THE STRUCTURE AND VALIDITY OF SELF- AND PEER-REPORTED PERSONALITY TRAITS

KENN KONSTABEL
Department of Psychology, University of Tartu, Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (in Psychology) on May 19, 2006, by the Doctoral Committee of the Department of Psychology, University of Tartu

Supervisors:  Jüri Allik, PhD, Professor, University of Tartu, Estonia
             Anu Realo, PhD, University of Tartu, Estonia

Opponent:  Jens B. Asendorpf, PhD, Professor
            Institut für Psychologie
            Humboldt-Universität Berlin

Commencement: June 19, 2006

ISSN 1024–3921
ISBN 9949–11–357–1 (PDF)

Autoriõigus Kenn Konstabel, 2006

Tartu Ülikooli Kirjastus
www.tyk.ee
Tellimus nr. 312
CONTENTS

LIST OF ORIGINAL PUBLICATIONS .......................................................... 6

INTRODUCTION........................................................................................... 7

1. Trait Ascriptions to Self and Others as Products of Ordinary Social Perception................................................................................................... 7

2. The Sources of Personality Trait Structure ................................................ 7

3. Measurement and Meaning of the Personality Traits................................ 10

4. Validity and Consensus............................................................................. 11

5. Summary and Conclusions........................................................................ 12

ACKNOWLEDGEMENTS ........................................................................... 14

REFERENCES............................................................................................... 15

SUMMARY IN ESTONIAN ......................................................................... 17

PUBLICATIONS ........................................................................................... 19
LIST OF ORIGINAL PUBLICATIONS

This dissertation is based on the following original publications which will be referred to in the text by their respective Roman numerals.


II Konstabel, K., & Virkus, A. (2006). How similar are the conceptual and empirical structures of personality traits? Manuscript accepted for publication, European Journal of Personality. DOI: 10.1002/per.591


INTRODUCTION

1. Trait Ascriptions to Self and Others as Products of Ordinary Social Perception

There is nothing unusual in the thesis that personality trait ascriptions (including scores on personality questionnaires) are products of ordinary social perception. In fact, it is almost a truism: personality traits are socially relevant dispositions; even the most specifically behavioural items in questionnaires usually involve a considerable abstraction (Block, 1989) and cannot be answered simply by referring to objectively available evidence. Moreover, considering the limitations of human memory and the typically very short time interval that is used to respond to a personality item, people cannot be expected to answer even a slightly abstract item by counting the number of times they have acted or thought of felt in a particular way. A growing body of evidence shows that people usually do not rely on their episodic knowledge when answering to abstract questions about their personalities; they may even be lacking episodic memory at all and still be able to describe their personality traits (Klein & Loftus, 1993; Klein, Loftus & Kihlstrom, 1996; Tulving, 1993).

Following this conceptualization, one can conclude that to a large part, it is the human observer rather than the test or questionnaire that is doing the measurement in personality research. During our everyday activities, we occasionally ‘measure’ our own or others’ qualities (including personality traits); a questionnaire can be used to induce the respondent to recollect these measurements, or to make new ones. In that sense, a questionnaire can only be considered to be a measurement instrument in conjunction with the social or self-perceptions from the part of the respondent; by itself, it is just a device for eliciting and recording the answers. In the following pages, I will elaborate some of the consequences of this thesis, concentrating on the aspects converging with my own research.

The overall goal of this dissertation is to explore the determinants of the structure and consensual validity of personality traits. The first two studies aim to find regularities in the cross-cultural variations in the five-factor structure of traits, and in the differences between the conceptual and empirical structures. The next two studies investigate the role of social desirability and self-enhancement in responses to personality items, and the role of these phenomena in consensual validity.

2. The Sources of Personality Trait Structure

In this section I describe two strategies for studying the sources of personality trait structure: comparing the empirical and conceptual structures of traits (e.g., the structures of self-reports and trait covariation ratings), and examining the
cross-cultural variations in structure. The basic tenet here is that cross-cultural similarities (McCrae & Costa, 1997) are not highly informative about the sources of structure because they could be caused by several kinds of human universals (cultural and ecological as well as biological); on the other hand, systematic variations and differences can be informative.

2.1. Comparing the Empirical and Conceptual Structures of Traits

Traits are concepts used to categorize people — therefore, it seems natural to assume that trait taxonomy, if truthful, reflects the covariation of personality attributes among people. In another sense, however, trait taxonomy is also about trait terms, and reflects the meanings given to these terms by the users of ordinary language. Trait taxonomy is thus about people in two different senses — people as objects of categorization (‘empirical structure’), and people as users of trait terms (‘conceptual structure’). It may seem that these two options cannot be disentangled in an empirical study, because when describing someone's traits, one by necessity uses ordinary language. However, at least a partial separation is possible when the question is made more specific: can it be shown that an empirical trait taxonomy contains information that is not available when considering the meaning similarities of trait terms alone?

This question was examined in Study II; briefly, it was found that there are at least two aspects in the empirical self-report structure that are notoriously difficult to explain if one presumes that the empirical structure reflects nothing more than meaning similarities or conceptual relationships between traits (Shweder, 1975; cf. Borkenau, 1992):

(a) the traits of extraversion and neuroticism are almost independent in the empirical structures, but in the three samples of Study II, their markers formed a bipolar dimension with positive loadings on extraversion items and negative loadings on neuroticism items;

(b) in the empirical structure, impulsiveness is related to both neuroticism and (lack of) conscientiousness, as well as (to a lesser extent) extraversion (Costa & McCrae, 1992; Kallasmää et al., 2000), but in the structure of covariation ratings in Study II, it was almost exclusively related to (lack of) conscientiousness.

It can be concluded, thus, that the empirical (self- and peer-report) and conceptual structures of personality traits are at least in part based on different sources of information; conceptual structure is not just a reflection of observed behavioral covariations, as it has been supposed (e.g., Block, Weiss & Thorne, 1979), and the empirical structure is at least partly based on the information that perceivers have about the target persons of ratings.
2.2. Cross-cultural Variations in the Empirical Structure

Another strategy for disentangling the sources of trait structure is to examine the cross-cultural or cross-sample correlates of variations in the structure. In Study I, we examined two variations of this kind: the position of the extraversion/agreeableness axis after the varimax rotation, and the varying correlations of impulsiveness with the extraversion and conscientiousness factors.

The first kind of variation — the position of the so-called ‘interpersonal axes’ — is significant because it may reflect a systematic difference in how the traits of extraversion and agreeableness are conceptualized in different cultures. In some samples, factors close to the North American extraversion and agreeableness emerge; in other samples, these blends of the markers of these dimensions form two different factors that have been labelled ‘love’ and ‘dominance’. It has been shown, though, that these differences are largely a matter of the rotation of axes, and can be almost eliminated with targeted rotation (Costa & McCrae, 1997). Nevertheless, Kallasmaa and colleagues (2000) have demonstrated that the position of interpersonal axes is correlated with a country's level of individualism/collectivism, possibly reflecting a preference of collectivists for the dimensions of evaluation and potency (Osgood, Suci & Tannenbaum, 1957) in explaining behaviour. In Study I, we found that in addition to a country-level association, the same pattern can be observed within one country (Estonia) when the sample is divided according to the mean levels of collectivism. In addition, it was found that the position of axes is related to the country's mean level of subjective well-being and life satisfaction.

With regard to the correlates of impulsiveness, it was found that the association of impulsiveness with extraversion could be predicted by the country’s mean level of individualism, and (with a reversed sign), conscientiousness. The association of impulsiveness with conscientiousness could be predicted by life satisfaction and normative life satisfaction. It was speculated that because one of the most important attributes of collectivism (as an opposite to individualism at the cultural level) is an emphasis on social norms and duty defined by the group (Triandis, 1995), the expression of extraversion in impulsive behavior may be inhibited by norms and rules in collectivistic and less open cultures, and that higher levels of conscientiousness may hinder the expression of extraversion in impulsive behavior. This reasoning needs to be complemented by data from conceptual structure: if the findings in conceptual structure would mimic the cross-cultural variations in the relationship of impulsiveness to extraversion, neuroticism, and conscientiousness, this would lend some indirect support to the idea that the characteristics of a culture may have an influence on the empirical structure of personality.
3. Measurement and Meaning of the Personality Traits

The classical measurement model states that the observed score consists of true score and random measurement error. One of the hidden assumptions here is that measurement occasions are repeatable: we can ask the same question twice and receive answers that differ only by random error. In this regard, a thought experiment by Lord and Novick (1968) is especially notorious:

Suppose we ask an individual, Mr. Brown, repeatedly whether he is in favour of the United Nations; suppose further that after each question we ‘wash his brains’ and ask him the same question again. Because Mr. Brown is not certain as to how he feels about the United Nations, he will sometimes give a favorable and sometimes an unfavorable answer. Having gone through this procedure many times, we then compute the proportion of times Mr. Brown was in favor of the United Nations. (Quoted by Borsboom, Mellenbergh & Van Heerden, 2002).

This assumption of repeatability makes perfect sense in the case of physical measurements (e.g., that of length) but in the context of personality measurement, it is just a theoretical possibility comparable to that of building a stainless steel ladder to the moon (cf. Dennett, 1991, p. 4 on ‘in principle’ possibilities). Asking questions in a questionnaire is “subject to the rules that govern social relations between strangers” (Sudman, Bradburn & Schwartz, 1996), and these rules do not generally allow us to ask one and the same question twice or two very similar questions in a row. Even if one would want, doing so would be useless because people could easily repeat their earlier answer, and can even be worse than useless because the respondents might become suspicious about the researcher's motives. Because of these practicalities, the idea of a scale consisting of ‘in principle’ interchangeable replications of the same question has been replaced by the idea of representative sampling of items from a content domain (Haynes, Richard & Kubany, 1995). If the trait (or the content domain of measurement) is even slightly abstract, then its indicators (e.g., responses to personality items) are likely to have multiple causes besides the latent trait, and, if at least one of these causes influences more than one person, also a multi-dimensional correlation structure. This conjecture is at odds with the unidimensionality assumption in classical test theory, but largely consistent with the existence of socially desirable responding described in Study III and the tendency of self-enhancement described in Study IV.

Finally, the formal measurement models are unlikely to provide a solution to a more fundamental problem in personality measurement: that of “cognitive reliability”, or the dependability of the processes by which respondents arrive at their answers (cf. Tourangeau, Rips & Rasinski, 2000). Studies III and IV might provide starting points for experimental studies of the process of responding to personality items. Two findings from Study III are relevant in this respect. First, it was found that the items' standard deviations varied as a
quadratic function of the items' average social desirability ratings: the responses were more uniform on evaluative items and more varied on neutral items. A tentative interpretation of this finding is that the social norms may not only shift the modal response, but they also make people respond in a more uniform way. Another finding was that controlling for an index of social desirability proposed by Hofstee (2003) heightened the self-peer and peer-peer correlations on most traits of the five-factor model. A likely interpretation is that a motivation to respond in a socially desirable manner (comprising both individual tendencies and situational factors) can shift the responses toward a socially desirable answer; this shift occurs to different degrees for different individuals, and can partly be balanced by controlling for the index of social desirability. A related finding from the Study IV was that in people with higher self-esteem, the within-individual agreement between self-ratings and social desirability ratings was also higher, even when controlling for the average ratings. This finding may reflect a tendency to respond in a socially desirable manner, although another interpretation, that of egocentric ratings of social desirability or self-serving trait definitions is also possible.

4. Validity and Consensus

In a recent paper, Borsboom, Mellenbergh and van Heerden (2004) have argued for a causal concept of validity: in their view, a measure can be called valid if the trait it purports to measure exists, and causally produces variations in the measurement outcomes. This is an important clarification of the concept of validity, but it also raises some difficult questions. What does it mean for a trait (or an ‘abstract tendency,’ McCrae & Costa, 1996) to ‘exist’? It is clear that individual differences in behaviour do exist and correlate with other individual differences, but which evidence is needed to prove the existence of a trait? According to Borsboom and colleagues (2004), one should look at the processes that generate the item responses, but here the difficulty is that for investigating whether a trait causes the responses, one should somehow be able to identify the trait before measuring it. One could think of two ways to do it: the traits could be traced back to (a) socially relevant categories (e.g., Buss & Craik, 1983; Borkenau, 1992), or to (b) psychobiological structures and processes (e.g., Zuckerman, 2005). In the long run, these approaches should, of course, complement each other.

Picking the first option for the time being, there is a large choice of different criteria of validity, and none of them should be preferred on a priori grounds. However, there is a sense in which peer-reports are among the most relevant validity criteria for self-reported traits. Objectively measurable validity criteria — for instance frequency counts of behaviors and life outcomes — can hardly ever be thought of as conceptually equivalent to any given personality
trait; therefore, an ideal correspondence between the trait scale and the criterion cannot be expected even in the imaginary condition of perfect measurement accuracy. Reports by acquaintances, although based on ordinary social perception just like self-reports, constitute an important validity criterion because their informational basis, as well as the category breadth of the trait descriptors, is at least comparable to those of the self-reports.

In principle, the consensus between self- and peer-reports could capitalize on response tendencies, for instance, socially desirable responding: because people and their acquaintances who provide peer-reports are likely to share a social background and common values, these response tendencies might be shared too. However, in Study III, we showed that at least for socially desirable responding, this is probably not the case. Partialling out the social desirability index actually increased the consensus for both self- and peer-reports, which shows that to a large part, the social desirability ‘bias’ is not shared. In Study IV, it was found that when the desirability of a trait or an item is difficult to judge, people tend to rely on a ‘self-based heuristic’; this heuristic was first described by Ready and colleagues (2000) to be used in peer-ratings of difficult-to-judge traits. From the social perception viewpoint, the existence of a reverse, ‘norm-based’ heuristic could be predicted which would guide self-ratings for items for which clear social norms exist but the individual has no clear self-knowledge.

5. Summary and Conclusions

In sum, the main results and conclusions in the present dissertation are:

- There are systematic cross-cultural variations in the structure of personality traits. In Study I, it was found that at least two kinds of these variations (the position of axes that define the extraversion and agreeableness factors, and the relationship of impulsiveness with extraversion and conscientiousness) are related to country-level variables in a meaningful way.

- There are systematic and replicable differences between the empirical and conceptual structures of personality traits. In Study II, it was found that Neuroticism and Extraversion had a weak negative correlation in self-reports, but were judged to be almost bipolar opposites in covariation ratings; Impulsiveness was judged to be a negative indicator of Conscientiousness in covariation ratings, but was equally strongly related to Neuroticism component in self-ratings. These systematic differences demonstrate that the structure of self-rated traits is not reducible to semantic similarities of traits descriptors.
Socially desirable responding can be efficiently measured by an index proposed by Hofstee (2003): mean cross-product of an individual's self-ratings and average social desirability ratings to the items, where both are expressed as deviations from the scale midpoint (study III). Socially desirable responding as measured by this index was found to be a suppressor of consensus on personality ratings. This is consistent with the finding of Study IV that especially in people with high self-esteem and low neuroticism there is a tendency to self-enhance, either by claiming to possess socially desirable qualities, or by describing their own traits as desirable ones. An important implication here is that scores on broad personality traits consist of at least two components, which can be labeled evaluative and descriptive (cf. Saucier, 1994). In multidimensional questionnaires, these components can, to some degree, be separated by statistical means.

Highly evaluative personality items have lower standard deviations, especially when there is a strong motivation to respond in a socially desirable manner (Study III). This finding probably reflects the influence of social norms on the item responses. However, when there are no clear social norms for what is considered to be a ‘good’ response to a given item, people tend to use a ‘self-based heuristic’ (Ready et al., 2000) to estimate the item's favorability (Study IV).
ACKNOWLEDGEMENTS

Besides my family and friends (what would I do without you?), there are a number of people who have more or less directly helped me in doing the research reported in this dissertation. I would like to thank my advisors, Jüri Allik and Anu Realo, for fruitful discussions and constructive criticism, as well as for not being authoritarian. Several other people have (in some cases perhaps without noticing it themselves) influenced my thinking in important ways, and/or helped me in writing the papers; without them (in approximate chronological order: Bruno Mölder, Jaan Valsiner, Aaro Toomela, Peter Borkenau, Robert R. McCrae, Krista Fischer), this dissertation would be worse. Thanks are also due to my co-authors (Talvi Kallasmaa, Ann Virkus, Toivo Aavik), and — last but definitely not the least, I would like to express my gratitude to the thousands of research participants who were selfless enough to devote a part of their invaluable time for describing their own or their friends’ personalities.
REFERENCES


Väitekirja peamised tulemused ja järeldused on järgmised:

- **Isiksuseomaduste struktuuris on süstemaatilised kultuuridevahelised erinevused (I uurimus).** Vähemalt kaks tüüpi neist erinevustest (ekstravertsuse ja sotsiaalsuse faktortelgede asend pärast Varimax pööramist ning impulsivsuse seosed ekstravertsuse ja meelekindlusega) on interpreteeritavalt seotud maade või kultuuride üldiste omadustega.

- **Isiksuseomaduste empiiriline ja kontseptuaalne struktuur on hoolimata varem leitud sarnasustest ka süstemaatiliselt erinevad (II uurimus).** Omaduste koosnesemise tõenäosuste hinnangute põhjal leitud kontseptuaalsete struktuuris olid neurotism ja ekstraverts ja kirjeldades sotsiaalsuse, kusjuures nii nõustumise määr kui sotsiaalne soovitavus on sõltuva erinevalt. Impulsiivsus peeti koosnesemise tõenäosuse hinnangutes meelekindluse negatiivseks indikaatoriks ning sõltuva erinevalt. Need süstemaatilised ja korratavad erinevused näitavad, et isiksuseomaduste empiiriline struktuur ei ole taandatav ainult nende omaduste kirjelduste semantiliste sarnasustede ja erinevuste ja erinevuste.

Väga hinnangulistel isiksuseväädetel on väiksem standardhälve, eriti olu-korras, kus vastajal on tugev motivatsioon sotsiaalselt soovitavaks vastamiseks (III uurimus). See tulemus näitab tõenäoliselt sotsiaalsete normide mõju vastustele. Teiselt poolt, kui selged sotsiaalsed normid “hea” vastuse kohta puuduvad, kiputakse isiksuseväädete sotsiaalset soovitavust hindama iseenda omaduste põhjal (IV uurimus).
PUBLICATIONS
CURRICULUM VITAE

Kenn Konstabel

Citizenship: Estonian
Date of birth: June 11, 1975
Address: University of Tartu, Tiigi 78, 50410 Tartu, Estonia
Telephone: +372 737 5903
E-mail: kenn.konstabel@ut.ee

Education

1993–1997 Bachelor's studies, Department of Psychology, University of Tartu
1997–1999 Master's studies, Department of Psychology, University of Tartu
1999 Magister scientiarum (MSc) in psychology, University of Tartu
1999–2005 Doctoral studies, Department of Psychology, University of Tartu

Professional employment

2005– Researcher in personality psychology (0.5), Department of Psychology, University of Tartu
2002–2005 Coordinator of the Estonian Centre of Behavioural and Health Sciences (0.5), Faculty of Social Sciences, University of Tartu
2002– Lecturer of personality psychology, Department of Psychology, University of Tartu

Research interests

The structure of personality traits; implicit personality theory; socially desirable responding in personality questionnaires
CURRICULUM VITAE

Kenn Konstabel

Kodakondsus: Eesti
Sünniaeg: 11. juuni 1975
Aadress: Psühholoogia osakond, Tartu Ülikool, Tiigi 78, 50410 Tartu
Telefon: 737 5903
E-post: kenn.konstabel@ut.ee

Haridus

1993–1997 bakalaureuseõpe, Tartu Ülikooli psühholoogia osakond
1997–1999 magistriõpe, Tartu Ülikooli psühholoogia osakond
1999 magister scientiarum (MSc) psühholoogia erialal, Tartu Ülikool
1999–2005 doktoriõpe, Tartu Ülikooli psühholoogia osakond

Teenistuskäik

2005– Isiksusepsühholoogia erakorraline teadur (0.5), Tartu Ülikooli Psühholoogia osakond
2002–2005 Eesti Käitumis- ja Terviseteaduste Keskuse peaspetsialist (0,5), Tartu Ülikooli Sotsiaalteaduskond
2002– Isiksusepsühholoogia lektor (0,5), Tartu Ülikooli psühholoogia osakond

Teadustöö põhisuunad

Isiksuseomaduste struktuur; implitsiitne isikusseteooria; sotsiaalselt soovitav vastamine isiksuseküsimumustikes