40 – .47; McCrae et al., 2004), even the agreement in low-ability group. One reason for the relatively high agreement in low-ability individuals might be the use of linguistically simple measure. Thus, observing one's own behavior, thoughts and feelings and making valid personality judgments on their basis is not a particularly demanding task. Even intellectually not very sophisticated individuals can make relatively valid ratings on their personality traits, at least when items are formulated in a readable form. Nevertheless, as developmental data demonstrate, a minimal amount of ability is required for this task (Allik et al., 2004). Several studies have shown that self-report personality measures can be used in adolescents as young as 10 or 12 years (Markey, Markey, Tinsley, & Ericksen, 2002; De Fruyt, Mervielde, Hoekstra, & Rolland, 2000) although their applicability also

depends on rater's level of intellectual development (Allik et al., 2004). Along with these results, the present findings that generally support the use of linguistically simple self-report measures with intellectually less talented and educated adults are in no doubt good news for personality assessment. Of course, the situation would probably be different if individuals with extremely low cognitive ability were studied. In the present study, the participants were most likely in the range of normal ability.

However, a minor decline of personality data quality was still observed in the group of individuals with lower ability. This is not surprising because it is relatively well documented that the ability to judge personality traits is higher in those persons who are intelligent, have wide range of interests and who are both emotionally and socially well-adjusted (Taft, 1955). Analogously, Davis and Kraus (1997) concluded that good judges are intelligent, they see the world in a cognitively complex, sophisticated way, and describe themselves as gregarious and responsible. The internal reliability of some facet scales of the EPIP-NEO was remarkably lower in the group of individuals with lower ability and this cannot be ignored. For this reason, when studying individuals with potentially lower level of ability it is highly desirable to obtain ratings from multiple sources. Aggregated ratings are likely to be more reliable and valid than self-report ratings only (Kolar et al., 1996).

Another important implication of this study is related to the demonstration that cognitive ability does not have remarkable effect on the structure of human personality. If the opposite was true, studying personality would probably be more complicated. According to the nomothetic hypothesis (Eysenck, 1954), all personality traits are applicable to all subjects. If cognitive ability had a substantial effect on personality structure, this assumption would turn out to be questionable. In this case a model would hold only for individuals who have reached a certain level of cognitive development. For different ability levels different personality models or, even worse, no model would be appropriate. Fortunately, results of this study indicate that there is no obvious need to work out different models of personality for the description of individuals at different levels of cognitive ability. According to present results, FFM is almost equally well applicable to individuals, at least in the range of normal cognitive ability. Not less important is the conclusion that cognitively less able individuals are not demonstrably more "primitive" and uniform, having "less" personality (Harris et al., 2005).

The present study, however, has several limitations. Since the CAT as a measure of cognitive ability has not been standardized in a large and representative sample, it was

difficult to ascertain into which IQ-range the participants exactly fell. Although there was no clinical information available about the participants, it was not likely that the sample included individuals with subnormal intellectual functioning since those with obvious cognitive deficit (e.g. dementia) or severe substance abuse problems were excluded. However, considering the relatively wide range of the CAT scores, the diverse age and educational background of participants, it is highly likely that the sample was rather heterogeneous in relation to cognitive ability. Thus, although the present results cannot automatically be generalized to individuals with extreme ability scores, it seems adequate to draw conclusions about large proportion healthy individuals. The relationship between cognitive ability and personality structure or variability in more extreme ability groups is a matter of future research. Another limitation is a relatively modest size of the sample. Unfortunately, it was not possible to divide the sample into five ability groups like in the study of Toomela (2003). However, the differences between two ability groups were, if present, relatively small. This makes it possible to hypothesize that there would be no dramatic differences in the personality structure, when for example individuals with IQ score of 70 – 85 were compared to individuals with IQ score of 115 – 130. Nevertheless, this hypothesis should be tested on larger samples. Finally, it is difficult to demonstrate to which extent the present results were affected by the use of linguistically minimalist personality measure (Mõttus et al., submitted). In order to provide such demonstration a linguistically more sophisticated measure (such as the NEO-PI-R) should have been administered in parallel with the EPIP-NEO. In the present study, this was difficult because due to the individually administered CAT participants could have been overloaded with test materials.

Acknowledgements

Authors are most grateful to Maie Kreegipuu, Kristo Burk and numerous other enthusiastic persons for their efforts with organizing and carrying out testing. This project was supported by the Estonian Science Foundation Grant 5677 and an Estonian Ministry of Education and Science Grant 0180543 provided for the second author.

References

Ackerman, P.L., & Heggestad, E.D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 212,* 219 – 245.

Allik, J., Laidra, K., Realo, A., & Pullmann, H. (2004). Personality development from 12 to 18 years of age: Changes in mean levels and structure of traits. *European Journal of Personality, 18,* 445-462.

Allik, J., & McCrae, R.R. (2004). Escapable conclusions: Toomela (2003) and the universality of trait structure. *Journal of Personality and Social Psychology, 87,* 261-5.

Allik, J., & Realo, A. (1997). Intelligence, academic abilities, and personality. *Personality and Individual Differences, 23,* 809-814.

Austin, E.J., Deary, I.J., & Gibson, G.J. (1997). Relationships between ability and personality: Three hypotheses tested. *Intelligence, 25,* 49-70.

Austin, E.J., Deary, I.J., Whiteman, M.C., Fowkes, F.G.R., Pedersen, N.L., Rabbitt, P., Bent, N., & Mcinnes, L. (2002). Relationships between ability and personality: Does intelligence contribute positively to personal and social adjustment? *Personality and Individual Differences, 32,* 1391-1411.

Austin, E.J., Hofer, S.M., Deary, I.J., & Eber, H.W. (2000). Interactions between intelligence and personality: Results from two large samples. *Personality and Individual Differences, 29,* 405-427.

Brand, C., Egan, V., & Deary, I.J. (1994). Intelligence, personality, and society: "Constructivist" versus "essentialist" possibilities. In D.K. Detterman (Ed.), *Current Topics in Human Intelligence.* Volume 4. (pp. 29-42). Norwood, New Jersey: Ablex.

Cattell, R.B. (1957). *Personality and motivational structure and measurement*. New York: World.

Church, A.T., & Burke, P.J. (1994). Exploratory and confirmatory tests of the big five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology, 66,* 93-114.

Costa, P.T.Jr., & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEOFFI) – Professional Manual.* Odessa, FL: Psychological Assessment Resources, Inc.

Costa, P.T., Terracciano, A., & McCrae, R.R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology, 81,* 322-331.

Davis, M.H., & Kraus, L.A. (1997). Personality and empathic accuracy. In W. Ickes (Ed.), *Empathic accuracy.* (pp. 144-168). New York: Guilford Press.

De Fruyt, F., Mervielde, I., Hoekstra, H.A., & Rolland, J.P. (2000). Assessing adolescents' personality with the NEO PI-R. *Assessment, 7,* 329-345.

Eysenck, H.J. (1954). The science of personality: nomothetic. *Psychological Review, 61,* 339-342.

Eysenck, H.J. (1994). Personality and intelligence: Psychometric and experimental approaches. In R. Sternberg & P. Ruzgis (Eds.), *Personality and intelligence.* (pp. 3-31). New York: Cambridge University Press.

Funder, D.C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102,* 652-70.

Funder, D.C., & Colvin, C.R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55,* 149-158.

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.

Jackson, D.N. (1989). *Personality Research Form manual* (3rd ed.). Port Huron, Michigan: Sigma Assessment Systems.

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.

Kaufman, A.S., & Lichtenberger, E.O. (1999) Essentials of WAIS-III assessment. New York: Wiley.

Kolar, D.W., Funder, D.C., & Colvin, C.R. (1996). Comparing the accuracy of personality judgements by the self and knowledgeable others. *Journal of Personality, 64,* 311-37.

Markey, P.M., Markey, C.N., Tinsley, B.J., & Ericksen, A.J. (2002). A preliminary validation of preadolescents' self-reports using the Five-Factor Model of Personality. *Journal of Research in Personality, 36,* 173-181.

McCrae, R.R. (1982). Consensual validation of personality-traits - evidence from self-reports and ratings. *Journal of Personality and Social Psychology, 43,* 293-303.

McCrae, R.R., Costa, P.T., Martin, T.A., Oryol, V.E., Rukavishnikov, A.A., Senin, I.G., Hrebickova, M., & Urbanek, T. (2004). Consensual validation of personality traits across cultures. *Journal of Research in Personality, 38,* 179-201.

McCrae, R.R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: data from 50 cultures. *Journal of Personality and Social Psychology, 88,* 547-61.

McCrae, R.R., Zonderman, A.B., Costa, P.T., Bond, M.H., & 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.

McCrae, R.R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparisons. In R.R. McCrae & J. Allik (Eds.), *Five-Factor model of personality across cultures.* (pp. 105-125). Kluwer Academic/Plenum Publishers.

McCrae, R.R., & Costa, P.T. (1985). Updating Norman's "adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology, 49,* 710-721.

Mõttus, R., Pullmann, H., & Allik, J. (submitted for publication). Towards More Readable Five-Factor Personality Inventories. *European Journal of Psychological Assessment*.

Shure, G.H., & Rogers, M.S. (1963). Personality factor stability for three ability levels.: Personality factor stability for three ability levels. *Journal of Psychology: Interdisciplinary & Applied, 55,* 445-456.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin, 52,* 1-23.

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,* 723-35.

Woodruffe, C. (1985). Consensual validation of personality traits: Additional evidence and individual differences. *Journal of Personality and Social Psychology, 48,* 1240-1252.