JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH
577

Mikael Nederström

Personality Assessment and Self-other Rating Agreement

Moderators and Implications of Agreement

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH 577

Mikael Nederström

Personality Assessment and Self-other Rating Agreement

Moderators and Implications of Agreement

Esitetään Jyväskylän yliopiston kasvatustieteiden ja psykologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa S212
helmikuun 18. päivänä 2017 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Education and Psychology of the University of Jyväskylä,
in Auditorium S212, on February 18, 2017 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2017

Personality Assessment and Self-other Rating Agreement

Moderators and Implications of Agreement

JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH 577

Mikael Nederström

Personality Assessment and Self-other Rating Agreement

Moderators and Implications of Agreement

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2017

Editors
Timo Suutama
Department of Psychology, University of Jyväskylä
Pekka Olsbo, Sini Tuikka
Publishing Unit, University Library of Jyväskylä

Cover picture: "Even dogs' personality traits can be reliably assessed" by Mikael Nederström

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-6961-5

URN:ISBN:978-951-39-6961-5
ISBN 978-951-39-6961-5 (PDF)

ISBN 978-951-39-6960-8 (nid.)
ISSN 0075-4625

Copyright © 2017, by University of Jyväskylä

Jyväskylä University Printing House, Jyväskylä 2017

ABSTRACT

Nederström, Mikael
Personality assessment and self-other rating agreement: moderators and implications of agreement
Jyväskylä: University of Jyväskylä, 2017, 59 p.
(Jyväskylä Studies in Education, Psychology and Social Research
ISSN 0075-4625; 577)
ISBN 978-951-39-6960-8
ISBN 978-951-39-6961-5 (PDF)

One essential aspect of personality test validation is to estimate the magnitude of self-other agreement of personality ratings. In this method, external observers form their personality judgements of the target person. The self-other agreement coefficients obtained provide approximations of the validity of the measure used. The magnitude of self-other agreement on personality has also proven to be useful in explaining real-life criteria, such as managerial performance. The main aim in this research was to study self-other agreement of personality ratings in various organizational contexts. First, the overall magnitude of self-other agreement was examined in a personnel selection sample; second, the moderating factors affecting the level of agreement were estimated; third, the relationships between abnormal and normal personality measures via self-other agreement were explored; fourth, a new personality test was validated; and fifth, the consequences of agreement in leadership behavior were examined. The respondents were real-life job applicants (Studies I and II) and technology managers in a research organisation (Study III). Their personality was measured with four different instruments: Personality Research Form (PRF; Jackson, 1999), Stress Reaction Style (SRS; Nederström & Furnham, 2012), PK5 (2007) and Work Personality Inventory (WOPI; Leung & Zedeck, 2016).The main results demonstrated, first, that it is possible to obtain substantial self-other agreement on personality during a job interview. Second, moderating factors, such as demographic factors and the trait being judged, may affect the magnitude of this agreement. Third, self-other agreement can be employed in validating a new personality measure by using expert ratings as an external criterion. Fourth, the implications of managerial self-other agreement may be useful in predicting the occupational well-being of subordinates. From a practical point of view, it seems that self-other agreement on personality can be an important concept in validating new personality measures and examining organizational phenomena related to personality, such as leadership outcomes and personality assessments in a job interview.

Keywords: self-other agreement, personality assessment, leadership outcomes, dark side of personality

Author's address Mikael Nederström
Psycon Oy
Jaakonkatu 3 A, FIN-00100 Helsinki
mikael.nederstrom@psycon.com

Supervisors Professor Katariina Salmela-Aro
Department of Psychology
University of Jyväskylä, Finland

Professor Jari-Erik Nurmi
Department of Psychology
University of Jyväskylä, Finland

CEO Petteri Niitamo
Competence Dimensions

Reviewers Professor Jüri Allik
Department of Psychology
University of Tartu, Estonia

Docent Marjaana Lindeman
Department of Psychology
University of Helsinki, Finland

Opponent Professor Jüri Allik
Department of Psychology
University of Tartu, Estonia

TIIVISTELMÄ (FINNISH ABSTRACT)

Nederström, Mikael
Persoonallisuuden arviointi, arvioiden yhteneväisyys, arviointia moderoivat tekijät ja yhtenevyyden seuraukset
Jyväskylä: University of Jyväskylä, 2017, 59 p.
(Jyväskylä Studies in Education, Psychology and Social Research
ISSN 0075-4625; 577)
ISBN 978-951-39-6960-8
ISBN 978-951-39-6961-5 (PDF)

Eräs keskeinen näkökulma uusien persoonallisuustestien validointiin on tutkia itsearvioiden ja ulkopuolisten arvioiden yhteneväisyyttä. Näiden kahden arvioinnin yhteneväisyys tuottaa tietoa persoonallisuustestin validiteetista. Yhteneväisyyden avulla voidaan ennustaa myös ns. tosielämän kriteereitä, kuten johtamiskäyttäytymistä. Tämän tutkimuksen tavoite oli tutkia eri lähteistä saatavien persoonallisuusarviointien yhteneväisyyttä erilaisissa työympäristöissä. Ensimmäinen osatutkimus liittyi henkilöarviointitilanteeseen, jossa tutkittiin persoonallisuusarvioiden yhteneväisyyttä moderoivia tekijöitä. Toinen osatutkimus liittyi ns. persoonallisuushäiriöpiirteiden ja vakiintuneempien Big Five –persoonallisuusmallien yhteneväisyyteen, minkä yhteyksien selvittämisessä hyödynnettiin itsearvioinnin ja ulkopuolisen arvioinnin yhdistämisen metodologiaa. Lisäksi haluttiin validoida uutta, työ- ja organisaatiopsykologista persoonallisuushäiriöpiirteisiin liittyvää testiä. Kolmannessa osatutkimuksessa tutkittiin, kuinka esimiehen persoonallisuuden itsearvioinnin ja ulkopuolisen arvioinnin yhteneväisyys vaikuttaa alaisten työtyytyväisyyteen. Tutkimuksen koehenkilöt olivat henkilöarviointiin tulevia työnhakijoita (osatutkimukset 1 ja 2) sekä suuren suomalaisen tutkimusorganisaation esimiehiä ja työntekijöitä (osatutkimus 3). Persoonallisuutta mitattiin neljällä eri persoonallisuustestillä: Personality Research Form (PRF; Jackson, 1999), Stress Reaction Style (SRS; Nederström & Furnham, 2012), PK5 (2007) ja Work Personality Inventory (WOPI; Leung & Zedeck, 2016). Tulokset osoittivat, että parhaimmillaan työhaastattelussa persoonallisuusarviointien yhteneväisyys voi olla suhteellisen suurta, vaikka mukana on moderoivia tekijöitä, kuten arvioitava piirre, arvioitavan ikä, sukupuoli ja sosiaalisesti suotava vastaamistyyli. Lisäksi uuden persoonallisuustestin (SRS) validiteetille saatiin tukea löytämällä useita merkitseviä yhteyksiä sen ja vakiintuneemman, Big Five –malliin perustuvan testin (PK5) väliltä. Tulokset osoittivat myös, että persoonallisuusarviontien yhteneväisyys esimiehen ja alaisen välillä voi ennustaa alaisten työtyytyväisyyttä.

Avainsanat: persoonallisuuden arviointi, henkilöarviointi, johtaminen, persoonallisuuden pimeä puoli, työtyytyväisyys

ACKNOWLEDGEMENTS

This has been a long journey. It started as an experimental project, when my first intention was to study "something interesting" about personality. At the time, I didn't know whether this project would ever end or not. Now it seems that it wasn't a never-ending story after all!

One of the greatest inspirations along the way came from my, chronologically, first supervisor, Petteri Niitamo, whose optimism has always made things look a little bit smoother and brighter than one might think. And his optimism was truly needed in many instances when the obstacles seemed overwhelming. Not to forget Petteri's many intriguing theoretical ideas about the phenomenon of personality that I was able to draw on in deepening my research.

Luckily, my main supervisor, Katariina Salmela-aro, was able to contribute a dash of realism, which was needed to finalize the articles and get them published. Without the efforts of these two different personalities, this thesis would never have materialized. Furthermore, Jari-Erik Nurmi also gave me some very valuable advice during the final stages.

I also wish to thank the excellent reviewers of this thesis. I appreciate the time and expertise devoted by Professor Jüri Allik and Docent Marjaana Lindeman to reviewing my thesis. They made some very percipient and encouraging comments during the last few steps.

Private discussions with Professor David Funder and Professor Filip De Fruyt were very enlightening and inspiring. These two scholars encouraged me in feeling that my studies were on the right track.

Michael Freeman did an excellent job understanding some complicated and messy thoughts and turning them to simple, elegant and grammatically correct English.

A great deal of inspiration, encouragement and patience came from my employer, Psycon Oy. I have made use of a great amount of Psycon's time, resources and data, all of which, however, enabled some really significant, real-life research to be done. Many discussions with experienced and brilliant consultants have also inspired my research and helped me in reaching my conclusions. Our firm's assistants were very valuable with the data (Vappu Rautio with SRS & PK5) and editing the thesis (Kristiina Vehkomäki).

And, last but not least, my family has been a great source of support. Jemina, my beloved wife, has patiently listened to my weird speculations for many years. Lilia, Eedit and Peppi have been helpful by just being there. Not to forget all grandparents (including "pappa"), who have taken care of the children.

Helsinki 8.12.2016
Mikael Nederström

LIST OF ORIGINAL PUBLICATIONS

I Nederström, M. & Salmela-Aro, K. (2014). Self-other agreement of personality judgements in job interviews: Exploring the effects of trait, gender, age and social desirability. *Scandinavian Journal of Psychology, 55*, 520-526.

II Nederström, M. & Furnham, A. (2012). The relationship between the FFM traits and personality disorders in a personnel selection sample. *Scandinavian Journal of Psychology*, *53*, 421-429.

III Nederström, M., Furnham, A., Yli-Pärri, K., Auvinen, E., & Weis, L. (2016). Self-other agreement on personality ratings in manager-subordinate relationships as a predictor of occupational well-being. Submitted manuscript.

Taking into account the instructions and comments made by co-authors, the author of the thesis collected the data, conducted the analyses, and wrote the reports of the three publications, independently.

FIGURES

TABLES

CONTENTS

1 INTRODUCTION

Millions of people around the world are being assessed with different types of personality tests (Paul, 2005). It has been estimated that in Finland alone, tens of thousands assessment tests are administered every year (Lavonen, Myyry, & Helkama, 2004). Besides the assessment of mental health, personality tests are used to evaluate job applications, criminal matters, and even the fitness of spouse candidates (Paul, 2005). It is extremely important, therefore, that the tests are as socially fair and valid as possible.

One essential aspect of personality test validation is to estimate the magnitude of self-other agreement of personality ratings (e.g., Funder, 1999), which is a well-established component of the convergent validation process (Campbell & Fiske, 1959). The recent research generally concludes that personality judgments can reach a high level of self- and peer agreement on the basis of considerably thin information when the congruence of target self-reports and external judge ratings is used as a criterion (Allik, de Vries, & Realo, 2016; Biesanz, et al., 2011; Connelly & Ones, 2010; Funder, 1999; John & Robins, 1993; Paunonen, 1989; Zebrowitz & Collins, 1997). This methodological approach has been employed throughout this research in examining convergent validity. The main aim of this research was to study self-other agreement on personality ratings and its moderators and real-life consequences. First, the overall magnitude of the self-other agreement in a personnel selection sample was examined; second, the moderating factors affecting the level of agreement were estimated; third, the relationships between abnormal and normal personality measures via self-other agreement were explored; fourth, a new personality test was validated by applying self-other agreement coefficients to estimate convergent/divergent validity of the test; and fifth, the consequences of agreement in leadership behavior were examined. To answer these questions, this research employed data drawn from a real-life selection process in a large Finnish assessment center. These data sets were derived from several psychological personality measures and from expert ratings of personality from structured selection interviews. In addition, to examine the consequences of self-other agreement, data from a personnel survey combined with ratings of personality in a Finnish expert organization were used (see Figure 1: The relationships between the studies).

The main aim in Study I was to examine self-other agreement on personality judgments in a personnel selection process. What demographic and personality factors affect agreement? I concentrated on the general magnitude of self-other agreement and its five potential moderators: the trait being judged, the evaluativeness of the trait, the effect of social desirability, the gender, and age of the target.

In Study II, my main purpose was to study the relationships between the Big Five model of personality, which is the most well-established answer to the plurality of personality classifications (Costa & McCrae, 2006), and the DSM personality disorders (Diagnostic and Statistical Manual of Mental Disorders; American Psychiatric Association, 2013) in a psychological assessment center. My interest was in extending the research methodology beyond correlational self-report studies. This was achieved by using expert observations by psychologists as an additional criterion. Using self-reports alone may artificially inflate the trait-trait relationships because of common method variance, that is, the proportion of the variance in a that is attributable to the measurement method used (Doty & Glick, 1998). One way of overcoming this problem of bias is through comparing self-report personality scales to external judgments of personality (cf. Funder, 1999; Funder & Colvin 1988; John & Robins, 1993). I used the self-other agreement coefficients obtained between the Big Five model and DSM ratings to examine the convergent and divergent validities of the new DSM personality measure.

In Study III, my main purpose was to examine self-other agreement on personality between leader and subordinate ratings and leadership outcomes. This was done by examining self-other agreement and the direction of the discrepancy between the two types of ratings as predictors of the occupational well-being of subordinates. Building on the idea that agreement on personality can predict leadership outcomes, I proposed that the level of self-other agreement and the direction of the discrepancy between ratings are likely to contribute an important perspective on leadership behavior, as this brings incremental validity above and beyond self-reported personality.

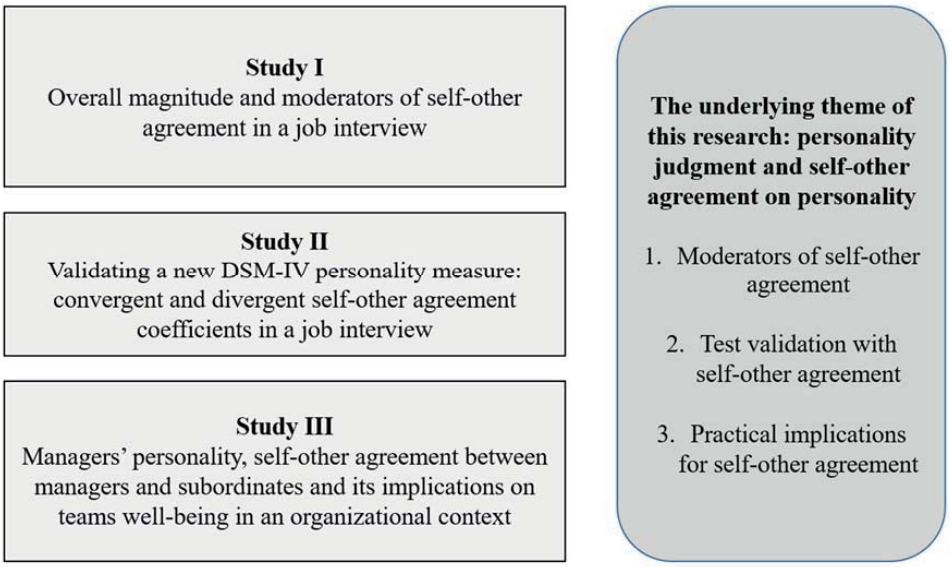


FIGURE 1 The relationships between the studies.

1.1 The essence of personality

While there might not be a single, consensual definition of personality, most researchers presently agree on broad definitions, such as that offered by Pervin (1996) or Funder (1997). Pervin (1996, p. 6) defines personality as "the complex organization of cognitions, affects, and behaviors that give direction and pattern to a person's life". Funder (1997, p. 1) states that "personality refers to an individual's characteristic patterns of thought, emotion, and behavior, together with the psychological mechanisms – hidden or not – behind those patterns".

These definitions are compatible with the ontologically realistic approach that explains personality with inner dispositions. These dispositions are actual attributes of individuals rather than just reflections of specific social constructs in the situation (Borkenau & Liebler, 1992; Funder, 1987; McCrae & Terracciano, 2005). Dispositions, such as personality traits, are not only linguistic concepts but they also show considerable longitudinal consistency (Soto, John, Gosling, & Potter, 2011), can predict behavior (Poropat, 2009), and can also be correlated with genetic factors (Bouchard & McGue, 2003). Growing evidence of the biological and genetic underpinnings of personality (Borkenau, Riemann, Angleitner, & Spinath, 2001; Bouchard, Lykken, McGue, Segal, & Tellegen, 1990; McCrae & Terracciano, 2005) provides empirical confirmation for the realistic, dispositional approach. In their meta-analysis, Roberts and DelVecchio (2000) demonstrated that despite potential changes in individual personality traits, the rank-order, or profile, of personality factors remains largely unchanged from

play-age to old age (Roberts & DelVecchio, 2000; see also Löckenhoff et al., 2008).

The realistic interpretation of personality is also close to the layman's intuition. Throughout history, a dispositional and trait-based definition has probably provided humans with the most effective tool of judging personality and predicting others' behavior (Haselton & Funder, 2005). The notion of personality dispositions is extremely useful when deciding whom to cooperate with, whom to avoid or whom to select as a long-term mate. More accurate personality judgments usually lead to better predictions and a better chance of surviving during evolution. The dispositional and trait-based pattern seems to accord with the prevailing layman's theory of personality in almost every culture (McCrae & Terracciano, 2005).

1.2 Personality tests: their use and validity

According to the definition by Hogan, Hogan and Roberts (1996, p. 470), "personality measurement is any procedure that systematically assigns numbers to the characteristic features of a person's interpersonal style according to some explicit rules. These numbers can then be used to make predictions about that person's responses in future settings." The most widely adopted personality test formats in the selection context are self-report questionnaires (Barrick & Mount, 1991; Schmidt & Hunter, 1998). A common component of all self-report tests is that they provide a limited number of structured items for the target to select from. Unlike more controversial projective personality tests (cf. Parker, Hunsley, & Hanson, 1988), self-report inventories have attained empirically substantiated standing and a relatively high predictive validity, as has clearly been shown in several meta-analyses (Barrick & Mount, 1991; Connelly & Ones, 2010; Poropat, 2009). A simplistic procedure for constructing self-report questionnaires could be defined as follows (Rorer & Widiger, 1983, p. 433):

> If you want information from someone, the best way to get it is to ask them. […] Assuming that they understand the question, that they have the information, and that they are not motivated to deceive you, that is not only the simplest and least expensive, but also the most accurate procedure. That fact is often threatening to those who are at pains to make the enterprise seem complicated and mysterious so as to support that one needs highly trained professional to administer and interpret obtuse and highly convoluted protocols.

Self-report questionnaires are particularly useful in industrial and organizational psychology: they are widely applied in personnel selection, leadership development and predicting job performance (Barrick & Mount, 1991; De Fruyt et al., 2009; Hogan, Hogan & Roberts, 1996; Schmidt & Hunter, 1998). Prior research has also found relationships between well-established leadership frameworks and personality questionnaire profiles (Judge & Bono, 2000; Judge, Bono, Ilies, & Gerhadt, 2002).

To measure self-reported aspects of personality, the reliability and validity estimates of self-report questionnaires need to be demonstrated by applying several methods. Using self-reports alone may artificially inflate the trait-trait relationships because of common method variance, i.e., the proportion of the variance that is attributable to the measurement method employed (Doty & Glick, 1998). One way of overcoming this bias problem is comparing self-report personality scales to external judgments of personality. In addition to more established internal validity analyses (e.g., relationships between self-reports, factor analyses), this method of validating a personality measure includes computation of self-other agreement, which, in turn, can be interpreted as a convergent validation (Funder, 1999).

Ratings by others, often called other-ratings, have also proven to be useful as an external real-life criterion for the measurement of self-reported traits (Connelly & Ones, 2010). External observers form their judgements through independent observation of the conditions affecting the target person. Self-other agreement coefficients provide approximations of the convergent and discriminant (divergent) validity of a measure (Campbell & Fiske, 1959). This method is also well in line with the realistic interpretation of personality: if both self-reports and external ratings show substantial agreement on certain personality dimensions, it is unlikely that these dimensions are merely an artificial product of contextual factors (Funder, 1987).

A main emphasis in the present research was on studying the convergent/divergent validity (both of which are considered subtypes of construct validity) of personality assessments. This was enabled by exploring the magnitude of self-other agreement on personality ratings. Convergent validity helps to establish construct validity by demonstrating that the measures of constructs that theoretically should be related to each other (e.g., self-reports and external judgments) are in fact correlated with each other. Divergent validity helps to establish construct validity by demonstrating that the construct the judges are assessing (e.g., anxiety) is different from other constructs that might emit similar trait-relevant cues (e.g., guilt feelings).

In the present research, other-ratings of personality were applied as an external and independent criterion. The self-other agreement procedure also helps to explore divergent and convergent validity as separate constructs (Campbell & Fiske, 1959). Hence, in an ideal judgment situation, the highest convergent correlation should exceed the highest divergent correlation, which indicates that the scales are clearly separable from each other. Thus, support for the validity of the instrument is obtained by showing that the convergence between self- and other-ratings exceeds the divergence on the correlated scales (cf. Bastiaansen, Rossi, & De Fruyt, 2012). This approach was employed in Studies I and II in a personnel selection context, while Study III focused on the real-life implications of self-other agreement and demonstrated its predictive value.

1.2.1 ”Normal” measures of personality

Personality can be studied and measured using many different classifications of dispositions. The most well-established answer to the plurality of classifications is probably the Big Five theory, which describes personality with five rather stable and universal traits: neuroticism, extraversion, openness to experiences, agreeableness, and conscientiousness (McCrae & Costa, 1985; De Fruyt, McCrae, Szirmák, & Nagy, 2004; McCrae & Terracciano, 2005). After years of debate, the Big Five has gradually become one of the most widely employed normal trait classification instruments (Costa & McCrae, 2006; De Fruyt et al., 2004; Goldberg, 1990; Salgado, 2003).

The Big Five framework represents personality at the broad level of abstraction, but by applying factor analysis it is also possible to find other solutions and subscales which measure personality factors on more or less discrete levels (e.g., Ashton, Jackson, Helmes, & Paunonen, 1998; Lord, 2007; Nederström & Niitamo, 2010). The final choice of the factors to be measured is always partially a convention. However, it seems that more than five factors are needed, if the purpose is to obtain a nuanced description (Paunonen & Jackson, 2000) or a predictive outcome (Hough, 1992) rather than just a global factor description. The Big Five domains can be further decomposed into primary factors or facets, when more detailed measurement is necessary (e.g., Costa & McCrae, 2006). The Finnish PK5 measure (PK5, 2007), utilized in Study II, is an example of the Big Five measure describing five global traits and 15 facets.

Among the well-known Big Five inventories currently in use, Allport’s (1937) distinction between expressive-stylistic traits and motivational traits and Murray’s famous motivational taxonomy (Murray, 1938) have inspired several instruments measuring personality, such as Douglas Jackson’s Personality Research Form (PRF) (Jackson, 1999) and Work Personality Inventory (WOPI) (Nederström & Niitamo, 2010), both of which were applied in the present research. The focus in Study I was on 15 motivational traits of the Finnish version of PRF (Niitamo, 1997), which includes two neuroticism-related traits (anxiety and guilt feelings). The conceptual structure of Finnish PRF is also comparable to some earlier self-other agreement studies (Paunonen, 1989).

Study III employed WOPI, a standardized self-report questionnaire of personality which approaches the construct of personality from an organizational angle (Leung & Zedeck, 2016; Nederström & Niitamo, 2010). This measure has been developed specifically for work-life settings. This frame-of-reference effect has several benefits when used as predictor construct. This effect on personality scales may enable the attainment of incremental validity above and beyond the more well-established, non-contextual personality measures (Bing, Whanger, Davison, & VanHook, 2004). Antonakis, Avolio, and Sivasubramaniam (2003) demonstrated that the psychometric properties of leadership instruments are affected by the context in which leadership is observed and evaluated. Use of a frame of reference that is conceptually relevant to the criterion reduces within-person inconsistency (Lievens, Corte, & Schol-

laert, 2008). The 14 WOPI scales are readily interpretable within the Big Five framework, as described in detail in the measures section of Study III.

1.2.2 Measures of personality disorders and the spectrum hypothesis

From the very outset of personality research, some theorists, usually clinicians, have been interested in measuring abnormal traits. This tradition is found in both psychology and psychiatry, but with a different history, taxonomies, and research instruments than in the case of the Big Five (De Fruyt et al., 2009). In many instances, these measures have been applied for clinical purposes. One of the most recent classifications of personality pathologies in clinical applications can be found in the regularly updated Diagnostic and Statistical Manual of Mental Disorders (DSM), now in its fifth edition (American Psychiatric Association, 2013). The DSM IV lists 10 different personality disorders. These disorders are characterized as persistent, inflexible, and maladaptive ways of relating to oneself and one's environment (APA, 2000). Unlike dimensional measures of personality, such as self-report trait questionnaires, the DSM has historically considered mental disorders to be qualitatively distinct conditions, diagnosed through a psychiatric interview. As the text-revised version of the manual puts it, the DSM "divides mental disorders into types based on criteria sets with defining features" (American Psychiatric Association, 2000, p. xxxi).

Despite of their theoretical and diagnostic roots, several DSM-based questionnaires have increased in popularity outside of clinical applications (e.g., Hogan & Hogan, 2001a, b; De Fruyt et al., 2009). Researchers who have looked at the theoretical and psychometric relationship between "normal" (universal) and "abnormal" (psychopathological) measures have suggested that many DSM personality disorders are, in fact, extreme poles of normal personality, dimensional rather than qualitative in their nature. According to this notion, sometimes called the spectrum hypothesis, personality traits constitute a continuum from normal personality to disorders, and there is no qualitative distinction between the two concepts (Bagby, Costa, Widiger, Ryder, & Marshall, 2005; Samuel & Widiger, 2008; Widiger & Trull, 2007). Widiger and Trull (2007, p. 71) propose this integrative goal as follows:

> It may be time to consider a shift to a dimensional classification of personality disorder that would help address the failures of the existing diagnostic categories as well as contribute to an integration of the psychiatric diagnostic manual with psychology's research on general personality structure.

Some research projects have concentrated on the relationships between personality disorders, leadership and organizational outcomes (Furnham & Taylor, 2004; Hogan & Hogan, 2001a; Kets de Vries, 2006; Ullrich, Farrington, & Coid, 2007). The question of "what makes a leader successful" has been replaced with "what kind of traits cause leaders to fail" (Babiak & Hare, 2006; Furnham, 2010; Hogan & Hogan, 2001a). These studies have suggested that dark-side tendencies are not only useful in screening out individuals, but might also be helpful

in improving incumbents' functioning during their development (cf. De Fruyt et al., 2009) in the same way as normal personality measures. This dimensional approach is compatible with the idea of measuring personality with self-report questionnaires.

One important goal in Study II was to establish the conceptual relationships between the Big Five and the DSM framework by utilizing expert ratings of personality disorders in a personnel selection sample. The choice of a personnel selection sample was considered to be especially appropriate for validation purposes because the new types of DSM-based questionnaires are applied largely in the industrial and organizational (I/O) psychology context.

1.2.3 Leaders' personality: the incremental value of other ratings

Previous research has shown great interest in identifying the personality traits of a good leader (Judge et al., 2002). A close connection has also been found between leaders' self-reported personality and leadership performance (for a meta-analysis, see Judge et al., 2002). Recently, however, increased interest has also been shown in observers' (subordinates', colleagues' or peers') perceptions of their leaders' personality (e.g., Connelly & Ones, 2010). This is consistent with studies (e.g., Oh, Wang, & Mount, 2011) indicating that the predictive validities of certain personality traits are higher when other-ratings of personality are used. In addition to adopting self-assessment in research on a leader's personality, Connelly and Ones (2010) have proposed that the perspective of others should be considered an important source of added value when performing a personality assessment. Moreover, it seems that taking the congruence of self-reports and other-ratings into account can further contribute to the prediction of organizational outcomes. Consequently, many studies have turned to exploring the practical implications of self-other agreement (cf. Connelly & Ones, 2010).

The previous self-other agreement research has mostly concentrated on leader performance ratings and its relationships to behavioral outcomes (Fleenor, Smither, Atwater, Braddy, & Sturm, 2010). Although self-other agreement in relation to personality (i.e., rating congruence) is a widely-explored topic, the number of studies on agreement on personality ratings and its relations to behavioral outcomes is limited (Colbert, Judge, Choi, & Wang, 2012; Connelly & Ones, 2010).

Study III examined agreement in self- and other-ratings of personality and its relation to subordinates' occupational well-being. To achieve these objectives, WOPI (Leung & Zedeck, 2016; Nederström & Niitamo, 2010), was utilized. This measure has been developed specifically for work-life settings.

1.3 Self-other agreement on personality

Perception of other people has been central in the tradition of social cognition since the famous studies by Asch in the 1940's (Kihlstrom & Haste, 1997). An

even older tradition of studying peer-judgments can be traced to the reports published by Estes in the 1930's (Estes, 1938). Estes asked subjects to judge stimulus persons on film, and assessed the accuracy of their judgements by comparing these to the judgements made by trained clinicians. The studies showed that subjects were able to predict the performance of themselves, friends and strangers on various tests of intelligence, personality, and artistic tendency. Even in cultures which place a greater emphasis on relationships than on traits, people make judgments on their own and other peoples' personality traits (McCrae & Terracciano, 2005; Triandis & Suh, 2002).

The recent research generally concludes that personality judgments can show considerable agreement, even when formed on the basis of relatively thin information, when the congruence of target self-reports and external judge ratings is adopted as a criterion (Allik et al., 2016; Biesanz, et al., 2011; Connelly & Ones, 2010; Funder, 1999; John & Robins, 1993).

Attention has been directed to exploring the specific conditions under which agreement may be higher or lower. These conditions have sometimes been labeled moderators (Funder, 1999) or determinants (John & Robins, 1993) of self-other agreement. Previous studies have suggested several moderators of agreement, such as the trait being judged, target or judge gender, level of acquaintanceship between the target and the judge, and the context of judgment (Funder, 2012; Funder & Dobroth, 1987; John & Robins, 1993; Paunonen, 1989). Barrick, Patton and Haugland (2000) examined the moderators of agreement in the recruitment interview by focusing on the moderating effects of the interview design (structure and content).

1.3.1 Models and moderators of self-other agreement

Several theoretical models [e.g., the PERSON (personality, error, residual, stereotype, opinion, and norm) and the WAM (weighted-average model) of Kenny, 2004; Brunswik's lens model (Zebrowitz & Collins, 1997); the SOKA (self-other knowledge asymmetry) of Vazire, 2010], and many empirical studies have proposed physical mechanisms mediating agreement in judgments and information flow and the moderators enabling the agreement Several potential moderators can be theoretically inferred from David Funder's Realistic Accuracy Model (RAM; Funder, 1999). The RAM is a synthesis, which aims to bring together the theoretical roots of the cognitive, personality and social psychology of person perception (Funder, 1999). The RAM explains the theoretically necessary (but not sufficient) conditions for accurate personality judgments. The judge must observe trait-relevant cues and appropriately assemble the cues to form an impression of the target. Accurate judgments can be formed when these conditions of the process are met. These conditions enable several moderators of agreement to be inferred. With respect to the RAM, these moderators include factors relating to the amount of information required, the target, the trait, and the judge.

The fundamental question concerning the amount of information is this: what are the accuracy-mediating mechanisms in minimal information condi-

tions? Several mediating physical mechanisms have been suggested to explain the relatively high accuracy achieved in the absence of previous acquaintanceship (Albright, Kenny, & Malloy, 1988; Borkenau & Liebler, 1993; Kenny, 2004). On the very first encounter with a stranger, one is instantly presented with much information about her, such as sex, race, physical attractiveness and style of dress. After this, tone of voice, behavior, and several other cues come into play (Albright et al., 1988; Ambady & Rosenthal, 1992).

Some people are better targets for personality judgment: they may be somehow easier to judge than others. They seem to be relatively easy to figure out and more predictable, even from very few observations of their behavior. Thus, targets' own personality might act as a factor reducing or increasing agreement; some people are more open and consistent in their interaction and behavior and hence easier targets to judge. It seems that active and extraverted people tend to express their personality freely, whereas passive or shy people, in particular, do not give enough cues for the observer to form a valid judgement (Funder, 1999).

The flip side of a good target is the problem of a good judge. The question of a judge's ability to rate personality had already arisen in the 1920's, but thus far little research has been published on accuracy differences, and the findings have been very tentative (Funder, 1999; Schmid Mast, Bangerter, Bulliard, & Aerni, 2011; Vogt & Colvin, 2003). Funder (1999, p. 142) states that "despite the research attention it has received, the good judge is the potential moderator concerning which the accuracy literature has the scarcest data and fewest firm findings to report." However, some studies have proposed that a judge's individual accessibility to a certain trait would probably increase the accuracy of judging it (Funder & Colvin, 1997). This phenomenon has been described as a kind of chronic activation, which helps some individuals to "tune in" to certain information cues, such as intelligence, physical attractiveness, athletic ability (Kihlstrom & Haste, 1997) or cues related to a global personality profile of the target (Schmid Mast et al., 2011).

Finally, some personality traits are considered more visible and easier to judge than others. The traits connected to social behavior, such as sociability and responsibility, seem to be relatively easy to observe and understand accurately (Funder, 1999). These socially-weighted traits are, almost by definition, the most visible, understandable, and available even to the occasional observer, whereas other traits, such as neuroticism, are almost impossible to observe during a short period. This variation in trait visibility has been demonstrated in many studies (Albright, Kenny, & Malloy, 1988; Borkenau & Liebler, 1992; Connelly & Ones, 2010; Funder & Colvin, 1988).

Previous research suggests that personality traits related to extraversion are most likely to become visible in overt social interaction, while neurotic traits are usually considered more covert dimensions (Funder, 1999, 2012; Funder & Colvin, 1988; John & Robins, 1993; Zebrowitz & Collins, 1997). In line with the numerous self-other agreement works on the Big Five, Paunonen (1989) suggested that the most observable motivational traits of the Personality Research

Form (PRF; Jackson, 1999) framework are harm avoidance, play, achievement, nurturance, and sentience. From the perspective of the RAM, these traits are the ones that are closely connected to the cues that are usually the most relevant and easily available for the judge (Funder, 1999).

1.3.2 Context-specific moderators

In addition to the pre-conditions of the RAM, several studies have empirically explored the moderators of self-other agreement. These are not in contradiction to the broader RAM suggestions but instead are rather nuanced and context-specific moderators, and their effect on agreement may depend on the assessment context. Furthermore, most of the prior findings on these moderators remain tentative (Allik et al., 2016).

Despite visibility *an sich*, some traits are considered more valuable than others, and this social value can be emphasized in the assessment context. Unlike social desirability (e.g., Paulhus & John, 1998), which is a target-level individual difference in responding style, trait evaluativeness (or desirability of a trait) is a trait-level construct (Funder, 1995). Several studies have demonstrated differences in trait evaluativeness (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Feldt & Honkaniemi 2008; John & Robins, 1993). In real-life settings, applicants across all job types scored significantly higher than non-applicants, particularly on emotional stability and conscientiousness (Birkeland et al., 2006). John and Robins (1993) suggested agreeableness to be the most evaluative trait. Because assessment center targets may try to emphasize behavior related to desirable traits and suppress less desirable trait behaviors, judges are likely to observe fewer genuine cues for the most evaluative traits (Funder, 1995), such as conscientiousness, agreeableness and emotional stability.

Potential moderators can be found among demographic factors related to targets, such as the gender and age of the target. Some self-other agreement studies have suggested that, on average, women are easier to judge than men (Funder, 1999), but so far the results are tentative (Little & Perrett, 2007). One hypothesis accounting for this difference relates to targets' expressed behavior: women are considered to be more nonverbally expressive (Buck, Miller, & Caul, 1974; Funder, 1999; Hall, 1984). On the other hand, the expression of behavioral aggression in interaction is culturally more tolerated with male than female targets (Eagly & Steffen, 1986), which may lead to higher self-other agreement on this particular PRF trait with male targets. Taken together, these findings have not been consistent, and the gender difference in judgability has not been studied among individuals applying for a job (see, however, Furnham, Jensen, & Crump, 2008).

Targets' age might also act as a moderator of agreement. One reason for this may be that people's facial expressions, over time, form lasting features on their faces, increasing the judgability of older targets (Malatesta, Fiore, & Messina, 1987). It is also plausible to assume that a young person is more difficult to judge than an older person, whose personality is no longer likely to change (Allik et al., 2016). However, these hypotheses have been tentative and the research

results mostly unpublished (e.g., Hubbard, 1994). The existing research on this topic has not been either systematic or well-integrated, and thus far there no comprehensive results have been reported on the effect of target age (cf. Allik et al., 2016; Borkenau, et al., 2004; Little & Perrett, 2007); hence this demographic moderator clearly merits further investigation.

Finally, one target moderator can be identified in the judgment context from its effect on certain targets, i.e., social desirability, which can be defined as a positive self-rating bias. Social desirability can be further decomposed into a two factors: self-deception, which is the unconscious tendency of a person to see him or herself in a favorable light, and impression management, which is a deliberate distortion of self-presentation (Paulhus & Reid, 1991). In personnel selection, some targets' behaviors and response styles may be based rather on self-deception or impression management than natural manners (Birkeland et al. 2006; Hogan, Hogan, & Roberts, 1996; Honkaniemi & Feldt, 2008). While aforementioned trait evaluativeness is a trait-level construct, social desirability is a target-level individual difference in responding style (e.g., Paulhus & John, 1998). Despite of the importance of this moderator, research appears to have seldom been conducted in a selection assessment center (however, see Barrick et al., 2000; Blackman, 2002; Blackman & Funder, 2002; Schmid Mast et al., 2011).

1.3.3 Self-other agreement and leadership outcomes

Self-other agreement of judgments has many practical implications in organizational settings. For example, predicting leadership behavior and job performance has recently grown in importance (e.g., Connelly & Ones, 2010; Church, 1997; Fleenor et al., 2010). Self-other agreement research has demonstrated a relationship between managerial performance and congruence in self-other behavior ratings (for a review of the topic, see Fleenor et al., 2010).

Church (1997) suggested that one reason for the predictive validity of self-other agreement can be found in the concept of self-awareness. Atwater and Yammarino (1997) proposed that self-aware individuals (i.e., individuals who see themselves as others see them), owing to their competence in self-observation and self-reflection, provide more accurate self-ratings. The ability to self-observe one's behavior or performance and compare these observations to feedback given, for example, from colleagues, can be understood as self-awareness (Wicklund, 1975). For this reason, self-other agreement is viewed as an important way of operationalizing self-awareness (Fleenor et al., 2010).

One potential theoretical framework for connecting self-other agreement and subordinates' well-being can be found in the concept of authentic leadership. Recent definitions of authenticity include four core components: self-awareness, relational transparency, internalized moral perspective and balanced processing (Walumbwa, Avolio, Gardner, Wernsing, & Peterson, 2008). Furthermore, it has been suggested that these four components influence the well-being of both leader and follower, while "self-awareness includes being aware of one's strengths and weaknesses as well as understanding one's emotions and personality" (Ilies et al., 2005, p. 378). Therefore, it can be hypothe-

sized that self-other agreement is related to one of the core components of authentic leadership and, consequently, to well-being.

However, to the best of my knowledge, no research has yet been published on the links between agreement on leader personality and the occupational well-being of subordinates. Building on the idea that agreement on personality can predict leadership outcomes, I also proposed, in Study III, that the direction of the discrepancy between ratings is likely to contribute an important perspective on leadership behavior. As Nichols and Cottrell (2014, p.711) demonstrate in their study, "leaders greatly affect many important employee outcomes, and discrepancies between the traits subordinates desire and the traits superiors possess may ultimately lead to negative organizational outcomes".

1.4 Research questions and hypotheses

1.4.1 Study I

My main aim was to examine self-other agreement of personality judgments in a personnel selection process and what demographic and personality factors moderate this agreement. Despite the importance of these moderators (Funder, 1999), research has seldom been conducted in a selection assessment center (however, see Barrick et al., 2000; Blackman, 2002; Blackman & Funder, 2002; Schmid Mast et al., 2011). Furthermore, despite the practical importance of social desirability in selection assessment (Honkaniemi & Felt, 2008), its effect on self-other agreement has not previously been explored in this particular context. In this research, I focused on the general magnitude of self-other agreement and its five potential moderators: the trait being judged, evaluativeness of the trait, the effect of social desirability, and the gender and age of the target. Six research questions were examined:

1. What degree of self-other agreement is achieved in job interviews, when the judges are experts in assessing personality? I hypothesized that expert judges, when the targets are strangers, would yield higher mean agreement coefficients than reported in previous studies (Connelly & Ones, 2010; John & Robins, 1993; Paunonen, 1989).

2. What are the most judgable traits? I hypothesize that the most judgable traits will be related to extraversion and most unjudgable traits to neuroticism, as demonstrated in previous studies (Barrick et al., 2000; Connelly & Ones, 2010; Funder, 2012).

3. Does the evaluativeness of a trait moderate agreement? John and Robins (1993) demonstrated that self-other agreement increases as traits become more neutral and decreases with more undesirable/desirable and less neutral traits, which is consistent with the RAM model (Funder, 1999). I hypothesized that this tendency will decrease self-other agreement for highly evaluated traits (conscientiousness, agreeableness, neuroticism) and increase agreement for

more neutral traits (extraversion, openness), as earlier research has suggested (Connelly & Ones, 2010; Funder, 1995).

4. Does the socially desirable response style of a target moderate agreement? The effect of desirability was considered particularly relevant due to the job-seeking context in which the assessments were conducted. As studies accounting for the effects of social desirability in personality judgment tasks have been scarce, no hypothesis was formulated.

5. Does the gender of a target act as a moderator for agreement? I hypothesize that there will be a significant difference between female and male targets. On average, higher mean agreement, except on aggression, will be reached for female targets. Higher agreement on aggression will be reached with male subjects, since the expression of behavioral aggression in interaction is culturally tolerated in males (Eagly & Steffen, 1986).

6. Is the target's age a moderator in the judgment situation? One reason for this may be that people's facial expressions form lasting features on their faces, increasing the judgability of older targets (Malatesta, Fiore, & Messina, 1987). Furthermore, it is plausible to assume that a young person is more difficult to judge than an older person, whose personality is no longer likely to change (Allik et al., 2016). However, studies on this potential moderator of self-other agreement in organizational settings are virtually non-existent. Therefore, no hypothesis was formulated.

1.4.2 Study II

My main purpose was to study the relationships between the Big Five and the DSM personality disorders in a psychological assessment center with real-life job applicants. I wanted to extend the research methodology beyond correlational self-report studies. This was done by using expert psychologists' observations as an additional criterion. Using self-reports alone may artificially inflate the trait-trait relationships because of common method variance, that is, the proportion of the variance that is attributable to the measurement method used (Doty & Glick, 1998). One way of overcoming this problem is to compare self-report personality scales to external judgments of personality (cf. Funder, 1999; Funder & Colvin 1988; John & Robins, 1993).

I utilized two different self-report inventories and expert ratings of the DSM-IV disorders as an external criterion. Hence one important methodological objective in this study was to examine both convergent and divergent correlations between external expert ratings and self-reports of personality. Thus, three different research questions were examined:

1. How strong are the associations between two self-report inventories with different theoretical backgrounds: a Big Five self-report inventory and a DSM self-report inventory?

2. What degree of overall self-other agreement can be achieved between expert ratings of personality disorders and the DSM inventory? Will the convergent correlations between self-reports and expert ratings exceed the divergent correlations? I hypothesized that the agreement coefficients will be similar

or higher than have been reported in previous self-other agreement studies with strangers as targets. I also hypothesized that the median coefficients of the convergent correlations will exceed those of the divergent correlations, thus demonstrating the construct validity of the DSM inventory.

3. What degree of self-other agreement is achieved between expert ratings of personality disorders and the self-reported Big Five inventory? Which of the two inventories will predict the expert ratings better, the DSM or the Big Five questionnaire? Primarily, in the spirit of the dimensional model, I hypothesized that the expert ratings of the DSM disorders would be equally accounted for by both self-reports, regardless of the theoretical background of the inventory. In other words, the Big Five inventory scores will predict the expert observations to a similar order of magnitude as the dedicated DSM inventory.

However, it should be emphasized that the present study was conducted with normal adult working population. No comprehensive clinical interviews were made. Hence, the term "disorder" is used rather in statistical and dimensional than in diagnostical sense.

1.4.3 Study III

Atwater and Yammarino (1997) posit that managers whose self-ratings are in alignment with others' ratings are more likely to be linked to more positive individual and organizational outcomes, such as the occupational well-being of subordinates. To the best of my knowledge, no research has yet been published on the links between rating agreement on leader personality and the occupational well-being of subordinates. I studied subordinates' occupational well-being and its relationships to the overall magnitude of self-other agreement, the incremental validity of agreement above and beyond managers' self-reported personality, and the direction of the discrepancy, when self- and other-ratings are not in agreement (overestimators vs. underestimators). These themes were examined with three research questions:

1. Is self-other agreement in personality ratings associated with subordinate occupational well-being? I hypothesized that the overall magnitude of agreement is associated with the occupational well-being of subordinates, as predicted by the theory of authentic leadership (Walumbwa et al., 2008).

2. Does self-other agreement in ratings explain more of the variance in subordinate occupational well-being than managers' self-reported personality alone? Some studies (e.g., Morgeson, Campion, Dipboye, Hollenbeck, Murphy, & Schmitt, 2007) have suggested that the modest relationships between personality and leadership outcomes may be due to the limited use of others' ratings of personality. I proposed that using self-other agreement as a predictor[1] of well-being will explain more of the variance in well-being outcomes than using managers' self-reports of personality alone.

---

[1] In this research, the term "predictor" was used in a purely statistical sense. Therefore, this term does not imply causality in either direction.

3. Does the direction of the discrepancy matter? Overestimating one's personality traits, particularly those related to leader-like behavior (Colbert et al., 2012) and empathy (Brutus, Fleenor, & McCauley, 1999), has been shown to decrease occupational well-being outcomes. Thus, I assumed that the effect of this discrepancy will be strongest on extraversion-related traits (competitive achievement, leadership, inspiration) and empathy, as these traits have been demonstrated to associate with positive other-ratings (Brutus et al., 1999; Judge & Bono, 2000; Skakon et al., 2010) and desired leadership traits (Nichols & Cottrell, 2014).

2 METHOD

2.1 Study I

2.1.1 Procedure

The data in Study I were gathered from a Finnish assessment center in 2006–2007. All the personality judgments were made during a 1-day psychological assessment during structured job selection interviews by expert psychologists. Research has shown that unstructured interview formats produce a richer quality response and non-verbal behavior from the job applicant and subsequently stronger levels of self-other agreement (Blackman, 2002); hence the restricted context was not ideal for judges. The judges were assigned to the targets randomly. The assessment day typically lasted 6–8 hours and included several personality and ability tests with supervising assistants, discussion group exercises, and a one-hour job selection interview. Interaction between the judges and targets was restricted to this job interview. The judges resorted to no test results before making their judgments. Care was taken to ensure that the judges had no previous contact with their targets.

2.1.2 Participants

The study targets comprised 139 job-seeking candidates (84 males and 55 females; sample age 23–56 years, mean 36 years, SD = 8.4 years). Most of the job applicants were seeking managerial positions; hence the emphasis in the assessment day was on managerial competencies. The judges (n = 14, seven males and seven females) were all professional psychologists with several years of working experience as assessment consultants and interviewers (5–25 years of experience; age 33–53). Their main expertise and background was in industrial/organizational psychology, and several of them had received additional clinical training. Hence, a total of 14 expert judges assessed a total of 139 targets individually (each job candidate met with only one judge). Most of the judges

rated 10 to 15 candidates, while the range was 1–18 judgments and the median 10.5 judgments per consultant judge.

2.1.3 Measures

The focus was on 13 motivational traits of the Finnish version of PRF (Niitamo, 1997), plus two neuroticism-related traits (axiety and guilt feelings). To investigate the self-reported personality, all targets filled in 15 scales of the Finnish version of the PRF Form E inventory (Jackson, 1999). The Finnish PRF was originally validated in 1997 at the Finnish Institute of Occupational Health (FIOH), and it has since been standardized and validated with over 10 000 recruitment candidates (cf. Furnham & Nederström, 2010; Honkaniemi & Feldt, 2008). The Finnish version only includes the scales with an internal consistency better than .60, otherwise the construction of the remaining scales is the same as in Jackson's version (Niitamo, 1997). Each personality scale is measured by 16 items (except anxiety, which is measured by 20 items) that the respondent is instructed to mark as either True or False. Answers to each scale were tallied to form a raw score. These scales were (reliability in the personnel selection sample is given in parentheses; Niitamo, 1997): achievement (Cronbach's alpha = .74), affiliation (.82), aggression (.73), cognitive structure (.69), defendence (.72), dominance (.86), exhibition (.85), harm-avoidance (.84), impulsivity (.82), nurturance (.72), order (.89), sentience (.74), and succorance (.70). The rest of the original scales (with reliabilities lower than a = .60; abasement, autonomy, change, desirability, endurance, infrequency, play, social recognition, understanding) are not included in the Finnish version (Niitamo, 1997). Two additional neuroticism-related trait scales were used instead (guilt feelings and anxiety alphas .72 and .90). These scales were factor-analyzed with the original scales.

The results of the factor analysis are detailed in the Finnish validation manual of PRF (Niitamo, 1997) and are available upon request from the author. In addition to content scales, PRF includes a scale for social desirability, which contains 12 items, and its reliability was .70 in a Finnish standardization sample (Niitamo, 1997). To rate the job-seekers' personalities all the judges filled an external personality rating form after the interview. The form included a short verbal description of the low and high poles of the PRF dimensions with a graphic 1–10 rating scale for each dimension. Thus, the 14 judges rated a total of 15 different personality traits of 139 targets.

2.1.4 Analyses

The statistical analyses were performed in several steps. First, Pearson's correlations between the PRF self-report and expert ratings were calculated for all targets and all personality traits. These associations were examined by using both divergent and convergent correlations, which are ways of assessing the construct validity of a measurement. Divergent validity helps to establish convergent validity by demonstrating that the construct the judges were assessing (e.g.,

anxiety) is different from other constructs that might induce similar trait-relevant cues (e.g., guilt feelings). In an ideal situation, each convergent correlation should be significant and exceed the highest divergent correlation (e.g., Bastiaansen, Rossi, & De Fruyt, 2012).

The next step was to calculate the self-other agreement correlations. Pearson coefficients were computed separately for male and female targets., Following the suggestion, based on a Monte Carlo study, by Alexander and DeShon (1994), that researchers should avoid using the moderated multiple regression (MMR) analysis when testing hypotheses regarding categorical moderators, the significance of these differences was computed with the procedure introduced by McNemar (1962).

In the third step, an estimation of evaluativeness was calculated for each PRF trait by correlating each PRF trait with the social desirability scale. Thus, the greater the magnitude of the association (either negative or positive) between the trait and social desirability, the higher was the evaluativeness of the trait. This estimation of trait evaluativeness was adopted for the present sample, as trait evaluativeness may depend on the assessment context and purpose (Birkeland et al., 2006). This procedure allowed to examine on which traits the greatest (and the lowest) social value is placed in the present job seeker sample.

In the fourth step self-other agreement was evaluated as a function of the target's age by employing moderated multiple regression analysis (see Howell, 2002). Predictor variables were entered into the equation in the following steps: self-reports, target age, and the product term "self-report x age," when using the expert ratings as an outcome variable. A moderator effect is evident if the product term adds significantly to the prediction beyond the linear contributions of the predictor variable (self-reports) and the moderator variable (target age) (Paunonen, 1989; Paunonen & Jackson, 1988). This procedure was repeated for each PRF trait.

The fifth step included a moderated multiple regression, where self-other agreement was evaluated as a function of the target's social desirability. The self-reports were entered into the regression equation, followed by the target's social desirability and the product term "self-report x social desirability." This procedure was again repeated for each PRF trait.

2.2 Study II

2.2.1 Procedure

All data were gathered from a large Finnish psychological assessment center during a one-day personnel selection process. The assessment day typically lasted 6–8 hours and included discussion group exercises, a one-hour job selection interview and several personality and ability tests administered by assistants. All the personality judgments were made in the one-day assessment during structured job selection interviews. The interaction between judges and tar-

gets was restricted to this job interview, where psychologists assessed targets' job competencies, motivation, and personality. The judges did not resort to any test results before making their judgments. It was of the essence that the judges did not know the targets previously, which is a preferred best practice in the assessment center. This was verified by the assessment center assistants, who gathered the information that could identify the candidates before the assessment day.

2.2.2 Participants

A sample (n = 229) of job-seeking candidates as targets and a sample (n = 15) of expert psychologists as judges were used in the study. As most of the job seekers were aiming at managerial positions, the emphasis in the interviews was on managerial competencies. The main job-seeker sample comprised both male (70%) and female (30%) candidates (23–56 years, mean 36 years) with varying educational backgrounds. These individuals filled in two self-report inventories: the Big Five and the DSM-based measures. A subsample (n = 103) of the 229 job seekers formed the targets in a personality judgment process. This sample was socio-demographically similar to the main sample. In addition to completing the DSM and Big Five self-reports, their personality was judged during the job interviews. The personality judgments were made by experienced personnel selection psychologists (8 men and 7 women), who interviewed the targets one at a time and filled in 103 external rating forms. The judges were all professionals, who had several years (5–25 years) of working experience as personnel selection experts and interviewers. They were specialized in industrial and organizational psychology, but they also had clinical education and training in identifying personality disorders. To sum up, 15 expert judges individually judged a total of 103 targets. Most of the judges rated 4 to 10 candidates.

2.2.3 Measures

PK5 self-report questionnaire

To investigate the job seekers' self-reported personality, all targets completed a Finnish Big Five inventory (PK5) and a DSM inventory (Stress Reaction Style, SRS). PK5 is a Finnish language version of the five-trait inventory (openness to experiences, conscientiousness, extraversion, agreeableness, neuroticism) (PK5, 2007). The neuroticism factor is coded reversely as emotional stability. PK5 contains three subscales (or facets) for each factor and 150 respondent-descriptive items, which are responded to on a 1–5 Likert scale. Some of the items are in reversed form to control for acquiescent responding. PK5 has been shown to have relatively high subscale reliabilities (alphas between .75 and .94, mean .82 in a Finnish standardization sample, see Table 1). PK5 has been normed upon responses from real-world recruitment and training samples (n = 3 644) with a fully representative work age distribution (20–59 years). The published PK5 manual reports a validation process in a Finnish population. The PK5 validation

has also been conducted with clinical samples (Saario, 2011). Construct validity has been assessed via associations with well-known personality and work behavior measures. All PK5 subscales and global factors have been extensively cross-validated with the Personality Research Form (Jackson, 1984; Niitamo, 1997), which, in turn, has been mapped with other measures of the Big Five model (Ashton et al., 1998). The Finnish PRF/PK5 cross-validation tables and sample items are available from the present author.

TABLE 1 PK5 scales and reliabilities (Cronbach's alpha)

PK-5 scale	Alpha
Extroversion: Lively	.85
Extroversion: Socially bold	.88
Extroversion: Leading	.82
Agreeableness: Friendly	.81
Agreeableness: Ingenuous	.88
Agreeableness: Trusting	.82
Conscientiousness: Responsible	.77
Conscientiousness: Systematic	.78
Conscientiousness: Prudent	.80
Emotional stability: Composed	.86
Emotional stability: Relaxed	.91
Emotional stability: Confident	.89
Openness: Open to new experiences	.75
Openness: Theoretical	.79
Openness: Sensitive	.83

SRS self-report questionnaire

The Stress Reaction Style inventory (SRS; Nederström & Furnham, 2012) was constructed to measure Axis II personality disorders of the DSM-IV in the form of a self-report questionnaire. These disorder tendencies become apparent under stress, which is a defining characteristic of many current work environments (however, it should be noted that some disorders can be fully expressed without any stress at all, e.g., antisocial or narcissistic personality disorders). The validation process started at Psycon Corporation in 2008, where this measure is used in a personnel selection process. It should therefore be emphasized that SRS was intended for use in normal adult working populations. It was not meant for clinical use in any circumstances.

The 10 scales of the SRS inventory, each with 14 items, have their content in behavioral descriptions of 10 personality disorders, although these have been reformulated to better fit the organizational context and are scored dimensionally. The questionnaire contains 152 respondent-descriptive items, which are answered on a 0–4 Likert scale. The internal consistency of the scales in the present sample was on a reasonable level, ranging from .68 to .87. These reliabili-

ties are comparable to those obtained for the Hogan Development Survey (HDS; Hogan & Hogan, 2001b). The mean scores, SDs, reliabilities and sample items of the SRS scales are listed in Table 2. These data were collected from a total of 1 697 job applicants. All the SRS scales were also cross-validated with the Personality Research Form (Jackson, 1984; Niitamo, 1997) before the present study. A detailed correlation matrix is available from the author. Theoretically, it could be argued that no one would actually admit to personality disorder symptoms when applying for a job. To investigate socially desirable responding in completing the SRS content scales, I constructed a scale for desirability as a control scale. This scale was constructed by combining item contents from the two factors of socially desirable responding (self-deception and impression management) introduced by Paulhus and Reid (1991). The desirability scale contains 12 items, and in the present sample its alpha was .84. However, I did not expect to find strong negative associations between social desirability and personality disorders, as the inability to perceive a disorder in oneself is symptomatic of many personality disorders and thus would not have much effect on the level of social desirability (American Psychiatric Association, 1994).

TABLE 2 SRS scales, reliabilities (Cronbach's alpha) and sample items

SRS scale	Alpha	Sample item
Paranoid	.77	I trust most people from the first time I meet them. (Reversed item)
Schizoid	.76	I am not that interested in what is on other people's mind.
Schitzotypal	.79	I have some very original habits and hobbies.
Antisocial	.68	Winning comes before playing by the rules, if the situation requires.
Borderline	.76	I often explode over small matters and later regret it.
Histrionic	.87	I like to turn people's attention to myself.
Narcissist	.74	My ideas have been excellent without exception.
Avoidant	.71	I carefully control my behavior in the company of others.
Dependent	.70	Even with small matters, I ask other people's opinion before taking action.
Obsessive-Compulsive	.76	Sometimes I would like to try to improve a particular job endlessly.
Social desirability	.84	I have never held prejudices against anybody.

DSM external rating form

To judge the targets' personality disorder traits, all the psychologists completed an external rating form of the DSM with a graphic 1–10 rating scale for each dimension. This rating form contained a short verbal description of the 10 personality disorder dimensions of the SRS/DSM and a 1–10 estimate of candidates' proneness to answer in a socially desirable way. Thus, the psychologists judged the targets on 10 personality disorder traits and assessed their responding style. The judges used the dimensional model to assess the disorders, without any diagnostic thresholds or cutoff points. Hence, the Axis II disorder terms

in the present study should be interpreted as disorder symptoms or traits rather than diagnostic labels.

2.2.4 Analyses

The statistical analyses were performed with SPSS 20.0 software in three steps. In the first step, I calculated Pearson's correlations between the Big Five and the personality disorders. The correlational analyses drew on both the self-reports and expert ratings. These associations were evaluated by using the global Big Five factors and facet-level constructs and examined both divergent and convergent correlations.

In the second step, I used canonical correlational analysis to explain the overall relations of the two sets of variables. In my analysis, I used pooled redundancy coefficients (Thompson, 1984) to compute the sum of all the redundancy coefficients for all the variables in a set. This procedure collects information on the overlap between two sets of variables in a comparable way. The total overlap between the two sets was computed with the total redundancy index recommended by Alpert and Peterson (1972, p. 189), who state that the researcher may ''use canonical correlation coefficients to test for the existence of overall relationships between sets of variables, but for a measure of the magnitude of relationships, redundancy may be more appropriate.''

In the third step, the findings were elaborated on by means of a hierarchical regression analysis. This allowed comparison of whether PK5 had incremental validity beyond SRS (and vice versa) in predicting the expert ratings on personality.

2.3 Study III

2.3.1 Procedure

The data were collected by a Finnish human resources consulting company (Psycon Corp.) in a Finnish technology research organization. The data collection was carried out as a part of a broader personnel survey in which personnel received an online invitation to participate. The invitation included a link to the survey questionnaire. Of the 2 780 persons working in the organization, 2 067 participated in the personnel survey, yielding a response rate of 74.4%. The participants represented different organizational levels and units. Although total sampling was used, participation in the study was voluntary.

Both managers and subordinates filled in the personnel survey, although the present report concerns only the subordinates' responses. In conjunction with the survey, the managers were asked to complete a work-related personality inventory, WOPI. The subordinates were in turn asked to evaluate their manager's personality by rating the same set of personality dimensions appearing in the WOPI inventory.

2.3.2 Participants

The target managers (n = 180) were individuals with the job titles of team leaders (80%) and technology managers (20%), all of whom had responsibility for leading their teams. More than half of the managers were males (70% males, 30% females). All the managers filled in the WOPI questionnaire online. Subordinates rated their managers individually and where a manager had more than one subordinate in his/her team (as was the case with most of the managers), the ratings were averaged across the responding subordinates for better reliability.

Team sizes varied from 1 to 16 subordinates (median 4, while 80% of the teams had 2–6 subordinates). Thus, for the 180 managers, a total of 1 951 individual personality ratings, expressed as 180 average team values, formed the other-ratings data set. An interrater agreement analysis was conducted employing intraclass correlation coefficients (ICC) to explore the reliability of the team level aggregates. The ICC2 analysis demonstrated values from .57 (mean) to .66 (median). These values are regarded as "fair" (.40–.59) or "good" (.60–.75) (Cicchetti, 1994). Thus, it seems that aggregating teams' personality judgments was statistically justified.[2]

2.3.3 Measures

Self-reported personality

The Work Personality Inventory (WOPI) is a standardized self-report questionnaire on personality with 14 scales deemed important for work settings (Leung & Zedeck, 2016; Nederström & Niitamo, 2010). The 224 respondent-descriptive item statements are answered on a dichotomous (true-false) scale, yielding a trait score range of 0 to 16. The validation study (Nederström and Niitamo, 2010) reports scale reliabilities collected from an occupational sample. All the scale reliabilities are comparable to those of established personality inventories: (KR-20/re-test reliability in parentheses): focus (.76/.71), competitive achievement) (.78/.83), leadership (.80/.89), inspiration (.74/.86), sociability (.74/.84), empathy (.75/.85), reliance (.69/.76), orientation (.77/.89), perception (.78/.80), thinking (.81/.80), decision making (.77/.91), ambiguity-change (.71/.85), optimism (.81/.85), and self-reflection (.78/.73) (see Table 3 for Cronbach's alphas and sample items). Concurrent and predictive validation has been shown in studies investigating the convergence of peer and spousal ratings of personality, and construct validity in studies on the relations between WOPI and more established personality inventories as well as miscellaneous other measures, such as values, learning styles, organizational culture, etc. (Leung & Zedeck, 2016;

---

2 Sometimes aggregating individual responses may raise questions about their dependency and the "nesting" of data (Zyphur, Zammuto, & Zhang, 2016). However, the teams in the present study were randomly selected from the total population and there were no obvious reasons to expect any nesting effects among or between the teams.

Nederström & Niitamo). Factor analysis has demonstrated that the 14 scales in the inventory are readily interpretable in terms of the Big Five framework, which helps in connecting the present findings to prior studies on self-other agreement.

TABLE 3 WOPI scales, reliabilities (Cronbach's alpha) and sample items

WOPI scale	Alpha	Sample item
Focused achievement	.85	I always want to work thoroughly, even if it wasn't necessary.
Competition	.88	I strive for top results in everything I do.
Leadership	.82	I like to give orders and get things going.
Inspiration	.81	I want to be noticed when I'm with other people.
Sociability	.88	I'd never want to miss an opportunity to be with other people.
Empathy	.82	I'm happy to put my own things aside to do someone else a favor.
Reliance	.77	I carefully sound out others' opinions before I make a decision.
Orientation	.78	I'm more interested in ideas than facts.
Perception	.80	Theories help me enormously to understand things.
Thinking	.86	I let my feelings influence my decision making a great deal.
Decision-Making	.91	I enjoy taking risks.
Ambiguity-Change	.89	I don't like unexpected situations.
Optimism	.75	I always look at the positive side of things.
Self-Reflection	.79	I can always make the right decisions, even in difficult situations.

Others' ratings

The same set of 14 work-related personality dimensions were rated by the managers' direct subordinates (n = 1 951). The subordinates rated their managers' personality on an observer-rating form with a graphic 1–10 Likert scale. For each dimension, short verbal descriptions of the behavioral markers of low- and high-scale points were given on the rating form (e.g., leadership: 1 = "Not willing to lead, withdraws, unwilling to take the initiative, soft, not actively in charge", 10 = "Willing to lead, directive, leads the way, initiator, decision-maker).

Personnel survey and occupational well-being

The occupational well-being of subordinates was assessed through the 77-item personnel survey. The survey themes ranged from evaluating one's own job-related aspects and the functioning of the team to assessing the management style of one's own manager. All the items were answered on a five-point Likert scale where the opposite ends of the scales were described verbally (e.g., "Team spirit in our team is 1 = bad … 5 = good"; "My workload is 1 = excessive … 5 = reasonable"; "I feel enthusiasm and joy at work 1 = very rarely … 5 = almost always"). The respondents were instructed to use the whole range of the response scale. The answers to the 77 questions were analyzed with explorative

factor analysis using Principle Axis Factoring as an extraction method and Promax with Kaiser Normalization rotation in order to explore the underlying factors. For the present purposes, only the theoretically relevant factors were examined. This analysis helped to construct three reliable, interpretable and relatively independent (correlation between variables r < .49**) outcome scales for occupational well-being: 1) satisfaction with one's direct superior and team spirit (superior support, sample item: "Superior's time to listen to subordinates"); 2) trust in the management, sharing information and involving employees in decision-making (trust in management, sample item: "Trust in the management of the function"); and 3) satisfaction with one's work and opportunities for development (job satisfaction, sample item: "Enjoyment of one's work"). The internal consistencies (Cronbach's alpha) of the scales were: superior support .94 (12 items), trust in management .92 (10 items) and job satisfaction .88 (6 items).

2.3.4 Analyses

I used polynomial regression analysis and response surface tests to answer the research questions on agreement and its outcomes. This is considered the best method of analyzing data of this kind as it overcomes methodological problems (e.g., those related to using difference scores) found in earlier research on agreement (Fleenor et al., 2010). The analytical procedure, then, consisted of polynomial regression analysis followed by response surface analysis, which provides three-dimensional graphs of the results of the polynomial regression analysis (e.g., Edwards, 1994; Edwards & Parry, 1993). By using this procedure, one can investigate how the two predictor variables (self- and other-ratings of personality) and in particular the discrepancy between them, relate to an outcome variable (the occupational well-being of subordinates) (Shanock, Baran, Gentry, Pattison, & Heggestad, 2010). As the two ratings are kept separate, the computing of the higher-order terms makes it possible to examine linear and non-liner relations. Because the predictor variables were not measured on the same scale (WOPI self-reports 0–16 vs. other-ratings 1–10), I transformed the variables to a standardized scale, thereby placing them on a common metric (Shanock et al., 2010).

Polynomial regression

In polynomial regression analysis, the dependent variable (Z) is regressed on each independent variable (X and Y), the interaction between the independent variables (X × Y) and the squared terms for the independent variables (X^2 and Y^2). Thus, three new variables were created for each of the present analyses: the square of the subordinates' rating, the cross-product of the subordinates' and manager's rating, and the square of the manager' rating. Next, polynomial regression analyses were conducted by regressing the dependent variables on the independent variables, the product of the independent variables, the subordinate rating squared, and the manager rating squared terms.

Response surface analysis

In response surface analysis, the slope and the curvature of two lines graphically illustrate the phenomenon (Shanock et al., 2010). First, the slope of the line of perfect agreement (X = Y) illustrates how agreement between the independent variables (subordinates' rating and manager's rating) relates to the dependent variable (i.e., occupational well-being of subordinates). In addition, the curvature along the line of perfect agreement shows whether or not the relationship between the in-agreement ratings and the dependent variable is nonlinear. The line of incongruence (X = -Y) represents a situation where X and Y are not in agreement. A significant curvature along the line shows how the degree of discrepancy between the independent variables relates to the dependent variable. Furthermore, the slope along the line of incongruence indicates the direction of the discrepancy, which shows that the dependent variable can be more strongly influenced when the discrepancy is in a given direction (X > Y or X < Y). This helps to compare discrepancies between two variables in predicting outcomes and the degree to which one type of discrepancy (X > Y) better predicts that outcome than another type (Y > X). I examined all the combinations with three response surface test values (a^1, a^3, a^4).

The results allowed three questions to be formulated: How does the agreement between the independent variables (self- and other-ratings) relate to the dependent variable (occupational well-being)?; How does the degree of discrepancy between the independent variables relate to the dependent variable?; How does the direction of the discrepancy (self-rating > other-rating or self-rating < other-rating) between the independent variables relate to the dependent variable (Shanock et al., 2010)?

Figure 2 summarizes the participants, judging contexts, and measures across all studies.

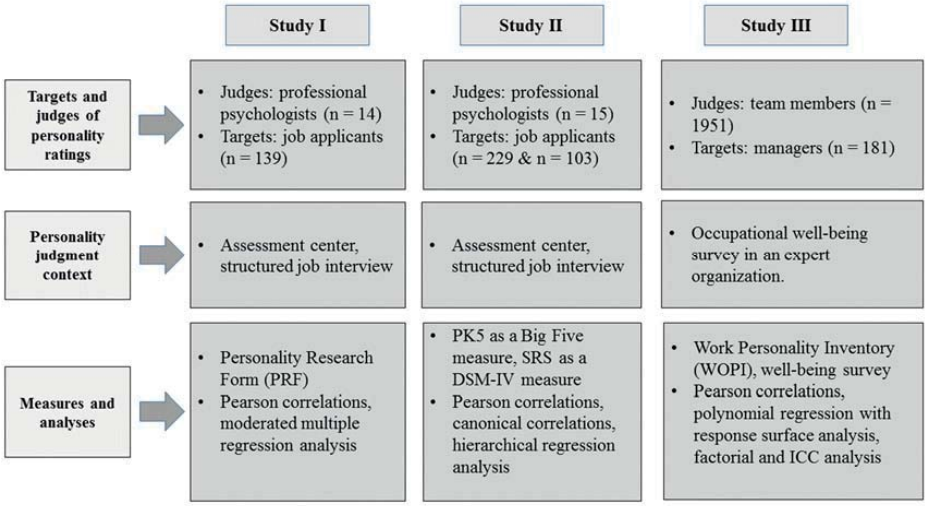


FIGURE 2 Summary of the participants, judgment context and measures in Studies I-III.

3 SUMMARY OF THE RESULTS

3.1 Study I

The study examined self-other agreement of personality judgments in a personnel selection process. I concentrated on the general magnitude of self-other agreement between judges and targets and five potential moderators of agreement: the trait being judged, the evaluativeness of the trait, social desirability, and the gender and age of the target, and examined these relationships through six research questions. The results revealed that the general magnitude of self-other agreement was in line that found in previous studies (Connelly & Ones, 2010; Paunonen, 1989). The mean and median self-other trait correlations in the present study were .25. At the trait level, 12 of the 15 dimensions judged yielded significant self-other correlations.

The results also indicated that some of the hypothesized moderators affected the judgment situation. On the one hand, when the socially desirable responding style of the interviewees was studied, the moderator results suggested that self-other agreement was negatively related to social desirability on two traits, dominance and defendence: higher social desirability in the questionnaire responses led to lower self-other agreement on these traits. On the other hand, I found no systematic relationships between the level of agreement and the evaluativeness of the other traits. This finding seems to suggest that self-other agreement is less vulnerable to the moderator effects of trait evaluativeness than individual variability in social desirability. In other words, social desirability was a more important factor than trait evaluativeness in the current judgment context. Target age and gender were significant moderators of agreement on certain traits. On anxiety, the moderating effect of age was positive: age increased the magnitude of the relationship between the self-reports and expert ratings. In line with my hypothesis, I also found one substantial gender difference, which was higher agreement on aggression in the male targets.

3.2 Study II

The study investigated the relationships between the Big Five and the DSM personality disorder traits in a psychological assessment center with real-life job applicants. My aim was to extend the research methodology beyond correlational self-report studies, and hence I employed the self-other agreement procedure to explore the conceptual relationships between the Big Five and DSM frameworks. This was done by using the expert psychologists' observations as an additional criterion. Furthermore, my aim was to validate the new DSM-based SRS measure, which maps the disorder traits by means of a self-report questionnaire.

The self-report relationships demonstrated considerable overlap between the Big Five trait measures of normal and the DSM measures of psychopathological personality, as hypothesized. In addition to the relationships between the self-report measures, the self-other agreement between the interviewers and job applicants yielded substantial correlations on the DSM personality disorder traits. The overall magnitude of the agreement was generally on the same level as the coefficients found, using normal personality measures, between close acquaintances. When the correlations of the expert-rated disorder traits were calculated, first with the PK5 and then with the SRS self-reports, the predictive power of PK5 appeared to be comparable to that of SRS: All the expert-rated disorders seemed to have at least one substantial association ($r > .40$) with the PK5 facets. This supports the general hypothesis of that a strong relationship exists between the Big Five and the DSM models. The results also support the validity of the SRS questionnaire, demonstrating its convergent and divergent validities.

3.3 Study III

The study investigated subordinates' occupational well-being and its relationships to the overall magnitude of the rating agreement on leader personality, the incremental validity of agreement above and beyond the managers' self-reported personality, and the direction of the discrepancy when self- and other-ratings are not in agreement (overestimators vs. underestimators). The main results of the present study seemed to be generally consistent with the previous findings on rating agreement and its beneficial effects on leadership outcomes (e.g., Fleenor et al., 2010). Superior support was significantly predicted by self-other agreement on seven traits (competitive achievement, leadership, inspiration, thinking, ambiguity-change, optimism and sociability), trust in management on two traits (competitive achievement, inspiration) and job satisfaction on three traits (competitive achievement, perception, sociability). This indicates the strong predictive value of self-other agreement compared to self-reports on several well-being variables. In addition, also as hypothesized, the surface re-

sponse analysis revealed that as the discrepancy between self-reports and other ratings increased, well-being decreased.

The results also suggested that overestimators of the Work Personality Inventory dimensions are more fatal for the well-being of their subordinates than underestimators. This effect appeared particularly for one facet of extroversion (inspiration) and for empathy. Other facets of extraversion were also significant, as predicted, but the effect sizes were considerably lower than those for inspiration and empathy. Openness to experience and conscientiousness revealed similar effects, although the effect sizes were also lower.

4 DISCUSSION

The main purpose of this research was to examine the extent to which self-ratings and other-ratings of personality are in agreement, what kinds of variables moderate their relationship, and what organizational outcomes result from such agreement. All three studies demonstrated that self-other agreement of personality ratings can reach congruence in various organizational contexts regardless of the measures employed. Furthermore, the results demonstrated that this congruence is affected by situational and demographic moderators, and has real-life implications for leadership outcomes.

Study I examined the general level of self-other agreement and its moderators in a personnel selection context. Nine of the 15 traits exceeded the mean agreement coefficient reported by Paunonen (1989). Yet the mean coefficient was substantially lower than that reported by Connelly and Ones (2010), which is probably due to the large proportion of close acquaintances in their rater sample (e.g., family members, friends and cohabitants). A particularly notable exception compared to previous studies, and contrary to my hypothesis, was the high agreement found on the anxiety scale, which is a subscale of neuroticism (or the low pole of emotional stability). Neuroticism is usually regarded as a relatively covert trait for strangers (Connelly & Ones, 2010). It is possible that a stressful job-seeking situation arouses maximal variance between the candidates being interviewed. In more relaxed situations, this transient activation of high anxiety would probably not take place, and this individual variation would remain hidden (see also Allik et al., 2016). This explanation is consistent with the Realistic Accuracy Model (RAM), which suggests that the visibility of a trait may be dependent on the contextual demands, and the visibility of a trait can vary across settings (Funder, 1999). Unfortunately, there was no Personality Research Form (PRF) data that would have allowed estimation and comparison of the scale variances across different rating contexts.

I also found that social desirability and the gender and age of the target moderated self-other agreement. Target age and gender were significant moderators of agreement on certain traits. On anxiety, the moderating effect of age was positive: age increased the magnitude of the relationship between the self-

reports and expert ratings. The direction of the association was as suggested in the literature (e.g., Hubbard, 1994). This could be regarded as tentative support for the hypothesis that people's facial expressions, over time, leave lasting traces on their faces, and hence increase self-other agreement. In line with the hypothesis, I also found one substantial gender difference, which was the higher agreement on aggression in the case of the male targets. This finding is in line with aggression research (Eagly & Steffen, 1986), which has suggested that males are culturally permitted to express their aggression more freely in interaction. Otherwise, the results contrast with the hypotheses suggested by Funder (1999) and Hall (1984) that females are easier targets of personality judgment. This finding might stem from the use of expert judges, who are probably less vulnerable to gender-related rating biases and, thus, better raters regardless of the target's gender (Funder, 1999).

When the socially desirable responding style of the interviewees was studied, the moderator results suggested that self-other agreement was negatively related to social desirability on two traits, dominance and defendence. In other words, higher socially desirable responding in the questionnaire responses led to lower self-other agreement on these traits. Defendence is an undesirable trait, so it is possible that interviewees scoring high on social desirability attempt to hide this trait in the selection process (cf. Funder, 1995; Schmid Mast et al., 2011). However, socially desirable responding had no effect on the correlations of self-other agreement with the other facets of neuroticism, guilt feelings and anxiety. These results seem to suggest that socially desirable responding does not systematically distort the key elements of self-other agreement. This finding is also consistent with that of Paunonen (1989), who concluded that social desirability is not a large component in the ratings either of friends or of strangers. Furthermore, I found no systematic relationships between the level of agreement and the evaluativeness (desirability) of the trait itself. Funder (1995) argued that high trait evaluativeness would endanger agreement because targets try to suppress undesirable information related to the less evaluated traits. My findings yielded no support for this hypothesis.

Study II investigated the relationships between the Big Five personality dimensions and personality disorder traits with the self-other agreement method. Thus, the aim was to extend the research methodology beyond correlational self-report studies. This was done by using expert psychologists' observations as an additional criterion. The results demonstrated considerable overlap between the Big Five measures of normal (PK5) and measures of abnormal (SRS) personality in both samples and regardless of the assessment method, which supports the spectrum hypothesis and construct validity of the SRS test. Study II also extended the previous research to the facet level of the Big Five in an organizational context.

Samuel and Widiger's (2008) meta-analysis mainly comprised student and inpatient/outpatient subsamples. Despite sample differences, many similarities were found when the direction and magnitude of the PK5 and SRS relationships were examined. As the regression results demonstrated, eight out of the

10 Axis II disorder ratings were predicted with equal or better accuracy by the Big Five self-reports than the SRS self-reports. The antisocial and schizotypal ratings were the only ones to gain significant incremental validity from the SRS self-reports. This indicates that the Axis II personality disorder traits and the "normal traits" of the Big Five are associated irrespective of the impact of the assessment sample or measure.

However, two notable exceptions were observed when the relationships between the Big Five at the facet level and the Axis II disorders were compared to previous findings. First, in the present study a substantial relationship was observed between the antisocial and emotional stability ratings, the main contributor to which was the relaxed facet of PK5. This relationship was neither found in previous empirical studies (Furnham & Crump, 2005; Samuel & Widiger, 2008), nor predicted by Widiger, Trull, Clarkin, Sanderson and Costa (2002). Yet this finding is compatible with clinical predictions, which have identified an exceptionally relaxed state between episodes of restless, anti-social behavior in psychopaths (Martens, 1997). In addition, I found a relatively strong negative relationship between schizotypal disorder and the prudent facet of PK5. This was not predicted by Widiger et al. (2002), and it was not found as clearly in the study by Furnham and Crump (2005). However, a recent meta-analysis (Samuel & Widiger, 2008), demonstrated a small but significant negative relationship between schizotypal disorder and all the facets of conscientiousness. One theoretical explanation can be derived from the different thinking styles of these two conditions: the DSM-IV describes one of the symptoms of schizotypal disorder as odd or ''magical thinking that influences behavior and is inconsistent with subcultural norms.'' Thus, this finding might be explained by the lack of common sense, order and control in the thinking styles of highly schizotypal people.

Study III investigated subordinates' occupational well-being and its relationships with the overall magnitude of the rating agreement on leader personality, the incremental validity of agreement over and beyond managers' self-reported personality, and the direction of the discrepancy. The findings extend the existing literature in some important ways. To the best of my knowledge, no previous self-other agreement research has investigated the connection between leaders' personality and subordinates' occupational well-being.

The main results of the Study III seemed to be generally consistent with the previous findings on self-other agreement and its beneficial effects on leadership outcomes (e.g., Fleenor et al., 2010). As hypothesized, the results revealed that as the discrepancy between self-reports and other ratings increased, well-being decreased. Although the evidence suggests that self-other agreement has a substantial relationship with well-being, the exact mediating process by which the effect occurs remains unclear. One explanation for the strong effect of high self-other agreement has been self-awareness. This explanation lends support to the view that authentic leadership is effective, because authenticity refers to awareness of one's strengths and weaknesses, ambiguities, inconsisten-

cies and limits of self-knowledge and acting in tune with one's true self (Luthans & Avolio, 2003).

The results demonstrated that self-other agreement can bring a significant incremental validity above and beyond self-reported personality. Colbert, Judge, Choi, and Wang (2012) suggest that there are two reasons why observer ratings can be expected to contribute to the prediction of work-related outcomes. First, self-reports and observer ratings are able to capture unique information on the target. Second, both sources of information are likely to predict unique variance in the outcomes as both personality and outcomes are based on observer ratings of trait expression.

Among the personality traits, self-other agreement on extraversion (competitive achievement, leadership, inspiration), was the most strongly related to the occupational well-being of subordinates. This outcome was not expected, and it seems that personality traits are not equal as self-other agreement predictors: agreement on extraversion seemed to be more important than agreement on the other traits when predicting organizational outcomes. It was also interesting to note that the extraversion-related traits were not the most observable traits but instead were of average magnitude. Nonetheless, they were the best self-other agreement predictor of well-being. This supports the previous findings that extraversion is an important trait in leadership behavior, emergence and leaderlikeness (Colbert et al., 2012; Offermann, Kennedy, & Wirtz, 1994) *an sich*, and this importance cannot be accounted for solely by its visibility.

It proved to be beneficial for subordinates' well-being that their assessments of their managers' work-related personality clustered towards the high pole of the assessment scale for certain traits, even if the manager's self-assessment tended towards the low pole of the assessment scale for the same traits. More specifically, when subordinates see a manager as leaderlike (facets of extraversion in WOPI) but the manager disagrees, the fact that the manager might be underestimating his/her leadership traits does not endanger the well-being of subordinates. This effect appears, in particular, for one facet of extroversion (inspiration) and for empathy, when superior support is used as a criterion. This finding suggests that underestimating one's own empathy or extraversion is less fatal for well-being outcomes than overestimating these traits: i.e., humility is better than hubris.

Taking all the findings from Studies I, II, and III into consideration, it would seem that self-other agreement of personality ratings can reach a high level of congruence in diverse organizational contexts regardless of the measures employed. Compared to previous studies, the mean effect size of self-other agreement seemed to fall somewhere between self-informant and self-stranger ratings (e.g., Paunonen, 1989; Connelly & Ones, 2010). All the personality measures utilized (PRF, PK5, SRS and WOPI) reached significant self-other agreement coefficients on several personality traits. This finding supports David Funder's RAM (Funder, 1999), which is based on a realistic and optimistic interpretation of personality judgments. Furthermore, as the findings in Study

III demonstrated, self-other agreement on personality can have significant real-life implications in manager-subordinate relationships.

4.1 Limitations

Both Studies I and II have their limitations. These mainly stem from the choice of sample population and the small number of personality judges. In Study I, the judgment context was exceptionally structured and controlled; hence, it was not possible to study the information amount or quality as a moderator of agreement, which are one important components of the RAM (Funder, 1999) and a relevant variable in a selection interview (Barrick et al., 2000).

In Study II, the validation sample consisted mainly of Finnish adults who were active in their work life. It is, therefore, unlikely that this sample population contained individuals with severe personality disorders. Additional research with different sample populations should be conducted before further generalizing the spectrum hypothesis. Second, the self-reports were based on particular questionnaires, namely PK5 and SRS. These are both well standardized and so far the most reliable self-report tests of the Big Five and DSM disorders used in Finnish psychological assessment centers. However, utilizing different inventories, even those with the same theoretical background, would probably have led to slightly different results. This incommensurability can give rise to theoretical concerns when comparing samples and results from different countries, even with more established measures. As Samuel and Widiger (2008) demonstrated, associations between the Big Five and personality disorders vary across measuring instruments.

It was not possible in either of these two studies to explore the inter-rater reliability, rating consensus or judge properties, mainly owing to the dyadic nature of the interviews and the small size of the interviewer sample (n = 14 and n = 15). The accuracy of the judges can act as a self-other agreement moderator, according to both the RAM and empirical research (Schmid Mast et al., 2011). Although the convergent effects were statistically significant in both studies (and thus indirectly supported the reliability of the expert ratings), the relatively small sample size of professional judges and the unequal number of targets rated by each judge did not allow us to explore possible differences in rating accuracy between raters. The large sample size required for moderator tests makes these effects difficult to detect. However, it is likely that the individual accuracy of the judges plays a significant role in the judgment process (the concept of "good judge", Funder, 1999).

The main limitations in Study III relate to the cross-sectional research design and the measures employed. The cross-sectional research design means that caution is needed when drawing conclusions about causality. Many previous studies (e.g., Nichols & Cottrell, 2014) have implied that high self-other agreement can be expected to lead to positive outcomes. Sosik and Megerian (1999) suggested that managerial self-other agreement moderates the relation-

ships between some aspects of emotional intelligence, transformational leadership behavior and managerial performance. Yet it is theoretically plausible that a positive relationship between a leader and a subordinate will lead to increased interaction, which, in turn, will lead to a better knowledge of the other person and, thus, higher self-other agreement. Therefore, the direction of causality and the effect of mediators should be interpreted tentatively.

In the present sample, the outcome variable (well-being of subordinates) and the other-ratings were not obtained from separate sources. Thus, it is possible that common method variance at least partially explains the strong associations found between the observer ratings and the leadership outcomes. The limitations of our sample prevented us from using relative weight analysis (Johnson, 2000) or similar analytical methods, which may have revealed the proportionate contribution of each rating type in explaining the total variance of the well-being outcomes (Connelly & Ones, 2010). It would be useful to adopt less subjective measures of well-being and performance that focus on outcomes of competent leadership behavior. This would help to obtain data from statistically independent sources.

4.2 Strengths

In both Studies I and II, the main contribution was to seek answers to the traditional research questions about self-other agreement in a real-life assessment center with expert judges and target moderators. The judges and targets were highly-educated professionals rather than students conducting a mock interview (cf. Barrick et al., 2000; Schmid Mast et al., 2011). This gives confidence in generalizing the results to an applicant population of this type and improves the applicability of the results in practice. Surprisingly few studies have considered a real-life selection context when investigating agreement on personality judgments (Blackman, 2002; Blackman & Funder, 2002; Furnham et al., 2008), despite of its practical importance in selection decisions.

The results in both studies demonstrated that expert ratings in the case of job interviews can reach significant agreement on some traits and moderate agreement on many others. Interestingly, in Study I, self-other agreement also emerged on less visible traits also of importance in predicting job performance, e.g., neuroticism (Barrick et al., 2000). This suggests that the assessment center setting may play a major role when judging certain traits (cf. Furnham et al., 2008), and that the visibility of a trait may be dependent on the demands of the assessment situation.

Study II demonstrated considerable overlap between the Big Five trait measures of normal and DSM measures of psychopathological personality. The key contribution of this study was its use of expert industrial and organizational psychologists as an external criterion to explore this overlap. It also extended previous research to an organizational context and to the facet level of the Big Five. These results are also well in line with current efforts to restructure the

description and diagnosis of personality disorders in the spirit of the dimensional model of personality disorders, as outlined in the DSM-5 (American Psychiatric Association, 2011). The results also indicate that the SRS measure has a significant amount of construct validity, when the magnitude of self-other agreement coefficients and relationship between convergent/divergent validity are used as criteria.

The findings in Study III extend the existing literature in some important ways. To the best of my knowledge, no previous self-other agreement research has investigated the connection between leaders' personality and subordinates' occupational well-being. The methods chosen for the statistical analyses can be considered a technical strength of the study. Furthermore, it revealed the incremental validity of self-other agreement above and beyond self-reports of personality. If the concept of self-awareness is to be more useful, additional research is needed to clarify the distinctions between different types of discrepancy and personality traits. For example, both over- and underestimators of extraversion are lacking self-awareness, but the organizational outcomes may be completely different, as our results demonstrated.

4.3 Practical implications

Interviewer ratings of applicant personality are useful because they are related to important outcomes at work (Barrick et al., 2000; Connelly & Ones, 2010). Study I contributes valuable nuances to discussion on the impact of moderators in the personality judgment process. From a practical perspective, it seems likely that the validity of personality judgments is not seriously threatened either by the socially desirable responding of targets or the evaluativeness of traits. Furthermore, the effect of targets' age played only a minor role in self-other agreement.

The results in Study II are potentially useful for taxonomists of personality disorders as well as psychometricians attempting to devise valid self-report measures of personality disorder traits. The SRS self-report inventory provides a new perspective on the Big Five dimensions that are already included in many selection and development assessments. In addition, the relationships between SRS and PK5 suggest one possible pattern in mapping the Big Five traits with personality disorders in a personnel assessment context. These associations can prove valuable both in making more effective personnel selection decisions and in designing conceptually elaborated leadership development programs.

The results in Study III are potentially useful for organizational developers and industrial and organizational psychologists who want to increase the subjective well-being of teams. Given that the analysis found evidence of significant relationships between self-other agreement on certain personality traits and hypothesized outcomes, I suggest that personality (in addition to 360-degree or performance measures) should be recognized as a variable with im-

portant implications for self-other agreement. Moreover, if the concept of self-awareness is to be more useful, additional research is needed to clarify the distinctions between different types of discrepancy and personality traits. For example, while both over- and underestimators of extraversion are lacking in self-awareness, their impact on organizational outcomes may be completely different, as the results demonstrated.

YHTEENVETO (SUMMARY)

Persoonallisuuden arviointi, arvioiden yhteneväisyys, arviointia moderoivat tekijät ja yhtenevyyden seuraukset

Persoonallisuusarvioiden yhteneväisyys on vanha tutkimusaihe, jota voidaan lähestyä monelta eri puolelta. Nykyisin on yleisesti hyväksyttyä, että itsearvioinnit ja ulkopuolisen tekemät persoonallisuusarvioinnit korreloivat vahvasti keskenään. Varsinkin tietyillä, helposti ulospäin näkyvillä persoonallisuudenpiirteillä, yhteneväisyys on voimakasta. Vielä ei kuitenkaan tiedetä, minkälaiset asiat ja minkälaisessa ympäristössä tämän yhteneväisyyden määrään vaikuttavat. Näitä vaikuttavia asioita kutsutaan moderaattoreiksi.

Yhteneväisyyden tutkimus nojaa vahvasti ns. realistiseen tulkintaan ihmisen persoonallisuudesta: persoonallisuus on suhteellisen pysyvä, tilanteesta toiseen ja ylitse ajan säilyvä ominaisuus, joka joiltakin osin näkyy myös ulospäin. Tämä mahdollistaa toisen ihmisen piirteiden luotettavan arvioinnin myös hyvin lyhyen tuttavuuden perusteella. Yhteneväisyyden näkökulma tarjoaa myös mahdollisuuden validoida persoonallisuustestejä muutenkin kuin niitä keskenään korreloimalla: kun mukana on riippumaton arvioija, löydetty korrelaatio itsearvioinnin ja ulkopuolisen arvioinnin välillä on vähemmän triviaali kuin pelkkien itsearviointitestien välillä.

Tämän tutkimuksen tarkoitus oli hyödyntää arviointien yhteneväisyyden tutkimista työ- ja organisaatiopsykologian keskeisissä kysymyksissä. Metodia sovellettiin persoonallisuuden arvioinnin tutkimiseen henkilöarvioinnissa sekä esimiestyössä asiantuntijaorganisaatiossa. Pitkästä tutkimusperinteestään huolimatta yhteneväisyyden tutkiminen juuri näiden kysymysten osalta on ollut suhteellisen niukkaa.

Ensimmäinen osatutkimus keskittyi selvittämään, kuinka vahvoja korrelaatioita voidaan eri piirteiden välillä saavuttaa henkilöarviointipäivän aikana, kun arviointi tehdään suhteellisen lyhyen työhaastattelun perusteella. Lisäksi tutkittiin yhteneväisyyttä moderoivia tekijöitä: piirrettä, jota arvioidaan, arvioitavan ikää, sukupuolta ja sosiaalisesti suotavaa vastaamistyyliä. Tutkimuksessa ulkopuolisen arvion muodosti kokenut henkilöarviointipsykologi ja itsearvion henkilöarviointipäivään osallistunut työnhakija PRF-testin avulla. Tulokset osoittivat, että keskimäärin yhteneväisyys itsearvioiden ja ulkopuolisten arvioiden välillä oli samaa luokkaa kuin aikaisemmissa tutkimuksissa, kun ulkopuolisen arvion tekee tuntematon arvioija. Moderaattorit vaikuttivat enimmäkseen odotetun suuntaisesti, mutta efektit olivat lieviä, eikä systemaattista eroa ei syntynyt kuin piirteen osalta: arvioitava piirre vaikutti keskimäärin kaikkein vahvimmin arvioiden yhteneväisyyteen.

Aikaisempi tutkimus on osoittanut, että varsinkin johtamiskäyttäytymisessä ns. persoonallisuuden pimeällä puolella on suuri merkitys johtamistyön onnistumisen kannalta. Toinen osatutkimus selvitti itsearvioiden ja ulkopuolisten arvioiden yhtenevyyttä tutkimalla, kuinka vahvasti persoonallisuuden pimeä puoli (persoonallisuushäiriöpiirteet) selittyy persoonallisuustutkimuksessa

jo vakiintuneiden Big Five -piirteiden avulla. Tutkimuksessa ulkopuoli-sen arvion muodosti kokenut henkilöarviointipsykologi, ja itsearvion itsestään teki henkilöarviointipäivään osallistunut työnhakija. Kliinistä diagnoosia arvioitaville ei tehty, vaan persoonallisuushäiriöpiirteiden voimakkuus arvioitiin työhaastattelun avulla. Työhaastattelussa tehtyjä arvioita puolestaan selitettiin vuorotellen sekä SRS- että PK5-persoonallisuustestien itsearviointien avulla. Tulokset osoittivat, että PK5-testi on monien piirteiden osalta lähes yhtä hyvä tai yhtä hyvä persoonallisuushäiriöpiirteiden selittäjä kuin varta vasten niitä mittaamaan kehitetty SRS-testi. Samalla saatiin uutta tietoa SRS-testin validiteetista ja soveltuvuudesta persoonallisuuden pimeän puolen kartoittamiseen henkilöarviointitilanteessa.

Kolmas osatutkimus selvitti, minkälaisia käytännön vaikutuksia persoonallisuusarviointien yhteneväisyydellä on suuressa suomalaisessa asiantuntijaorganisaatiossa. Tässä tutkimuksessa arvioitavina olivat asiantuntijaorganisaation esimiehet, ja esimiestyön onnistumisen kriteerinä oli alaisten kokema työtyytyväisyys. Esimiehet arvioivat omaa persoonallisuuttaan WOPI-testillä, ja alaiset arvioivat esimiestensä persoonallisuutta vastaavien piirteiden avulla. Tulokset osoittivat, että korkea yhteneväisyys alaisten ja esimiesten välisissä persoonallisuusarvioinneissa tuotti positiivisia arvioita myös työtyytyväisyydessä. Vastaavasti heikko yhteneväisyys persoonallisuusarvioinneissa oli yhteydessä matalampaan työtyytyväisyyteen. Tulos oli sikäli yllättävänkin vahva, että persoonallisuusarviointien yhteneväisyys selitti laajemmin alaisten työtyytyväisyyttä kuin esimiehen persoonallisuus sinänsä.

Yhteenvetona voidaan todeta, että kaikissa osatutkimuksissa persoonallisuuden itsearviointien ja ulkopuolisen arvion välillä löytyi selvä yhteys useammalla eri piirteellä. Tämä löytö tukee realistista tulkintaa persoonallisuudesta; se on suhteellisen stabiili, tilanteesta toiseen säilyvä ominaisuus, joka vaikuttaa ihmisen käyttäytymiseen ja siten näkyy ulospäin myös suhteellisen lyhyen vuorovaikutuksen aikana. Lisäksi näyttäisi siltä, että persoonallisuuden itsearvioinnin ja ulkopuolisen arvion välinen yhteneväisyys on ilmiö, joka vaikuttaa merkittävästi myös käytännön tilanteisiin (kuten esimerkiksi esimiestyön onnistumiseen, kun kriteerinä on alaisten työtyytyväisyys). Yksi selitys tähän vaikutukseen voi olla esimiehen itsetuntemus, joka tulee näkyviin persoonallisuusarvioiden yhteneväisyytenä.

REFERENCES

Albright, L., Kenny, D., & Malloy, T. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology, 55,* 387–395.

Alexander, R. A. & DeShon, R. P. (1994). Effect of error variance heterogeneity on the power of tests for regression slope differences. *Psychological Bulletin, 115,* 308–314.

Allik, J., de Vries, R. E., & Realo, A. (2016). Why are moderators of self-other agreement difficult to establish? *Journal of Research in Personality, 63*, 72–83.

Allport, G. W. (1937). *Personality: A psychological interpretation.* New York: Holt, Rinehart & Winston.

Alpert, M. I. & Peterson, R. A. (1972). On the interpretation of canonical analysis. *Journal of Marketing Research, 9,* 187–192.

Ambady, N. & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111,* 256–274.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders DSM-IV* (4th edition). Washington, DC: American Psychiatric Association.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders DSM-IV-TR* (Text Revision). Washington, DC: American Psychiatric Association.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders DSM-5* (5th edition). Washington, DC: American Psychiatric Association.

Antonakis, J., Avolio, B. J., & Sivasubramaniam, N. (2003). Context and leadership: An examination of the nine-factor full-range leadership theory using the Multifactor Leadership Questionnaire. *The Leadership Quarterly, 14*, 261–295.

Arthur, W. Jr., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, *56*, 125–154.

Ashton, M. C., Jackson, D. N., Helmes, E., & Paunonen, S. V. (1998). Joint factor analysis of the Personality Research Form and the Jackson Personality Inventory: Comparisons with the Big Five. *Journal of Research in Personality*, *32*, 243–250.

Atwater, L. E., Waldman, D. A., Robie, C., Ostroff, C., & Johnson, J. (2005). Self-other
agreement: Comparing its relationship with performance in U.S. and Europe. *International Journal of Selection and Assessment*, *13*, 25–40.

Atwater, L. & Yammarino, F. (1997). Self-other rating agreement: A review and model. *Research in Personnel and Human Resources Management*, *15*, 121–174.

Babiak, P. & Hare, R. D. (2006). *Snakes in suits. When psychopaths go to work.* NewYork: Harper Collins.

Bagby, R. M., Costa, P. T. Jr, Widiger, T. A., Ryder, A. G., & Marshall, M. (2005). DSM-IV personality disorders and the five-factor model of personality: A multi-method examination of domain- and facet-level predictions. *European Journal of Personality*, *19*, 307–324.

Barrick, M. R. & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*, 1–26.

Barrick, M. R., Patton, G. K., & Haugland, S. N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychology*, *53*, 925–951.

Bastiaansen, L., Rossi, G., & DeFruyt, F. (2012). Comparing five sets of five-factor model personality disorder counts in a heterogeneous sample of psychiatric patients. *European Journal of Personality*, *27*, 377–388.

Biesanz, J. C., Human, L. J., Paquin, A., Chan, M., Parisotto, K. L., Sarracino, J., & Gillis, R. L. (2011). Do we know when our impressions of others are valid? Evidence for realistic accuracy awareness in first impressions of personality. *Social Psychological and Personality Science*, *2*, 452–459.

Bing M. N., Whanger J. C., Davison, H. K., & VanHook J. B. (2004). Incremental validity of the frame-of-reference effect in personality scale scores: A replication and extension. *Journal of Applied Psychology*, *89*, 150–157.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14*, 317–335.

Blackman, M. C. (2002). The employment interview via the telephone: Are we sacrificing accurate personality judgments for cost efficiency? *Journal of Research in Personality*, *36*, 208–223.

Blackman, M. C. & Funder D. C. (2002). Effective interview practices for accurately assessing counterproductive traits. *International Journal of Selection and Assessment*, *10*, 109–116.

Block, J. (1965). *The challenge of response sets: Unconfounding meaning, acquiescence, and social desirability in the MMPI.* New York: Appleton-Century-Crofts.

Borkenau, P. & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, *62*, 645–657.

Borkenau, P. & Liebler, A. (1993). Consensus and self-other agreement for trait inferences from minimal information. *Journal of Personality*, *61*, 477–496.

Borkenau, P., Riemann, R., Angleitner, A., & Spinath, F. M. (2001). Genetic and environmental influences on observed personality: Evidence from the German Observational Study of Adult Twins. *Journal of Personality and Social Psychology*, *80*, 655–668.

Bouchard, T. J. Jr., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990) Sources of human psychological differences: The Minnesota Study of Twins Reared Apart. *Science*, *12*, 223–229.

Bouchard, T. J. & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology*, *54*, 4–45.

Brutus, S., Fleenor, J. W., & McCauley, C. D. (1999). Demographic and personality predictors of congruence in multi-source ratings. *Journal of Management Development*, *18*, 417–435.

Buck, R., Miller, R. E., & Caul, W. F. (1974). Sex, personality, and physiological variables in the communication of affect via facial expression. *Journal of Personality and Social Psychology*, *30*, 587–596.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

Church, A. H. (1997). Managerial self-awareness in high-performing individuals. *Journal of Applied Psychology*, *82*, 281–292.

Colbert, A. E., Judge, T. A., Choi, D., & Wang, G. (2012). Assessing the trait theory of leadership using self and observer ratings of personality: The mediating role of contributions to group success. *The Leadership Quarterly, 23*, 670–685.

Colvin, C. R. (1993). "Judgable" people: Personality, behavior, and competing explanations. *Journal of Personality and Social Psychology*, *64*, 861–873.

Connelly, B. S. & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136, 1092–1122.

Cook, M. (2009). *Personnel selection - Adding value through people* (Fifth ed.). Singapore: Wiley-Blackwell.

Costa P. T. Jr. & McCrae R. R. (2006). Trait and factor theories. In J. C. Thomas & D. L. Segal (Eds.), *Comprehensive handbook of personality and psychopathology. Volume 1. Personality and everyday functioning* (pp. 96–114). New York: Wiley.

De Fruyt, F., De Clercq, B., Miller, J., Rolland, J.-P., Jung, S.-C., Taris, R., Furnham, A., & Van Hiel, A. (2009). Assessing personality at risk in personnel selection and development. *European Journal of Personality*, *23*, 51–69.

De Fruyt, F., McCrae, R. R., Szirmak, Z., & Nagy, J. (2004). The five-factor personality inventory as a measure of the five-factor model: Belgian, American, and Hungarian comparisons with the NEO-PI-R. *Assessment, 11*, 207–215.

Doty, D. H. & Glick, W. H. (1998). Common methods bias: Does common methods variance really bias results? *Organizational Research Methods*, *1*, 374–406.

Eagly, A. H. & Steffen, V. J. (1986). Gender and aggressive behavior: A meta-analytic view of the social psychological literature. *Psychological Bulletin*, *100*, 309–330.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes, 58*, 51–100.

Edwards, J. R. & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*, 1577–1613.

Estes, S. G. (1938). Judging personality from expressive behavior. *Journal of Abnormal and Social Psychology*, *33*, 217–236.

Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self-other rating agreement in leadership: A review. *Leadership Quarterly*, *21*, 1005–1034.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, *101*, 75–90.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652–670.

Funder, D. C. (1997). *The personality puzzle*. New York: Norton.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception.* London: Academic Press.

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, *21*, 177–182.

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.

Funder, D. C. & Colvin, C.R. (1997). Congruence of others' and self-judgments of personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 617–647). San Diego, CA: Academic Press.

Funder, D. C. & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, *52*, 409–418.

Furnham, A. (2010). *The elephant in the boardroom: The psychology of leadership derailment.* Basingstoke: Palgrave MacMillan.

Furnham, A. & Crump, J. (2005). Personality traits, types, and disorders: An examination of the relationship between three self-report measures. *European Journal of Personality*, *19*, 167–184.

Furnham, A., Jensen, T., & Crump, J. (2008). Personality, intelligence and assessment centre expert ratings. *International Journal of Selection and Assessment*, *16*, 356–365.

Furnham, A. & Nederström, M. (2010). Ability, demographic and personality predictors of creativity. *Personality and Individual Differences*, *48*, 957–961.

Furnham, A. & Taylor, J. (2004). *The dark side of behaviour at work: Understanding and avoiding employees leaving, thieving and deceiving.* Basingstoke: Palgrave MacMillan.

Graziano, W. G., Biesanz, J. C., & West, S. G. (1998). Moderators of self-other agreement: Reconsidering temporal stability in personality. *Journal of Personality and Social Psychology, 75,* 467–477.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*, 1216–1229.

Hall, J. A. (1984). *Non-verbal sex differences: Communication accuracy and expressive style.* Baltimore, MD: Johns Hopkins University Press.

Haselton, M. G. & Funder, D. C. (2005). The evolution of accuracy and bias in social judgment. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 15–38). New York: Psychology Press.

Hogan, R. & Hogan, J. (2001a). Assessing leadership: A view from the dark side. *International Journal of Selection and Assessment*, *9*, 40–51.

Hogan, R. & Hogan, J. (2001b). *Hogan Development Survey manual.* Tulsa, OK: Hogan Assessment Systems.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions - Questions and answers. *American Psychologist*, *51*, 469–477.

Honkaniemi, L. & Feldt, T. (2008). Egoistic and moralistic bias in real-life inventory responses. *Personality and Individual Differences*, *45*, 307–311.

Hough, L. (1992). The "Big Five" personality variables - Construct confusion: Description versus prediction. *Human Performance*, *5*, 139–155.

Howell, D. C. (2002). *Statistical methods for psychology*. Pacific Grove, CA: Duxbury.

Hubbard, J. (1994). *What's in a face: Correlations of self, spouse, and stranger ratings of personality*. Unpublished master's thesis. Macalester College.

Ilies, R., Morgeson, F. P., & Nahrgang, J. D. (2005). Authentic leadership and eudaemonic well-being: Understanding leader-follower outcomes. *The Leadership Quarterly*, *16*, 373–394.

Jackson, D. N. (1999). *Personality Research Form manual.* Port Huron, MI: Sigma Assessment Systems.

John, O. P. & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: the Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, *61*, 521–551.

Judge, T. A. & Bono, J. E. (2000). Five-factor model of personality and transformational leadership. *Journal of Applied Psychology*, *85*, 751–765.

Judge T. A., Bono, J. E., Ilies R., & Gerhadt M. W. (2002). Personality and leadership: A qualitative and quantative review. *Journal of Applied Psychology*, *87*, 765–780.

Kenny, D.A. (2004). PERSON: A general model of interpersonal perception. *Personality & Social Psychology Review*, *8*, 265–280.

Kets de Vries, M. F. R. (2006). *The leader on the couch – A clinical approach to changing people and organizations.* Chichester: Wiley.

Kihlstrom, J. F. & Hastie, R. (1997). Mental representations of persons and personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 711–735). San Diego, CA: Academic Press.

Lavonen, T., Myyry, L., & Helkama, K. (2004). *Yksityisyyden suoja työelämässä: Selvitys henkilöarviointimenetelmien käytöstä Suomessa [Act on the Protection of Privacy in Working Life: A report on the methods and use of the personnel assessments in Finland]*. Helsinki: Työministeriö.

Leung, S. A. & Zedeck, S. (in press). Reviews of the Work Personality Inventory. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The twentieth mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.

Lievens F., De Corte W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, *93*, 268–279.

Little, A. C. & Perrett, D. I. (2007). Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology, 98*, 111–126.

Lord, W. (2007). *NEO PI-R - A guide to interpretation and feedback in a work context*. Oxford: Hogrefe.

Luthans, F. & Avolio, B. (2003). Authentic leadership: A positive development approach. In K. S. Cameron, J. E. Dutton, & R. E. Quinn (Eds.), *Positive organizational scholarship* (pp. 241–258). San Francisco: Berrett-Koehler.

Löckenhoff, C. E., Terracciano, A., Bienvenu, O. J., Patriciu, N. S, Nestadt, G., McCrae, R. R., Eaton, W. W., & Costa P. T. Jr. (2008). Ethnicity, education, and the temporal stability of personality traits in the East Baltimore Epidemiologic Catchment Area study. *Journal of Research in Personality*, *42*, 577–598.

Malatesta, C. Z., Fiore, M. J., & Messina, J. J. (1987). Affect, personality, and facial expressive characteristics of older people. *Psychology and Aging*, *2*, 64–69.

Martens, W. H. J. (1997). *Psychopathy and maturation*. Maastricht: Shaker.

McAdams, D. P. (1997). A conceptual history of personality psychology. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 3–39). San Diego, CA: Academic Press.

McCrae, R. R. & Costa, P. T. (1985). Comparison of EPI and psychoticism scales with measures of the five-factor model of personality. *Personality and Individual Differences*, *6*, 587–597.

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*, 547–561.

McNemar, Q. (1962). *Psychological statistics* (3rd ed.). New York: Wiley.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, *60*, 683–729.

Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.

Nederström M. & Niitamo P. (2010). *Construction and validation of a work personality inventory*. Helsinki University of Technology, Department of Industrial Engineering and Management, Report 2010/1.

Nichols, A. L. & Cottrell, C. A. (2014). What do people desire in their leaders? The role of leadership level on trait desirability. *The Leadership Quarterly*, *25*, 711–729.

Niitamo, P. (1997). *PRF – Personality Research Form manual*. Helsinki: Psykologien Kustannus.

Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality: A meta-analysis. *Journal of Applied Psychology*, *96*, 762–773.

Parker, K. C. H., Hunsley, J., & Hanson, R. K. (1988). MMPI, Rorschach, and WAIS: A meta-analytic comparison of reliability, stability, and validity. *Psychological Bulletin*, *103*, 367–373.

Paul, A. M. (2005). *The cult of personality testing*. New York: Free Press.

Paulhus, D. L. & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, *66*, 1025–1060.

Paulhus, D. L. & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology*, *60*, 307–317.

Paunonen, S. V. (1989). Consensus in personality judgments: Moderating effects of target-rater acquaintanceship and behavior observability. *Journal of Personality and Social Psychology*, *56*, 823–833.

Paunonen, S. V. & Jackson, D. N. (1988). Type I error rates for moderated multiple regression analysis. *Journal of Applied Psychology*, *73*, 569–573.

Paunonen, S. V. & Jackson, D. N. (2000). What is beyond the Big Five? Plenty! *Journal of Personality, 68*, 821–835.

Pervin, L. A. (1996). *The science of personality*. New York: Wiley.

PK5 (2007). *PK5-persoonallisuustestin käsikirja [PK5 personality test manual]*. Helsinki: Psykologien Kustannus.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*, 322–338.

Roberts, B. W. & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3–25.

Robins, R. W. & John, O. P. (1997). The quest for self-insight: Theory and research on accuracy and bias in self-perception. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 649–679). San Diego, CA: Academic Press.

Rorer, L. G. & Widiger, T. A. (1983). Personality structure and assessment. *Annual Review of Psychology*, *34*, 431–463.

Saario, S. (2011). The relationship between general therapeutic orientation, Big Five personality traits, and interpersonal functioning in substance abuse therapists: An explorative study. *Addictive Disorders & Their Treatment*, *10*, 29–36.

Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organisational Psychology*, *76*, 323–346.

Samuel, D. B. & Widiger, T. A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review*, *28*, 1326–1342.

Schmid Mast, M., Bangerter, A., Bulliard, C., & Aerni, G. (2011). How accurate are recruiters' first impressions of applicants in employment interviews? *International Journal of Selection and Assessment*, *19*, 198–208.

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful

approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology, 25,* 543–554.

Skakon, J., Nielsen, K., Borg, V., & Guzman, J. (2010). Are leaders' well-being, behaviours and style associated with the affective well-being of their employees? A systematic review of three decades of research. *Work & Stress, 24,* 107–139.

Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. *Journal of Personality and Social Psychology*, *100*, 330–348.

Thompson, B. (1984). *Canonical correlation analysis. Uses and interpretation*. Beverly Hills, CA: Sage.

Triandis, H. C. & Suh, E. M. (2002). Cultural influences on personality. *Annual Review of Psychology, 53*, 133–160.

Ullrich, S., Farrington, D. P., & Coid, J. W. (2007) Dimensions of DSM-IV personality disorders and life-success. *Journal of Personality Disorders*, *21*, 657–663.

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98*, 281–300.

Vogt D. S. & Colvin, C. R. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality*, *71*, 267–295.

Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T. S., & Peterson, S. J. (2008). Authentic leadership: Development and validation of a theory-based measure. *Journal of Management*, *34*, 89–126.

Wicklund, R. A. (1975). Objective self-awareness. *Advances in Experimental Social Psychology, 8,* 233–275.

Widiger, T. A. & Trull, T. J. (2007). Plate tectonics in the classification of personality disorder: Shifting to a dimensional model. *American Psychologist*, *62*, 71–83.

Zebrowitz, L. A. & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, *1*, 203–222.

Zyphur, M. J., Zammuto, R. F., & Zhang, Z. (2016). Multilevel latent polynomial regression for modeling (in)congruence across organizational groups: The case of organizational culture research. *Organizational Research Methods*, *19*, 53–79.

ORIGINAL PAPERS

I

SELF-OTHER AGREEMENT OF PERSONALITY JUDGEMENTS IN JOB INTERVIEWS: EXPLORING THE EFFECTS OF TRAIT, GENDER, AGE AND SOCIAL DESIRABILITY

by

Mikael Nederström & Katariina Salmela-Aro, 2014

Scandinavian Journal of Psychology, *55*, 520–526

Reproduced with kind permission by Wiley.

ABSTRACT

The article investigated agreement between self-reports and stranger ratings of personality. A sample of 139 real-life job applicants was interviewed by expert psychologists upon entrance to assessment center. The applicants provided self-descriptions on 15 personality factors, and the psychologists rated the same traits of each target based on their impressions in the interview. The results demonstrated that professional judges can reach a substantial self-other agreement (SOA) on several traits even when the targets are strangers, and that the trait being judged, the target's gender, age and social desirability have an effect on the level of agreement.

INTRODUCTION

Perceptions of other people's personality have been in focal point in the tradition of social cognition since the famous Solomon Asch studies in 1940's (Kihlstrom & Haste, 1997). Even the older tradition of peer judgments can be traced to Estes' reports in 1930's (Estes, 1938). Sometimes extremely subtle cues, in combination with successful cue utilization, enable accurate judgments. Some studies have called this accuracy "zero-acquaintance conditions" (Albright, Kenny, & Malloy, 1988), some refer to "minimal information" (Borkenau & Liebler, 1992), whereas Ambady and Rosenthal (1992) have described this phenomenon as judgment of personality on "thin slices of behavior." The recent research generally concludes that personality judgments can reach considerable self- and peer-judge agreement on a basis of considerably thin information when the congruence of target self-reports and external judge ratings is used as a criterion (Connelly & Ones, 2010; John & Robins, 1993; Paunonen, 1989; Zebrowitz & Collins, 1997; Funder, 1999; Biesanz, Human, Paquin, Chan, Parisotto, Sarracino, & Gillis, 2011). Furnham, Jensen & Crump (2008) demonstrated that work-related assessment center constructs

and corresponding personality traits can reach substantial and significant self-other correlations. The self-other agreement (SOA) of judgments has many practical implications in organisational settings; e.g., predicting leadership behavior and job performance have recently grown in importance (e.g., Fleenor, Smither, Atwater, Braddy, & Sturm, 2010; Connelly & Ones, 2010; Church, 1997).

When the statistical significance or importance of SOA is no longer at issue, attention has been directed to exploring the specific conditions under which agreement may be higher or lower. These conditions have sometimes been labeled as moderators (Funder, 1999) or determinants (John & Robins, 1993) of self-other agreement. Previous studies have suggested several moderators of agreement, such as the trait being judged, target or judge gender, level of acquaintanceship between the target and the judge, and judgment context (Funder, 2012; Funder & Dobroth, 1987; Funder & Colvin; 1988; Paunonen, 1989; John & Robins, 1993). Barrick et al. (2000) examined the moderators of SOA in the recruitment interview by focusing on the moderating effects of interview design (structure and content). The present study aims to investigate the effect of moderating factors in a real-life personnel selection interview by focusing on the moderating effects of target traits, gender, age and social desirability.

Moderators of SOA

One comprehensive model of person perception is the Realistic Accuracy Model (RAM) (Funder, 1999; Funder, 2012). The RAM divides the SOA moderator effects into four classes: properties of the judge, the target, the trait, and the information on which the judgment is based. The judge must observe trait-relevant cues and appropriately assemble the cues to form an impression of the target. Accurate judgments can be formed when these conditions of the process are met.

In the present study the judgment context was highly structured (and thus similar for all targets); hence the information as a moderator of agreement was not studied. Furthermore, the limited amount of professional judges prevented us from considering the properties of the judge (cf. Schmid Mast, Bangerter, Bulliard, & Aerni, 2011). Hence, questions about variability of judges' accuracy and information sources are outside of the scope of this study. Instead we concentrate on the properties of the trait and the target. We also expand previous research with taking into account some novel moderator effects.

According to the RAM, one potential moderator of SOA is the personality trait being judged (Funder, 1999; Funder, 2012). The most widely adopted framework in SOA studies is the Five Factor Model (FFM) (Connelly & Ones, 2010) or closely related constructs (Nederström & Furnham, 2012). The FFM includes five global personality traits (Intellect/Imagination, Conscientiousness, Extraversion, Agreeableness, and Neuroticism) of personality (Costa & McCrae, 1992; Costa & McCrae, 2006). Previous research suggests that personality traits related to Extraversion are most likely to become visible in overt social interaction, while neurotic traits are usually considered more covert dimensions (Funder, 2012; Connelly & Ones, 2010). Barrick, Patton, & Haugland (2000) demonstrated that recruiters were able to judge applicants' Extraversion, Openness, and Agreeableness correctly in a 30 min job interview. In addition to the numerous works on the FFM, Paunonen (1989) demonstrated that the most observable motivational traits of the Personality Research Form (PRF; Jackson, 1999) framework are Harmavoidance, Play, Achievement, Nurturance, and Sentience.

Despite of visibility *an sich*, some traits are considered more valuable than the others and this social value can be emphasized in the assessment context. While social desirability (e.g., Paulhus & John, 1998) is a target-level individual difference, trait

evaluativeness is a trait-level construct (Funder, 1995). Several studies have demonstrated differences in trait evaluativeness (Birkeland et al., 2006; John & Robins, 1993; Feldt & Honkaniemi 2008). In real-life settings, applicants across all job types scored significantly higher than non-applicants particularly on Emotional stability and Conscientiousness (Birkeland et al., 2006). John & Robins (1993) suggested Agreeableness to be the most evaluative trait. Because assessment center targets may try to emphasize behavior related to desirable traits and suppress less desirable behavior, judges are likely to observe fewer genuine cues for traits (Funder, 1995), such as Conscientiousness, Agreeableness and Neuroticism.

Potential moderators can be found from demographic factors related to target, such as the gender and age of the target. Some SOA studies have suggested that, on average, women are easier to judge than men (Funder, 1999), but so far the results are tentative (Little & Perrett, 2007). One hypothesis accounting for this difference relates to targets' expressed behavior: women are considered to be more nonverbally expressive (Hall, 1984; Funder, 1999; Buck, Miller, & Caul, 1974). On the other hand, these findings have not been consistent, and the gender difference in judgability has not been studied with individuals applying for a job (however, see Furnham et al. (2008).

Targets' age might also act as a moderator of agreement. One reason for this may be that people's facial expressions, over time, form lasting features on their faces, increasing the judgability of older targets. However, this hypothesis is premature and research mostly unpublished (e.g., Hubbard, 1994). So far there have not been any comprehensive results on the effect of the target age (cf. Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004; McCrae & Costa, 1982; Little & Perrett, 2007), so this demographic moderator is worth studying.

Finally, one target moderator can be identified from the context of judgment and its effect on certain targets. Some targets' behaviors and response styles in personnel selection can be based rather on image management than natural manners (Birkeland, Manson, Kisamore, Brannick & Smith, 2006; Hogan, Hogan & Roberts, 1996; Honkaniemi & Feldt, 2008). This response bias is present when job applicants know they are under evaluation (Barrick et al., 2000; Honkaniemi & Feldt, 2008): they try to conceal undesirable traits and highlight desirable traits in their questionnaire responses. Despite of the importance of this desirability moderator regarding SOA, research has seldom been conducted, as far as the authors know, in a selection assessment center (however, see Schmid Mast et al., 2011; Barrick et al., 2000; Blackman & Funder, 2002; Blackman, 2002).

Aims and hypotheses of this study

Our main aim was to examine SOA of personality judgments in a personnel selection process. What demographic and personality factors affect agreement? We concentrated on the general magnitude of SOA and its five potential moderators: the trait being judged, evaluativeness of the trait, the social desirability, the gender, and the age of the target, and examined these relationships through six research questions. Four of the six questions were structured, two more explorative in their nature.

1) What level of SOA can be achieved in job interviews, when judges are experts in assessing personality, compared to previous research on agreement? We hypothesize that expert judges can yield higher mean agreement coefficients than previous studies have demonstrated, when the targets are strangers (Connelly & Ones, 2010; Paunonen, 1989; John & Robins, 1993).

2) What are the most judgable traits? We hypothesize that the most visible traits will be related to Extraversion and most covert traits to Neuroticism, as demonstrated in previous studies (Barrick et al., 2000; Funder, 2012; Connelly & Ones, 2010).

3) John & Robins (1993) demonstrated that self-other agreement increases as traits become more neutral and decreases with more undesirable/desirable and less neutral traits, which is a finding consistent with the RAM model (Funder, 1999). We hypothesize that this tendency will decrease SOA for highly evaluated traits (Conscientiousness, Agreeableness, Neuroticism) and increase SOA for more neutral traits (Extraversion, Openness), as earlier research has suggested (Funder, 1995; Connelly & Ones, 2010).

4) Does the socially desirable response style of a target act as a moderator for agreement? The effect of desirability was considered particularly relevant due to the job-seeking context of the assessments. As studies accounting for the effects of social desirability in personality judgment tasks have thus far been scarce and contradictory, we let this hypothesis remain an explorative question.

5) We hypothesize that there will be a significant difference between female and male targets. On average, female targets will reach higher mean agreement, except on Aggression. This trait will reach higher agreement with male subjects, due to cultural allowance for males to express their behavioral aggression more freely in interaction (Eagly & Steffen, 1986).

6) Can target age act as a moderator in a judgment situation? Borkenau et al. (2004) demonstrated that shared age stereotypes were mostly accurate and contributed negligibly to the agreement among judges. However, studies on this moderator on self-other agreement in

organisational settings are virtually non-existent. Therefore, we let this hypothesis remain an explorative question.

METHOD

Procedure

The data were derived from a Finnish assessment center in 2006-2007. All personality judgments were made in a 1-day psychological assessment during structured job selection interviews by expert psychologists. Research has shown that unstructured interview formats produce a richer quality response and nonverbal behavior from the job applicant and subsequently stronger levels of SOA (Blackman, 2002); hence the restricted context was not ideal for judges. The judges were assigned to the targets randomly. The assessment day typically lasted 6-8 hours and included several personality and ability tests with supervising assistants, discussion group exercises, and a one-hour job selection interview. The interaction between judges and targets was restricted to this job interview. The judges resorted to no test results before making their judgments. Care was taken to ensure that the judges had no previous contact with the targets.

Participants

The targets in the study included 139 job-seeking candidates (84 males and 55 females) in various age groups (23-56 years, mean 36 years, SD=8.4 years). Most of the job applicants were seeking managerial positions; hence the emphasis in the assessment day was on managerial competencies. Judges (N=14, seven males and seven females) were all professional psychologists, who had several years of working experience as assessment consultants and

interviewers (5-25 years of experience, ages 33-53). Their main expertise and background was in I/O psychology, and several judges had received additional clinical training. Hence, a total of 14 expert judges assessed a total of 139 targets individually (every job candidate met only with one judge). Most of the judges rated 10 to 15 candidates, while the range was 1-18 judgements and median 10.5 judgments per consultant.

Measures

In addition to the widespread FFM, Allport's (1937) distinction between expressive-stylistic traits and motivational traits and Murray's famous motivational taxonomy (Murray, 1938) have inspired many personality measuring instruments, such as Douglas Jackson's Personality Research Form (PRF) (Jackson, 1999), which is applied in the present study. We focus here on 15 motivational traits of the Finnish version of PRF (Niitamo, 1997), including two Neuroticism-related traits (Anxiety and Guilt feelings).

In order to investigate job-seekers' self-reported personality, all targets filled 15 scales of the Finnish version of the PRF Form E inventory (Jackson, 1999). The Finnish PRF was originally validated in 1997 in the Finnish Institute of Occupational Health (FIOH), and it has since been standardised and validated with over 10 000 recruitment candidates (cf. Furnhan & Nederström, 2010; Honkaniemi & Feldt, 2008). In the Finnish version, only the scales with better than α=0.60 internal consistency were included, but the construction of the remaining scales is the same as in Jackson's version (Niitamo, 1997). Each personality scale was measured by 16 items (except Anxiety, which is measured by 20 items) that the respondent is instructed to mark as either True or False. Answers to each scale were tallied to form a raw score. These scales were (reliability in personnel selection sample in parenthesis): Achievement (α=0.74), Affiliation (α=0.82), Aggression (α=0.73), Cognitive Structure (α=0.69), Defendence (α=0.72),

Dominance (α=0.86), Exhibition (α=0.85), Harmavoidance (α=0.84), Impulsivity (α=0.82), Nurturance (α=0.72), Order (α=0.89), Sentience (α=0.74), and Succorance (α=0.70). The rest of the original scales (with reliabilities lower than α=0.60; Abasement, Autonomy, Change, Desirability, Endurance, Infrequency, Play, Social Recognition, Understanding) were not included in the Finnish version (Niitamo, 1997). Two additional neuroticism-related trait scales were used instead (Guilt feelings and Anxiety, alphas 0.72 and 0.90). These scales were factor-analyzed with the original scales. The results are extensively introduced in a Finnish validation manual of PRF (Niitamo, 1997) and are available upon request from the first author. In addition to content scales, the PRF includes a scale for social desirability (SD), which contains 12 items, and its alpha is 0.70 in a Finnish standardization sample.

In order to rate job-seekers' personality, all judges filled an external personality rating form after the interview. The form included a short verbal description of low and high poles of PRF dimensions with a graphic 1-10 rating scale for each dimension. Thus, the 14 judges rated a total of 15 different personality traits of 139 targets.

Statistical analyses

The statistical analysis was performed in several steps. First we calculated Pearson's correlations between PRF self-report and expert ratings for all targets and all personality traits. These associations were examined by using both divergent and convergent correlations. Divergent validity helps to establish convergent validity by demonstrating that the construct the judges were assessing (e.g., Anxiety) is different from other constructs that might emit similar trait-relevant cues (e.g., Guilt feelings). In an ideal situation, each convergent correlation should be significant and exceed the highest divergent correlation (e.g., Bastiaansen, Rossi & DeFruyt, 2012).

The next step was to calculate SOA coefficients separately for male and female targets. The significance of these differences was computed with a procedure introduced by McNemar (1962). Alexander and DeShon (1994) suggested, based on a Monte Carlo study, that researchers should avoid using the moderated multiple regression (MMR) analysis when testing hypotheses regarding categorical moderators.

In the third step, we calculated a rank order of evaluativeness for PRF traits by correlating every PRF trait to the SD scale. This SD estimate for trait evaluativeness was adopted for the present sample, because trait evaluativeness may depend on the assessment context and purpose (Birkeland et al., 2006). This procedure allowed us to examine on which traits the greatest social value is placed in the present sample instead of relying on external data.

In the fourth step we evaluated SOA as a function of target age by employing MMR analysis (Howell, 2002). The self-reports were entered into the regression equation, followed by target age and the product term “self-report x age”, when using the expert ratings as a criterion variable. A moderator effect is revealed if the product term adds significant prediction beyond the linear contributions of the predictor variable (self-reports) and the moderator variable (target age) (Paunonen & Jackson, 1988; Paunonen, 1989). This procedure was repeated for each PRF trait.

The fifth step included an MMR, where we evaluated self-other agreement as a function of targets’ social desirability. The self-reports were entered into the regression equation, followed by target SD and the product term “self-report x SD”. This procedure was repeated, again, for each PRF trait.

RESULTS

To examine the hypothesis that job interviews by expert judges can achieve a substantial level of SOA, targets' self-reports were correlated with expert ratings. As the first column in Table 1 indicates, these findings were generally supportive for the hypothesis. 12 of the 15 dimensions judged and the other (Anxiety) of the two Neuroticism subscales yielded significant self-other correlations. The mean and median self-other trait correlations were 0.25 in the present study. The number of divergent correlations equal to or higher than the convergent correlation was relatively low. This convergence demonstrated the judges' ability to conceptually separate the traits from each other, although some individual divergent correlations were higher than the convergent ones for Defendence and Succorance, which were not differentiated accurately from the other traits.

[INSERT TABLE 1 HERE]

We examined the evaluativeness of the traits through their relationship to social desirability as a criterion. Seven of the 15 PRF traits (Aggression, Achievement, Nurturance, Impulsivity, Order, Cognitive Structure, Defendence) had a substantial ($r>.20$) and significant ($p<**$) association with the SD scale. Thus, these can be considered highly evaluative traits, which support earlier suggestions of evaluativeness (Birkeland et al., 2006; Connelly & Ones, 2010). However, these traits carried no relationship to the rank order of SOA as described in Table 1, but instead, were randomly distributed.

Table 2 shows the agreement coefficient difference between genders in absolute Fisher z-scores and the significance of the difference[1]. As the z-score shows, there was one significant gender difference in trait judgability, namely in Aggression. The results demonstrated that male targets were significantly easier to judge on this trait. There was also a minor (z-score >1) gender difference on Affiliation, Guilt Feelings, Sentience, Dominance, and Achievement. The male targets were easier targets on all of the aforementioned six traits except on Achievement, although the coefficient differences were not significant.

[INSERT TABLE 2 HERE]

The results of 15 MMR analyses can be seen in the fourth and fifth columns of Table 2. These analyses demonstrated a significant effect of target age on one PRF trait, namely Anxiety. On this trait, the moderating effect of age was positive: age increased the relationship between self-reports and expert ratings (β=.20, *t*=2.42*). The moderator analyses revealed no significant interaction effects for other traits. In addition, the MMR with SD as a moderator revealed two significant traits, where SOA was negatively related to SD, namely Dominance (β=.-158, *t*=-2.20*) and Defendence (β=-.23, *t*=-2.67**). On these two traits, lower SD led to higher SOA.

DISCUSSION

[1] The Fisher r to z transformation was used to estimate the significance of these differences.

Our results can be compared to previous studies of SOA: those reported by Paunonen (1989), with correlations of PRF scales across different levels of target-judge acquaintanceship (mean agreement coefficient 0.22, median 0.24, range 0.08-0.42), and agreement on the Five-Factor traits reported by Connelly & Ones (2010) (mean agreement coefficient across all rating groups 0.46, median 0.45, range 0.39-0.51). The mean and median self-other trait correlations were 0.25 in the present study. Nine of the 15 traits exceeded the mean agreement coefficient reported by Paunonen (1989). Yet our mean coefficient was substantially lower than what Connelly & Ones (2010) reported, which is probably due to a large amount of close acquaintances in their rater sample (e.g., family members, friends and cohabitators).

Several previous findings have suggested that traits related to Extraversion are easier to judge, whereas Emotional stability and Intellect/Imagination provide less visible cues for an accurate judgment (Borkenau & Liebler, 1992; John & Robins, 1993; Funder & Dobroth, 1987; Funder, 2012). Funder (1995) suggested that the difference in trait judgability is mainly due to differences in trait evaluativeness and level of visibility.

The following trait results are interpreted by grouping them according to the FFM (Ashton, Jackson, Helmes & Paunonen, 1998). The effect sizes on Extraversion-related traits Dominance (r=.52) and Exhibition (r=.42) were comparable to findings on SOA with cohabitators and higher than self-stranger agreement (Connelly & Ones, 2010), even when the expert judges had no prior interaction with their targets. The visibility of these two traits in the present study might have been inflated due to the job-interview situation, which is a highly interactive and self-representative context. Furthermore, this finding gives support to the hypothesis that these personality traits are relevant, available and non-evaluative (Funder, 1995).

One Extraversion-related trait, Affiliation, appear as less visible than the others. This could have been due to the item content of this trait in PRF, which is linked more closely to hidden Agreeableness than expressive Extraversion.

Conscientiousness, which has been considered a highly work-relevant (e.g., Barrick & Mount, 1991; Saldago, 2003) and also a socially desirable trait (Birkeland, 2006; Paulhus & John, 1998; Honkaniemi & Feldt, 2008) includes Cognitive Structure, Order, Achievement, and Impulsivity (reversed) of PRF. All other traits except Achievement yielded a significant agreement, although the effect size was smaller than within Extraversion. This finding is comparable to ratings of work colleagues and incidental acquaintances on Conscientiousness (Connelly & Ones, 2010).

On Agreeableness-related traits (Nurturance and Succorance,), the agreement appeared to be relatively low. This was in line with Paunonen (1989) with regard to Succorance, whereas he reported a substantial agreement (r=.36) on Nurturance. Agreeableness has reached somewhat mixed results also in studies using the FFM framework (Connelly & Ones, 2010). The contradictory results may stem from the divergent nature of the Agreeableness subscales; their number and content vary between measuring instruments. The high evaluativeness of this trait can also offer an explanation for the previous discrepant findings (John & Robins, 1993).

Harmavoidance and Sentience, which both are related to Intellect/Imagination, were judged almost with the same level of agreement as in Paunonen's study (1989). The agreement on Sentience (r=.37) was higher than found in most of the zero-acquaintance judgments on Intellect/Imagination trait (Connelly & Ones, 2010). Yet the agreement on this scale was similar to the agreement coefficients reported by Paunonen (r=.36; 1989). This result might be explained by item differences between the corresponding PRF and the FFM scales.

A particularly notable exception compared to previous studies, and contrary to our hypothesis, was high agreement on Anxiety scale, which is a subscale of Neuroticism (or the low pole of Emotional Stability). Neuroticism is usually regarded as a relatively covert trait for strangers (Connelly & Ones, 2010). On the other hand, this finding is consistent with those of Furnham et al. (2008), who demonstrated that work-related assessment center constructs and Neuroticism-related traits can reach substantial self-other correlations. It is possible that a stressful job-seeking situation arouses maximal variance between the interviewed candidates. In more relaxed situations this transient activation of the high Anxiety would probably not take place, and this individual variation would have remained hidden. This explanation is consistent with the RAM model, which suggests that the visibility of a trait may be dependent on the contextual demands, and the visibility of a trait can vary across settings (Funder, 1999).

The SOA appeared significantly also on another subscale of Neuroticism, namely Aggression, which reached greater agreement than reported by Paunonen (1989). The agreement on this scale was mainly contributed by male targets, which is discussed in the following paragraphs. The other facets of Neuroticism (Defendence, Guilt Feelings) were judged with zero or almost zero agreement, which is in line with the findings of Paunonen (1989). This could be due to the more hidden, indirect and passive nature of behaviors related to these scales.

When examining the evaluativeness of the traits, seven of the PRF traits (Aggression, Achievement, Nurturance, Impulsivity, Order, Cognitive Structure, Defendence) had a significant relationship with the SD scale. This result is consistent with previous studies, which suggest that in the FFM Conscientiousness (Achievement, Impulsivity, Order and Cognitive Structure), Agreeableness (Nurturance) and Neuroticism (Defendence and Aggression) are the most evaluatively loaded traits (Birkeland, et al., 2006; John & Robins,

1993). However, we found no systematic relationships between the agreement level and the desirability of the traits. Funder (1995) argued that high trait evaluativeness would endanger SOA because targets try to suppress undesirable trait information. Our findings yielded no support to this hypothesis.

When the socially desirable responding style of interviewees was studied, the moderator results suggested that SOA was negatively related to SD on two traits, Dominance and Defendence. In other words, higher social desirability in questionnaire responses led to lower SOA on these traits. Defendence is an undesirable trait, so it is plausible that interviewees high in SD attempt to hide this trait in the selection process (cf. Funder, 1995; Schmid Mast et al., 2011). This finding was in line with our hypothesis, even the SD had no effect on the correlations on other facets of Neuroticism. These results seem to suggest that SD does not systematically distort the key elements of SOA. This finding is consistent with that of Paunonen (1989), who concluded that social desirability was found not to be a large component in the ratings of friends or of strangers. One explanation for this result may be a consistent effort put to impression management: targets high in SD can distort both their questionnaire responses and interview behavior.

Target age and gender were significant moderators of agreement on certain traits. On Anxiety the moderating effect of age was positive: age increased the relationship magnitude between self-reports and expert ratings. The direction of association was as suggested earlier in the literature. This could be regarded as support for our hypothesis that people's facial expressions, over time, form lasting features on their faces and hence increase SOA.

We also found one substantial gender difference in line with our hypothesis, which was the higher agreement on Aggression with the male targets. This finding is also in line with

aggression research (Eagly & Steffen, 1986), which has suggested a cultural allowance for males to express their aggression more freely in interaction. Otherwise, our results contrast with the hypotheses suggested by Funder (1999) and Hall (1984) that females are easier targets of personality judgment. Furthermore, Furnham et al. (2008) found no substantial gender variance in SOA between personality and competence ratings in their assessment center sample.

Even the moderator effects were infrequent in the present sample, their direction was as hypothesized. As Paunonen (1989, p. 828) suggested, these effects can be considered noteworthy, because "reliable moderator and interaction effects have been notoriously elusive, even when proper multivariate procedures have been used to pursue them."

Limitations

The study has some limitations. The judgment context was exceptionally structured and controlled; hence the amount or quality of information as a moderator of agreement could not be studied, which is one important aspect of the RAM (Funder, 1999) and a relevant variable in a selection interview (Barrick et al., 2000).

Secondly, the inter-rater reliability, rating consensus and properties of judges could not be explored, mainly due to the dyadic nature of the interviews and the small sample size of interviewers (N=14). The accuracy of judges can act as a SOA moderator according to the RAM and empirical research (Schmid Mast et al., 2011). Although the convergent effects were statistically significant (and thus supported indirectly the reliability of the expert ratings), the relatively small sample size of professional judges and the unequal amount of targets for each judge did not allow us to explore if there were differences between raters in terms of rating

accuracy. The large sample size requirement of moderator tests makes these effects difficult to detect.

CONCLUSIONS

Our main contribution was to bring the traditional research questions of the SOA into a real-life assessment center with expert judges and target moderators that are seldom investigated (However, see Furnham et al., 2008, who examined relationships between assessment competency ratings and personality). Our judges and targets were highly-educated professionals rather than students in a mock interview (cf. Schmid Mast et al., 2011; Barrick et al., 2000). This gives confidence in generalizing our results to an applicant population of this type.

Surprisingly few studies have considered a real-life selection context when investigating the agreement of personality judgments (Furnham et al., 2008; Barrick et al., 2000; Blackman & Funder, 2002; Blackman, 2002), despite of its practical importance in selection decisions. Interviewer ratings of applicant personality are useful because they are related to important outcomes at work (Connelly & Ones, 2010; Barrick et al., 2000).

In general, these findings suggest that expert ratings on the basis of job interviews can reach significant and a substantial agreement on some and a moderate agreement on many traits. Compared to previous studies, the effect size of agreement on average seemed to fall somewhere between the self-informant and self-stranger (e.g., Connelly & Ones, 2010; Paunonen, 1989). However, contrary to our hypothesis, SOA took place also on less visible traits, which are important in predicting job performance, e.g., Neuroticism (Barrick et al., 2000). This seems to suggest that the assessment center context may play a major role when judging

certain traits (cf. Furnham et al., 2008), and the visibility of a trait may be dependent on the demands of the assessment situation. It would be interesting to study this effect closer. Can Neuroticism be regarded as a visible trait under typical assessment conditions which evoke the individual differences in neurotic behavior and, thus, provide more relevant cues for the professional judges?

We believe that the present study provides valuable nuances to discussion on the impact of moderators in a personality judgment process. From a practical perspective, it seems likely that the validity of personality judgments is not seriously threatened neither by social desirability of targets nor evaluativeness of traits. Furthermore, the effect of targets age played a minor role in SOA. The agreement coefficients were, to a larger extent, affected by the moderating effects of target gender. Further research is needed to explore the relationships between SOA, moderators and organisational outcomes. It would also be important to take intelligence, which is the single best predictor of work success in complex jobs (Schmidt & Hunter, 1998), into account as a potential moderator of SOA. Some researchers (Tett, Freund, Christiansen, Fox, & Coaster, 2012) have suggested that intelligent candidates are more prone to apply impression management tactics, which, in turn, can affect SOA.

References:

Albright, L., Kenny, D. & Malloy, T. (1988). Consensus in Personality Judgments at Zero Acquaintance. *Journal of Personality and Social Psychology,* 55, 387-395.

Alexander, R. A. & DeShon, R. P. (1994). Effect of error variance heterogeneity on the power of tests for regression slope differences. *Psychological Bulletin*, 115, 308-314.

Ambady, N. & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111, 256–274.

Ashton, M.C., Jackson, D.N., Helmes, E., & Paunonen, S.V. (1998). Joint factor analysis of the Personality Research Form and the Jackson Personality Inventory: Comparisons with the Big Five. *Journal of Research in Personality*, 32, 243–250

Barrick, M.R. & Mount, M.K. (1991). The Big Five personality dimensions and job performance: A Meta-Analysis. *Personnel Psychology*, 44, 1-26.

Barrick, M.R., Patton, G.K., & Haugland, S.N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychology*, 53, 925–951

Bastiaansen, L., Rossi, G. & DeFruyt, F. (2012). Comparing Five Sets of Five-Factor Model Personality Disorder Counts in a Heterogeneous Sample of Psychiatric Patients. *European Journal of Personality*, 27, 377–388.

Biesanz, J.C., Human, L.J., Paquin, A., Chan, M., Parisotto, K.L., Sarracino, J. & Gillis, R.L. (2011). Do we know when our impressions of others are valid? Evidence for realistic accuracy awareness in first impressions of personality. *Social Psychological and Personality Science*, 2, 452-459.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T. & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, 14, 317–335.

Blackman, M.C. (2002). The Employment Interview via the Telephone: Are We Sacrificing Accurate Personality Judgments for Cost Efficiency? *Journal of Research in Personality*, 36, 208–223.

Blackman, M.C. & Funder D.C. (2002). Effective Interview Practices for Accurately Assessing Counterproductive Traits. *International Journal of Selection and Assessment*, 10, 109-116.

Borkenau, P. & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology*, 62, 645-657.

Buck, R., Miller, R. E. & Caul, W. F. (1974). Sex, personality, and physiological variables in the communication of affect via facial expression. *Journal of Personality and Social Psychology*, 30, 587–596.

Church, A.H. (1997). Managerial Self-Awareness in High-Performing Individuals. *Journal of Applied Psychology*, 82, 281-292.

Colvin, C.R. (1993). Judgable people: Personality, behavior, and competing explanations. *Journal of Personality and Social Psychology*, 64, 861-873.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136, 1092–1122.

Costa, P.T. & McCrae, R.R (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) professional manual. Odessa, FL: PAR.

Costa P.T. Jr. & McCrae R.R. (2006). Trait and factor theories. In: J.C. Thomas and D.L. Segal A. (Eds.), Comprehensive Handbook of Personality and Psychopathology, Vol. 1. (pp. 96-114). New York: Wiley.

Eagly, A.H & Steffen, V.J. (1986). Gender and aggressive behavior: A meta-analytic view of the social psychological literature. *Psychological Bulletin*, 100, 309-330.

Estes, S. G. (1938). Judging personality from expressive behavior. *Journal of Abnormal and Social Psychology*, 33, 217-236.

Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self-other rating agreement in leadership: A review. *Leadership Quarterly*, 21, 1005–1034.

Funder, D. C. & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, 52, 409–418.

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

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670

Funder, D.C. (1999). Personality Judgment: A realistic approach to person perception. London: Academic Press.

Funder, D.C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21, 177-182

Furnham, A., Jensen, T. & Crump, J. (2008). Personality, Intelligence and Assessment Centre Expert Ratings. *International Journal of Selection and Assessment*, 16, 356–365.

Furnham, A. & Nederström, M. (2010). Ability, demographic and personality predictors of creativity. *Personality and Individual Differences*, 48, 957-961.

Hall, J. A. (1984). Nonverbal sex differences: Communication accuracy and expressive style. Baltimore: Johns Hopkins University Press.

Hogan, R., Hogan, J. & Roberts, B.W. (1996). Personality Measurement and Employment Decisions - Questions and Answers. *American Psychologist*, 51, 469-477.

Honkaniemi L. & Feldt T. (2008). Egoistic and moralistic bias in real-life inventory responses. *Personality and Individual Differences*, 45, 307–311.

Howell, D.C. (2002). Statistical Methods for Psychology. Pacific Grove, CA: Duxbury.

Hubbard, J. (1994). What's In A Face: Correlations Of Self, Spouse, And Stranger Ratings Of Personality. Unpublished master's thesis. Macalester College.

Jackson, D. N. (1999). Personality research form, manual. Port Huron: Sigma Assessment Systems.

John, O.P. & Robins, R.W. (1993). Determinants of interjudge agreement on personality traits: the Big Five domains, observability, evaluativeness, and the unique perspective of the self, *Journal of Personality*, 61, 521-551.

Kihlstrom, J. F. & Haste, R. (1997). Mental representations of persons and personality, in Hogan, R., Johnson J. & Briggs, S. (Eds.), Handbook of Personality Psychology, San Diego: Academic Press.

Little, A.C. & Perrett, D.I. (2007). Using composite images to assess accuracy in personality attribution to faces. *British Journal of Psychology*, 98, 111–126.

McCrae, R. R. & Costa, P T., Jr. (1982). Self-concept and the stability of personality: Cross-sectional comparisons of self-reports and ratings. *Journal of Personality and Social Psychology*, 43, 1282-1292.

McCrae, R. R. & Costa, P T., Jr. (1989). Different points of view: Self-reports and ratings in the assessment of personality. In J. P. Forgas & J. M. Innes (Eds.), Recent advances in social psychology: An international perspective (pp. 429-439). North-Holland, The Netherlands: Elsevier Science.

McNemar, Q. (1962). Psychological statistics. Third Edition. New York: John Wiley & Sons.

Murray, H. A. (1938). Explorations in personality. New York: Oxford University Press.

Nederström, M. & Furnham, A. (2012). The relationship between the FFM and personality disorders in a personnel selection sample. *Scandinavian Journal of Psychology*, 53, 5, 421–429.

Niitamo, P. (1997). PRF Personality Research Form, manual. Helsinki: Psykologien Kustannus Oy.

Nunnally, J.C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Paulhus, D. L., & John, O. P. (1998). Egoistic and Moralistic Biases in Self-Perception: The Interplay of Self-Deceptive Styles with Basic Traits and Motives. *Journal of Personality*, 66, 1025-1060.

Paunonen, S.V. (1989). Consensus in personality judgments: Moderating effects of target-rater acquaintanceship and behavior observability. *Journal of Personality and Social Psychology*, 56, 823-833

Paunonen, S. V., & Jackson, D. N. (1988). Type I error rates for moderated multiple regression analysis. *Journal of Applied Psychology*, 73, 569-573

Paunonen, S. V., & Jackson, D. N. (2000). What is beyond the Big Five? Plenty! *Journal of Personality,* 68, 821-835.

Piedmont, R.L., McCrae, R.R. Riemann, R., & Angleitner A. (2000). On the invalidity of validity scales: evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology,* 78, 582–593

Schmid Mast, M., Bangerter, A., Bulliard, C., & Aerni, G. (2011). How accurate are recruiters' first impressions of applicants in employment interviews? *International Journal of Selection and Assessment*, 19, 198-208.

Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.

Tett, R. P., Freund, K. A., Christiansen, N. D., Fox, K. E., & Coaster, J. (2012). Faking on self-report emotional intelligence and personality tests: Effects of faking opportunity, cognitive ability, and job type. *Personality and Individual Differences*, 52, 195-201.

Zebrowitz, L.A. & Collins, M.A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, 1, 203-222.

Table 1. Self-other agreement coefficients for 15 personality traits, median divergent correlations and 95% confidence intervals for each PRF trait (N=139).

	Convergent correlations	Median divergent correlations §	95% confidence intervals	
PRF scale	All targets		ConfL	ConfU
Dominance	.52***	.02 (0)	0.39	0.63
Exhibition	.42***	.00 (0)	0.27	0.55
Sentience	.37***	.01 (0)	0.22	0.51
Impulsivity	.35***	.01 (0)	0.20	0.49
Anxiety	.33***	.06 (0)	0.17	0.47
Aggression	.28**	.04 (0)	0.12	0.43
Harmavoidance	.28**	.10 (1)	0.12	0.43
Affiliation	.25**	-.03 (0)	0.09	0.40
Order	.23**	-.01 (0)	0.07	0.38
Cognitive structure	.22**	.08 (2)	0.06	0.37
Guilt feelings	.20*	.07 (1)	0.04	0.36
Achievement	.17*	-.03 (2)	0.00	0.33
Nurturance	.15	.00 (1)	-0.04	0.33
Succorance	.05	-.01 (9)	-0.14	0.24
Defendence	-.11	.00 (14)	-0.27	0.06

*Note. Probabilities are two-tailed values: *p <.05. **p <.01 ***p<.001.*

§ The number of correlations equal to or higher than the convergent correlation is given between parentheses.

Table 2. Gender differences on agreement computed as Fisher r-to-z transformation (N=84 for male targets, 55 for female targets), and age and SD as moderators (N=139).

	Gender differences			Age as a moderator	SD as a moderator
PRF scale	Males	Females	z difference		
Dominance	.58***	.43**	1.12	-	* (-)
Exhibition	.42***	.42**	0.01	-	-
Sentience	.36***	.14	1.37	-	-
Impulsivity	.40***	.27*	0.86	-	-
Anxiety	.35**	.34*	0.01	* (+)	-
Aggression	.40***	.08	1.93*	-	-
Harmavoidance	.23*	.37**	0.88	-	-
Affiliation	.33**	.05	1.63	-	-
Order	.18	.30*	0.72	-	-
Cognitive structure	.23*	.24	0.10	-	-
Guilt feelings	.29**	.02	1.57	-	-
Achievement	.09	.27*	1.06	-	-
Nurturance	.15	.03	0.72	-	-
Succorance	-.02	-.07	0.31	-	-
Defendence	-.07	-.11	0.23	-	** (-)

*Note. Probabilities are two-tailed values: *p <.05. **p <.01 ***p<.001.*

II

THE RELATIONSHIP BETWEEN THE FFM TRAITS AND PERSONALITY DISORDERS IN A PERSONNEL SELECTION SAMPLE

by

Mikael Nederström & Adrian Furnham, 2012

Scandinavian Journal of Psychology, *53*, 421–429

Reproduced with kind permission by Wiley.

ABSTRACT

The relationships between the Five Factor Model personality and personality disorders were investigated. A sample of real-life job applicants completed two personality questionnaires with different theoretical backgrounds in a psychological assessment center. The job applicants provided self-descriptions both on the Five Factor Model inventory and on the personality disorder trait inventory. A subsample of these candidates was interviewed by expert psychologists upon entrance to the assessment center. The psychologists assessed the same disorder traits of each target in job interviews. Both self-descriptions were used to predict the expert assessments. The results demonstrated considerable overlap between the FFM measures of normal and measures of abnormal personality in both samples and regardless of assessment method.

Keywords: Five Factor Model; personality disorders; selection assessment; personality judgments

INTRODUCTION

From the early years of personality research, a central aim has been to define, measure, and integrate the huge amount of divergence in concepts of human personality. From the very beginning of the research, some theorists, usually clinicians were interested in measuring abnormal traits, while differential psychologists seemed more interested in measuring normal traits. Certainly more theoretical and empirical work has been invested in the latter rather than the former. After years of debate, the Five Factor Model (FFM) has gradually become one of the most widely used normal trait classification instruments (Goldberg, 1990; Costa & McCrae, 1995; Salgado, 2003). This model includes five global traits of personality which seem prevalent in some form in every culture (Costa & McCrae, 2004). Another tradition found both in psychology and psychiatry, with a different history, taxonomies, and research instruments, has been related to clinical purposes. One of the main classifications of pathologies of personality in clinical applications can be found from the regularly updated Diagnostic and Statistical Manual of Mental Disorders, revised to its fourth version (American Psychiatric Association, 1994). Despite of their theoretical and diagnostic roots, several DSM-based personality classification systems and related inventories have increased their popularity outside clinical applications (e.g., Hogan & Hogan, 2001a; De Fruyt, De Clercq, Miller, Rolland, Jung, Taris, Furnham, & Van Hiel, 2009).

According to DSM IV, there are 10 different personality disorders. These disorders are characterized, as persistent, inflexible, and maladaptive ways of relating to oneself and one's environment (APA, 1994). Unlike dimensional measures of personality, such as self-report trait questionnaires, DSM has historically considered mental disorders to be qualitatively distinct conditions, diagnosed by psychiatric interview. As the text-revised version of the manual puts it, the DSM "divides mental disorders into types based on criteria sets with defining features" (APA, 2000, p. xxxi).

Those researchers who have looked at the theoretical and psychometric relationship between “normal” (universal) and “abnormal” (psychopathological) measures have suggested that many DSM personality disorders are, in fact, *extreme poles* of normal personality, rather dimensional than qualitative in their nature. According to this notion, sometimes called the spectrum hypothesis, personality traits constitute a continuum from normal personality to disorders, and there is no qualitative distinction between these two concepts (Samuel & Widiger, 2008; Bagby, Costa, Widiger, Ryder, & Marshall, 2005; Widiger & Trull, 2007). Our main goal in the present study is to predict these relationships in a personnel selection sample by utilizing expert ratings of personality disorders as an external criterion.

The spectrum hypothesis: from categories to traits

In recent years, a variety of studies have raised concerns regarding the validity of the categorical model of personality disorders and proposed alternative dimensional models. This dimensional notion, proposed by Eysenck (1947) over 60 years ago, has prevailed for some time being especially within trait theories. Livesley, Jackson, and Schroeder (1992) have shown that, in addition to normal traits, the main differences in personality disorders between clinical samples and normal samples have emerged more from quantitative than qualitative factors. Even its wide prevalence among clinicians, many authors have found serious empirical concerns with the categorical model of personality disorders. For instance, low stability, a great deal of comorbidity and low diagnostic agreement among Axis II instruments are “more the rule than the exception” (Trull & Durrett, 2005, p. 361). These problems suggest that the disorders may not represent entirely distinct diagnostic entities (Trull & Durrett, 2005). Hence, the dimensional approach is partly adopted in the upcoming DSM-5 (American Psychiatric Association, 2011).

In addition to a dimensional interpretation to personality disorder traits, there has been a constant tendency to integrate traditional trait-based personality models and Axis II of DSM-IV (Widiger & Samuel, 2005). As Widiger & Trull (2007, p. 71) put this integrative goal:

> It may be time to consider a shift to a dimensional classification of personality disorder that would help address the failures of the existing diagnostic categories as well as contribute to an integration of the psychiatric diagnostic manual with psychology's research on general personality structure.

One obvious candidate for general personality structure in this context is the FFM. Empirical research has already achieved encouraging results (De Fruyt et al., 2009; DeCuyper, De Pauw, De Fruyt, De Bolle, & De Clercq, 2009; Clark, 2007; Furnham & Crump, 2005; Widiger, Trull, Clarkin, Sanderson, & Costa, 2002; Lynam & Widiger, 2001). The FFM seems to provide a solution for problems of *comorbidity* with categorically diagnosed personality disorders. Unlike the more distinct FFM factors, many personality disorders seem to appear together. This comorbidity can be explained with certain common FFM factors, which underlie different disorders (Lynam & Widiger, 2001; Bagby et al., 2005). For example, Schizoid and Avoidant disorders can often be diagnosed together, because they have a common source of variance in introversion. On the other hand, a differential diagnosis between these disorders can be justified with a strong correlation between Avoidant personality and neuroticism, which does not exist between Schizoid and neurotic personality. (Widiger & Costa, 1994).

The five big domains ('super-factors') can be further decomposed to the primary factor or facet level, when more detailed measurement is necessary (e.g., Costa & McCrae, 1992). However, there are still few measures of the FFM that have been studied at the facet level and correlated with the DSM disorders until the meta-analysis of Samuel and Widiger (2008).

The present study focuses on relationships between the FFM facet level personality and DSM personality disorders in a personnel selection context. It answers De Fruyt and Salgado's (2003) call for more research on relationships between psychopathological personality and organisational behavior. The main contribution of this study is to demonstrate the convergence of

the FFM and DSM models. This is conducted by avoiding common method variance (using external expert criteria for assessing personality disorders) and utilizing detailed facet level of the FFM traits.

Personality models and organisational outcomes

In addition to conceptualizing personality, the FFM has many empirical implications. Empirical support for the construct validity of the FFM has grown extensive during the last 20 years, at both the global and the facet levels, including convergent, discriminant, and predictive validation across self-other agreement (Costa & McCrae, 1988), temporal stability (Costa & McCrae, 1994; Löckenhoff, Terracciano, Bienvenu, Patriciu, Nestadt, McCrae, Eaton, Costa, Jr, 2009), heritability (Plomin & Caspi, 1999), and work success (Barrick & Mount, 1991; Saldago, 2003; Hogan, Hogan & Roberts, 1996). Studies of the FFM and leadership behavior have established a close connection between leader personality and leadership style (Judge, Bono, Ilies, & Gerhadt, 2002; Judge & Bono, 2000) as well as at other organisational levels (Barrick & Mount, 1991; Salgado, 2003).

These suggestions have also relevant implications regarding the DSM and organisational psychology (Furnham, 2008). De Fruyt et al. (2009) administered five samples to investigate the usefulness of using FFM traits to screen out personality dysfunction in work settings. This research found evidence for both construct and predictive validity of the personality disorder compound scales constructed with the FFM. Some recent research projects have concentrated on the relationships between personality disorders and organisational outcomes (Ullrich, Farrington & Coid, 2007; Kets de Vries, 2006; Hogan, 2001a; Furnham & Taylor, 2004). The question of “what makes a leader successful” has been replaced with “what kind of traits cause leaders to fail” (Hogan & Hogan, 2001a; Furnham, 2010; Babiak & Hare, 2006). These studies have suggested that dark-side tendencies are not only useful in screening out individuals, but might also be helpful in improving incumbents’ functioning in the course of development (cf. De Fruyt et al., 2009).

Aims and hypotheses of this study

Our main purpose was to study relationships between the FFM and the DSM personality disorders in a psychological assessment center with real-life job applicants. We wanted to extend research methodology beyond correlational self-report studies. This was obtained by using expert psychologists' observations as an additional criterion.

Using only self-reports may artificially inflate the trait-trait relationships because of common method variance, i.e. a certain portion of the variance in a measure can be attributed to the method used (Doty & Glick, 1998). One way of overcoming this bias problem is comparing self-report personality scales to external judgments of personality (cf., John & Robins, 1993; Funder & Colvin 1996; Funder, 1999). We utilized two different self-report inventories and expert ratings of DSM-IV disorders as an external criterion. Hence one important methodological point in this study was to explore both convergent and divergent correlations between external expert ratings and self-reports of personality.

Thus the present study examined three different relationships: (1) correlations between two self-report inventories: a FFM self-report inventory and a DSM self-report inventory, (2) self-other agreement between expert ratings of personality disorders and the DSM inventory, and finally (3) self-other agreement between expert ratings of personality disorders and the FFM inventory. All analyses were conducted at the facet level of the FFM.

Primarily, we hypothesized that the expert ratings of the DSM disorders would be equally accounted for by both self-reports regardless of the theoretical background of the inventory. In other words, the FFM inventory scores will predict the expert observations with a similar magnitude than the dedicated DSM inventory. In addition, we examined associations between both self-report inventories in an explorative manner and compare these to previous studies between the FFM and the DSM.

METHOD

Procedure

All data were derived from a large Finnish psychological assessment center during a personnel selection process. The assessment day typically lasted 6-8 hours and included discussion group exercises, a one-hour job selection interview and several personality and ability tests supervised by assistants. All personality judgments were made in a one-day assessment during structured job selection interviews (e.g., Campion et al., 1997). The interaction between judges and targets was restricted to this job interview, where psychologists assessed targets' job competencies, motivation, and personality. The judges did not resort to any test results before making their judgments. It was of the essence that the judges did not know the targets previously, which is a preferred best practice in the assessment center. This was verified by assessment center assistants, who gathered the identification information of candidates before the assessment day.

Participants

A sample of job-seeking candidates was used as targets and one sample of expert psychologists as judges in the present study. As most job seekers were aiming at managerial positions, the emphasis in the interviews was on managerial competencies. The main sample of job seekers consisted of 229 male (70%) and female (30%) candidates in diverse age groups (23-56 years, mean 36 years) with varying educational backgrounds. This sample filled two self-report inventories: the FFM and the DSM-based measures. The subsample (N=105) of the aforementioned 229 job seekers was used as targets in a personality judgment process. This sample was sociodemographically similar than the main sample. In addition to filling the DSM and the FFM self-reports, their personality was judged during the job interviews.

The personality judgments were made by experienced personnel selection psychologists (N=15, 8 men and 7 women), who interviewed the targets one at a time and filled 105 rating forms. The judges were all professionals, who had several years (5-25 years) of working

experience as personnel selection experts and interviewers. They were specialized in industrial and organizational psychology, but they also had clinical education and training in identifying personality disorders. Subsequently, there were a total of 15 expert judges who individually judged a total of 105 targets. Most of the judges rated 4 to 10 candidates.

Measures

In order to investigate job seekers' self-reported personality, all targets filled a Finnish Five Factor Model inventory (PK-5) and a DSM inventory (SRS). The PK-5 is a Finnish language version of the five trait inventory (Openness to experiences, Conscientiousness, Extroversion, Agreeableness, Neuroticism). The Neuroticism factor is coded reversely as Emotional Stability. The PK-5 contains three subscales (or facets) on each factor and 150 respondent-descriptive items, which are responded on a 1-5 Likert scale. A portion of the items are in reversed form to control for acquiescent responding. The PK-5 has relatively high subscale reliabilities (alphas between .75-.94, mean .82 in a Finnish standardization sample). The PK-5 has been normed upon responses from real-world recruitment and training samples (N=3644) with a fully representative work age distribution (20-59 years). The published PK-5 manual reports an extensive validation process in Finnish population. (Psykologien Kustannus, 2006). The PK-5 validation has also been conducted with clinical samples (Saario, 2011). Construct validity has been assessed via relations to well-known personality and work behaviour measures. These tables are available from the first author upon request.

The SRS (Stress Reaction Style inventory) was constructed to measure Axis II personality disorders of the DSM-IV in a form of a self-report questionnaire. These disorder tendencies become apparent under stress, which is a defining characteristic of many current work environments (however, it should be noted that some disorders can be fully expressed without any stress at all, e.g., antisocial or narcissistic personality disorders). The validation process took place between 2008 and 2010 at Psycon Corporation, where this measure was used in a personnel

selection process. It should thereby be emphasized that SRS was intended for use in normal adult working populations. It was not meant for clinical use in any circumstances.

The 10 scales of SRS inventory, each grouping 14 items, have their content in behavioral descriptions of 10 personality disorders, although they have been formulated to better fit the organisational context and are scored dimensionally. The questionnaire contains 152 respondent-descriptive items, which are answered on a 0-4 Likert scale. Internal consistency of the scales in the present sample was on a decent level, ranging from .68 to .87. These reliabilities are comparable to Hogan and Hogan's (2001b) HDS. The mean scores, *SD*'s, reliabilities and sample items of the SRS scales are listed in the appendix. These data were collected from a total of 1697 job applicants. All SRS scales were also cross-validated with the Finnish version of Personality Research Form (Niitamo, 1997) before the present study. A detailed correlation matrix is available from the first author.

Theoretically it could be argued that no one would actually endorse personality disorder symptoms when applying for a job. In order to investigate social desirability of SRS content scales, we constructed a scale for desirability as a control scale. This scale was constructed by combining item contents from two factors of socially desirable responding (self-deception and impression management) introduced by Paulhus and Reid (1991). The desirability scale contains 12 items, and its alpha was .84 in the present sample. However, we did not expect strong negative associations between social desirability and personality disorders, as the inability to perceive a disorder in oneself is symptomatic of many personality disorders and thus it would not effect on level of social desirability. (American Psychiatric Association (1994).

In order to judge targets' personality disorder traits, all psychologists filled an external rating form of the DSM, with a graphic 1-10 rating scale for each dimension. This rating form had a short verbal description of the 10 personality disorder dimensions of the SRS/DSM and a 1-10 estimate of candidates' proneness to answer in a socially desirable way. Thus, the psychologists

judged targets on 10 personality disorders and assessed their responding style. The judges used dimensional model to assess the disorders, without any diagnostic thresholds.

RESULTS

Table 1 shows the pattern of Pearson's correlations between the self-rated FFM facets and the self-rated DSM personality disorders. We wanted to examine disorder differentiation as precisely as possible by using all lower-level facets in addition to the five global factors of the FFM (cf. Samuel & Widiger, 2008). All personality disorders had significant correlations and effect sizes greater than .40 at least with one facet of the global factors. When the two most contributing FFM facets in the correlation matrix are utilized to describe each DSM personality disorder, the adjective pattern looks as follows: Paranoids are suspicious and not confident, Schizoids are unfriendly and closed, Schizotypals are sensitive and imprudent, Antisocials are leading (or manipulating) and unsystematic, Borderlines are emotionally decomposed and unrelaxed, Histrionics are socially bold and lively, Narcissists are leading and socially bold, Avoidants are not confident and unrelaxed, Dependents are not leading (or submissive) and unrelaxed, and Obsessive-Compulsives are systematic and responsible. If these associations are interpreted conversely, the strongest correlation patterns (effect size greater than .50) between the Five-Factor facets and disorders appeared as follows (with the most contributing facet and the direction of its correlation in parentheses):

(1) Extroversion: Schizoid (-Lively), Histrionic (+Lively, +Socially bold), Narcissist (+Leading, +Socially Bold), Avoidant (-Lively, -Socially bold), and Dependent (-Leading).

(2) Agreeableness: Paranoid (-Trusting) and Schizoid (-Friendly).

(3) Conscientiousness: Avoidant (+Systematic) and Obsessive-Compulsive (+Systematic).

(4) Emotional Stability: Borderline (-Composed) and Avoidant (-Composed, -Relaxed).

(5) Openess to Experience did not yield any correlations over .50, while closest to this criterion were Schitzotypal (+Sensitive) and Narcissist (+Open to new experiences).

[INSERT TABLE 1 HERE]

Tables 2 and 3 show all examined relationships between the expert ratings of the DSM personality disorders and the two different self-report measures. Table 2 reports the self-other agreement correlations between expert ratings of the DSM disorders and self-report scores of the SRS, i.e., the convergent correlations. The second column lists the median divergent correlations, which account for all the other disorders rated by the judges. The figure between parentheses in these last columns refers to the number of divergent correlations equal to or greater than the convergent correlation coefficient. The expert judges reached a statistically significant agreement with SRS self-reports on eight of the 10 personality disorders. Convergent correlations were substantial and exceeded the median divergent correlations for all of the rated disorder traits: six of the convergent coefficients were greater than .40. Four individual divergent correlations were higher than the convergent ones: three for the Borderline and one for the Avoidant disorder. Similarly, these disorder traits were the most difficult to judge, with the convergent effect size used as a criterion.

[INSERT TABLE 2 HERE]

Self-other agreement correlations between the expert DSM ratings and self-report scores of the PK-5 can be found from Table 3. The overall magnitude of coefficients in this table was, not surprisingly, lower than correlations between the two self-report measures shown in Table 1, but there were significant relationships between several facets of the PK-5 and the 10 DSM disorders. It should be noted that the expert ratings of the DSM disorders could be explained (at

their best) with the same magnitude by both self-report inventories: Both inventories reach several agreement coefficients with the expert ratings with greater than .40 magnitudes. Seven of the 10 disorders yielded greater than .40 correlations with at least one facet of the PK-5. The clearest exception was Schizotypal disorder, which failed to reach any significant correlation between the PK-5 self-report and the expert ratings.

[INSERT TABLE 3 HERE]

Furthermore, we conducted a canonical correlation analysis between the three sets of personality ratings. The purpose of this analysis was to explain the overall relations of the two sets of variables. The 11 expert ratings of DSM disorders were thereby analyzed with 10 self-reported variables of SRS disorders and with 15 self-reported facets of the PK-5. In our analysis, we used pooled redundancy coefficients (Thompson, 1984) to compute the sum of all the redundancy coefficients for all the variables in a set. It collects the overlapping information of two sets of variables in a comparable way. The total overlap between the variable sets was computed with a total redundancy index recommended by Alpert and Peterson (1972, p. 189), who state that the researcher may "use canonical correlation coefficients to test for the existence of overall relationships between sets of variables, but for a measure of the magnitude of relationships, redundancy may be more appropriate." The total redundancy index was .26 between the expert ratings and the SRS scores, .28 between the expert ratings and the PK-5 scores, and .53 between the PK-5 and SRS self-reports. Hence both inventories seemed to explain the expert ratings roughly with the same magnitude, although the PK-5 was slightly better. As hypothesized, the strongest relationship appeared between the two self-report measures. When explaining the expert ratings with both the five global factors of the PK-5 and the SRS together, the analysis resulted in a

redundancy coefficient of .29. Accordingly, there was no significant increase in validity when both self-report inventories were operated together.

Finally, the results were elaborated with a hierarchical regression analysis. The regression was entered in two blocks: the PK-5 self-ratings and the SRS self-ratings in order to measure the incremental validity of the SRS over the PK-5. These results showed that the SRS gave no statistically significant incremental validity over the PK-5 when predicting expert ratings, except with two disorders: Schizotypal and Antisocial. The gain in validity (R Square Change) resulting from SRS was 7.5 % ($p<.001$) with Schizotypal ratings and 9.0 % ($p<.001$) with Antisocial ratings.

DISCUSSION

This study demonstrated considerable overlap between the FFM trait measures of normal and DSM measures of psychopathological personality. The main contribution of the current study was its use of expert I/O psychologists as an external criterion to explore this overlap. It also extended the previous research to the facet level of the FFM in an organisational context.

A number of the correlations between the FFM inventory (PK-5) and the DSM inventory (SRS) were highly significant and substantial in their magnitude ($r > .50$). These results can be compared with those reported by Furnham and Crump (2005), who used different questionnaires (NEO-PI-R and HDS) with the same theoretical backgrounds. The results were highly similar in both effect size and direction of association. They, too, found the strongest effects between Extroversion and personality disorders, and weaker relationships from disorders to Openness and Agreeableness. In the present study, Borderline, Avoidant, and Dependent disorders had the strongest negative relationships to Emotional Stability, whereas Histrionic and Narcissist disorders related positively to Emotional Stability, which, similarly, replicates Furnham and Crump's (2005) findings. In the current study, Avoidant had more relationships to the FFM traits

than any other disorder, when the effect size of .50 is used as a criterion, which is in line with findings by Furnham and Crump (2005) and Samuel & Widiger (2008). Our results also align well with current efforts to restructure the description and diagnosis of personality disorders as outlined in the upcoming DSM-5 (American Psychiatric Association, 2011). As the regression results demonstrated (p. 14), 8 of 10 Axis II disorder ratings were predicted with equal of better accuracy with the FFM self-reports than the SRS self-reports. Only Antisocial and Schizotypal ratings gained incremental validity from the SRS self-reports.

However, there were two notable exceptions when the relationships between the FFM and the disorders were compared to previous findings. First, in the present study there was a substantial correlation (r = .37) between Antisocial disorder and Emotional Stability, which was mainly contributed by Relaxed facet of the PK-5. This relationship was neither found in previous empirical studies (Furnham and Crump, 2005; Samuel & Widiger, 2008) nor predicted by Widiger et al. (2002). Yet this finding is compatible with clinical predictions which have identified an exceptionally relaxed condition of psychopaths between episodes of restless, anti-social behavior (Martens, 1997). In addition, we found a relatively strong negative relationship (r = -.37) between Schizotypal disorder and Prudent facet of the PK-5. This was not predicted by Widiger et al. (2002), and it was not found as clearly in Furnham and Crump's study (2005). On the other hand, a recent meta-analysis demonstrated (Samuel & Widiger, 2008) a small but significant negative relationship between Schizotypal disorder and all the facets of Conscientiousness. One theoretical explanation can be derived from the thinking styles: The DSM-IV describes one of the symptoms of Schizotypal disorder as odd or "magical thinking that influences behavior and is inconsistent with subcultural norms". Thus, this finding might be explained by the lack of common sense, order and control in thinking styles of highly Schizotypal people.

One novel methodological contribution was the use of expert ratings as real-life criteria. In addition to relationships between self-report measures, the self-other agreement on

personality yielded substantial correlations upon expert judges' ratings of interviewees' DSM personality disorders. The overall magnitude of the agreement was generally at the same level as the coefficients found between close acquaintances with normal personality measures (Funder & Colvin, 1988; Paunonen 1989; Funder, 1999). The relatively high self-other agreement across the several disorders traits can be regarded as support for the trained experts' ability to judge the DSM disorders in an interviewing context. In addition, this finding suggests that the SRS questionnaire has a certain amount of convergent validity in measuring the DSM disorders. This finding was supported when comparing convergent validity coefficients to divergent validity. In all of the 10 disorder traits, the convergent correlations were higher than median divergent correlations for the criterion measure (expert rating).

When the expert-rated disorder traits were first correlated with the PK-5 and then with the SRS self-report, the predictive power of the PK-5 appeared to be comparable to the SRS: All the expert-rated disorders seemed to have at least one substantial association ($r > .40$) to the PK-5 facets. This supports the general hypothesis of the strong relationship between the FFM and the DSM models. The regression analysis also demonstrated these findings. However, there were two exceptions in this predictive pattern, namely Schizotypal and Antisocial disorders.

According to the hypothesis of Widiger et al. (2002), Schizotypal disorder should have a negative relationship to Extraversion and positive relationships to Openness and Neuroticism. Contrary to this assumption, Schizotypal was the only disorder without any significantly corresponding FFM trait. Some support for the Widiger et al. (2002) findings was corroborated in the first analysis (Table 1, relationships between self-reports) but the second analysis with an independent criterion (expert rating) did not confirm this finding. One explanation for this finding could be related to the nature of Schizotypal disorder. Its core symptoms are related to cognitive aberrations and unusual perceptions (De Fruyt et al., 2009), which are described more

as schizophrenic symptoms than extreme poles of, e.g., Openness to experience or some other FFM trait.

In turn, Antisocial disorder has a close relationship to psychopathy, and there is not yet consensus whether psychopathy should be conceived as a categorical or a dimensional construct (Trull & Durrett, 2005). Some recent theories still support a categorical view (e.g., Trull & Durrett, 2005). The categorical (or at least multidimensional) nature of the Antisocial disorder could explain the lower validity of the PK-5 when predicting the expert ratings.

Analysis of canonical correlations was conducted to establish the overall level of overlap between these measures and the expert ratings. This analysis showed that the PK-5 and the SRS explained the expert judgments with a similar magnitude. We did not find any general predictive validity for the SRS over the PK-5, despite the latter questionnaire not being designed to measure personality disorders. The before-mentioned exceptions on Schizotypal and Antisocial disorders did not change the general association between these two measures. These results might help inform DSM-5 revisions and measurement practices in the future, which according to the information presented in the website, still lack strong empirical support (APA, 2011).

Limitations

A number of limitations should be considered when interpreting the results of this study. First, the validation sample consisted mainly of Finnish adults who were active in their work life. It is, therefore, unlikely that the current sample population would have constituted individuals with severe personality disorders. Additional research with different sample populations should be conducted before further generalizing the spectrum hypothesis. Second, the self-reports were drawn on the basis of particular questionnaires, namely the PK-5 and the SRS. These are the most reliable and well standardized self-reports tests of the FFM and DSM disorders used in Finnish psychological assessment centers. However, the validation process has just begun, hence the usage of these measures is conceptually highly explorative at this point. Utilizing different inventories,

even with the same theoretical backgrounds, would probably have caused slightly different results. This incommensurability can cause theoretical concerns when comparing samples and results from different countries. As Samuel and Widiger (2008) demonstrated, many associations between the FFM and personality disorders vary across measuring instruments. Third, the inter-rater reliability and individual rating accuracy of judges could not be explored, mainly due to the confidential nature of the interviews and a small sample size of interviewers (N=15). Although the convergent effects were statistically significant and thus supported indirectly the reliability of the expert ratings, the relatively small sample size of professional judges and the unequal amount of targets for each judge did not allow us to explore if there were differences between raters in terms of rating accuracy, which would be likely when judging strangers' personalities (Hammond, 1996; Funder, 1999). The accuracy of judges can act as a moderator variable between observed disorders and the self-reported FFM traits, and hence change the relationships between these two constructs.

Conclusions

These findings support the previous empirical studies (Furnham & Crump, 2005; Bagby et al.; 2005) and theoretical suggestions (Lynam & Widiger, 2001) that the FFM and personality disorders of the DSM mostly describe the same underlying constructs. The results are potentially useful for taxonomists of personality disorders as well as psychometricians attempting to devise valid self-report measures of personality disorders. The SRS self-report inventory provides a new perspective on the FFM data that are already included in many selection and development assessments. In addition, the relationships between the SRS and the PK-5 suggest one possible pattern how the FFM traits could be mapped with the personality disorders.

REFERENCES

Alpert, Mark I. & Peterson, Robert A. (1972). On the interpretation of canonical analysis. *Journal of Marketing Research* 9(2), 187-192.

American Psychiatric Association (1994). Diagnostic and Statistical Manual of Mental Disorders DSM-IV (4. edition). Washington, DC.

American Psychiatric Association (2000). Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR (Text Revision)

American Psychiatric Association DSM-5 Development website (2011). **http://www.dsm5.org/Pages/Default.aspx**

Babiak, P. & Hare, R. D. (2006). Snakes in Suits. When Psychopaths go to Work. NewYork: Harper Collins Publishers Inc.

Bagby, R. M., Costa, P. T. Jr, Widiger, T. A., Ryder, A. G. & Marshall, M. (2005). DSM-IV Personality Disorders and the Five-Factor Model of Personality: A Multi-Method Examination of Domain- and Facet-Level Predictions. *European Journal of Personality*, 307-324.

Barrick, M.R. & Mount, M.K. (1991). The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology*, 44, 1, 1-26.

Costa, P. T., Jr., & McCrae, R. R. (1988). Personality in adulthood: A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology*, 54, 853-863.

Costa, P.T, & McCrae, R.R (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) professional manual. Odessa, FL: PAR.

Costa, P. T., Jr., & McCrae, R. R. (1994). Set like plaster? Evidence for the stability of adult personality. In T. Heatherton & J. L. Weinberger (Eds.), *Can personality change?* (pp. 21-40). Washington, DC: American Psychological Association.

Costa, P. T., Jr & McCrae, R. R. (1995). Solid Ground in the Wetlands of Personality: A Reply to Block. *Psychological Bulletin*, 117, 216-220.

Costa, P. T., Jr & McCrae, R. R. (2004). Consensual validation of personality traits across cultures. *Journal of Research in Personality*, 38, 179–201.

DeCuyper, M., De Pauw, S., De Fruyt, F., De Bolle, M., & De Clercq, B. (2009) A Meta-Analysis of Psychopathy-, Antisocial PD- and FFM associations. *European Journal of Personality,* 23, 531-565.

De Fruyt, F., & Salgado, J. (2003). Applied personality psychology: Lessons learned from the IWO field. *European Journal of Psychology*, 17, 5123–5131.

De Fruyt, F., De Clercq, B., Miller, J., Rolland, J-P, Jung, S-C., Taris, R., Furnham, A., & Van Hiel, A. (2009). Assessing personality at risk in personnel selection and development. *European Journal of Personality*, 23, 51-69.

Doty, D. H., & Glick, W. H. (1998). Common methods bias: Does common methods variance really bias results? *Organizational Research Methods*, 1, 374-406.

Eysenck, H. (1947) Dimensions of Personality. New York: Praeger

Furnham, A., & Taylor, J. (2004). The dark side of behaviour at work: Understanding and avoiding employees leaving, thieving and deceiving. Hampshire: Palgrave MacMillan.

Furnham, A. (2008) Personality and Intelligence at Work. London: Routledge

Furnham, A. (2010) The Elephant in the Boardroom: The Psychology of Leadership Derailment. Basingstoke: Palgrave

Furnham, A. & Crump, J. (2005). Personality Traits, Types, and Disorders: An Examination of the Relationship Between Three Self-Report Measures. *European Journal of Personality*, 19, 167–184

Goldberg, L. R. (1990). An Alternative "Description of Personality": The Big-Five Factor Structure. *Journal of Personality and Social Psychology*, 59, 1216-1229

Hogan, R., Hogan, J. &, Roberts, B., W. (1996). Personality Measurement and Employment Decisions - Questions and Answers. *American Psychologist*, 51, 469-477.

Hogan, R. & Hogan, J. (2001a) Assessing Leadership: A View from the Dark Side. *International Journal of Selection and Assessment*, 9, 40-51.

Hogan, R., & Hogan, J. (2001b). Hogan development survey manual. Tulsa, OK: Hogan Assessment Systems.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102-138). New York: Guilford Press.

John, O.P. & Robins, R.W. (1993). Determinants of interjudge agreement on personality traits: the Big Five domains, observability, evaluativeness, and the unique perspective of the self, *Journal of Personality*, 61, 521-551.

Judge, T.A. & Bono, J.E. (2000). Five-Factor Model of Personality and Transformational Leadership. *Journal of Applied Psychology*, 85, 751-765.

Judge T.A., Bono, J.E., Ilies R., & Gerhadt M.W. (2002) Personality and Leadership: A Qualitative and Quantative Review. *Journal of Applied Psychology*, 87, 765-780

Kets de Vries, M. F. R. (2006) The Leader on the Couch – A Clinical Approach to Changing People and Organizations. Chichester: John Wiley and Sons Inc.

Mullins-Sweatt S. & Widiger, T.A. (2007). The Shedler and Westen Assessment Procedure from the perspective of general personality structure. *Journal of Abnormal Psychology*, 116, 618-23.

Livesley, J. W., Jackson, D. N. & Schroeder, M. L. (1992) Factorial Structure of Traits Delineating Personality Disorders in Clinical and General Population Samples. *Journal of Abnormal Psychology*, 101, 432-440.

Lynam, D. R. & Widiger, T. A. (2001) Using the Five-Factor Model to Represent the DSM-IV Personality Disorders: An Expert Consensus Approach. *Journal of Abnormal Psychology*, 110, 401-412.

Löckenhoff, C.E., Terracciano, A. Bienvenu, O.J., Patriciu, N.S, Nestadt, G., McCrae, R.R., Eaton, W.W., & Costa, Jr, P.T. (2008). Ethnicity, Education, and the Temporal Stability of Personality Traits In the East Baltimore Epidemiologic Catchment Area Study. *Journal of Research in personality*, 42, 577-598.

Martens, W.H.J. (1997). Psychopathy and Maturation. MD-thesis, Tilburg University, The Netherlands. Maastricht: Shaker Publishing.

Niitamo, P. (1997). PRF Personality Research Form, käsikirja. Helsinki: Psykologien Kustannus Oy.

PK5-persoonallisuustestin käsikirja (2006). Helsinki: Psykologien Kustannus Oy.

Plomin, R., & Caspi, A. (1999). Behavioral genetics and personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 251-276). New York: Guilford Press.

Saario, S. (2011). The Relationship Between General Therapeutic Orientation, Big Five Personality Traits, and Interpersonal Functioning in Substance Abuse Therapists: An Explorative Study. *Addictive Disorders & Their Treatment*. 10, 29-36

Salgado, J. F. (2003) Predicting Job Performance Using FFM and non-FFM Personality Measures. *Journal of Occupational and Organisational Psychology*, 76, 323-346.

Samuel, D.B., & Widiger, T.A. (2008). A meta-analytic review of the relationships between the five-factor model and DSM-IV-TR personality disorders: A facet level analysis. *Clinical Psychology Review*, 28, 1326-1342.

Shedler, J., & Westen, D. (2004). Dimensions of personality pathology: An alternative to the five-factor model. *American Journal of Psychiatry*, 161, 1743–1754.

Thompson, B. (1984): Canonical Correlation Analysis. Uses and Interpretation. Beverly Hills: SAGE.

Trull, T. J., & Durrett, C. (2005). Categorical and dimensional models of personality disorders. *Annual Review of Clinical Psychology*, 1, 355-380.

Ullrich, S., Farrington, D. P. & Coid, J. W. (2007) Dimensions of DSM-IV Personality Disorders and Life-Success. *Journal of Personality Disorders*, 21, 657-663.

Widiger T.A. & Costa P.T. (1994). Personality and personality disorders. *Journal of abnormal psychology*, 103, 78-91.

Widiger, T. A. & Samuel, D. B. (2005) Diagnostic Categories or Dimensions? A Question for the Diagnostic and Statistical Manual of Mental Disorders – Fifth Edition. *Journal of Abnormal Psychology*, 114, 494-504.

Widiger, T. A., Trull, T. J., Clarkin, J. F., Sanderson, C. & Costa, P. T., Jr (2002). A Description of the DSM-IV Personality Disorders with the Five-Factor Model of Personality. In P. T. Costa Jr & T. A. Widiger (eds.) *Personality Disorders and the Five-Factor Model of Personality* (2. edition). Washington, DC: American Psychological Association.

Widiger, T. A. & Trull, T. J. (2007) Plate Tectonics in the Classification of Personality Disorder. Shifting to a Dimensional Model. *American Psychologist*, 62, 71-83.

[INSERT APPENDIX HERE]

III

SELF-OTHER AGREEMENT ON PERSONALITY RATINGS IN MANAGER-SUBORDINATE RELATIONSHIPS AS A PREDICTOR OF OCCUPATIONAL WELL-BEING

by

Mikael Nederström, Adrian Furnham Kaisa Yli-Pärri, Elina Auvinen, & Laura Weis, 2016

Submitted manuscript.

Reproduced with kind permission by the authors.

Self-other agreement on personality ratings in manager-subordinate relationships as a predictor of occupational well-being

Abstract

Previous research has demonstrated that managers' self-reported personality has a significant impact on the organization and the well-being of its personnel. The managers (N=180) in a technology research organization provided self-reports of their personality, while their subordinates (N=1951) rated their superiors' personality and their personal occupational well-being. The results demonstrated that self-other agreement (SOA) between self-reports and subordinate ratings and the direction of the discrepancy were significant predictors of subordinates' occupational well-being, supporting earlier conceptualizations of self-awareness and leadership theories, such as authentic leadership. Higher SOA and a tendency towards self-reporting low trait values (when combined with higher other-ratings) predicted higher occupational well-being. Furthermore, SOA was a more accurate predictor compared to managers' self-reported personality on several facets of subordinate well-being. Implications and limitations are discussed.

Keywords: Personality, self-other agreement, management, leadership, self-awareness, occupational well-being

Previous research has shown great interest in identifying personality traits of a good leader. It has also established a close connection between leaders' self-reported personality and leadership performance (for a meta-analysis, see Judge, Bono, Ilies, & Gerhardt, 2002). Recently, however, increased interest has also been shown in observers' (subordinates, colleagues or peers) perceptions about their leaders' personality (Van Vugt, Hogan, & Kaiser, 2008; Furnham, Race, & Rosen, 2014). Connelly and Ones (2010) proposed that the perspective of others adds a great value to personality assessment.

Moreover, it seems that taking into account the congruence of self-reports and other-ratings (self-other agreement, SOA) can further contribute to the prediction of organizational outcomes. There is indeed an academic literature on self-awareness at work, as measured in part by self-other ratings, and its relationship to management success (Furnham & Baldry, 2000; Furnham & Bailey, 2005; Smither, London & Reilly, 2005). Consequently, many studies have turned to exploring the practical implications of SOA (cf. Connelly & Ones, 2010). Research has demonstrated a relationship between managerial performance and congruence in self-other ratings (for a review of the topic, see Fleenor, Smither, Atwater, Braddy, & Sturm, 2010; Atwater & Yammarino, 1997; Bass & Yammarino, 1991). One important criterion of managerial performance is occupational well-being in the organization (Ilies, Morgeson, & Nahrgang, 2005) which will be the dependent variable used in this study.

Church (1997) suggested that one explanation for the incremental validity of SOA can be found considering the concept of *self-awareness*. Atwater and Yammarino (1997) proposed that individuals who score high on self-awareness (i.e., individuals who know

themselves well and perceive themselves accurately) have more accurate self-ratings because of good self-observation and self-reflection skills. The ability to self-observe one's behavior or performance and compare these observations to feedback given, for example, from colleagues, can be understood as self-awareness (Wicklund, 1975). As such, SOA can be seen as an important way of operationalizing self-awareness (Fleenor et al., 2010).

One potential theoretical framework to connect SOA and subordinates' well-being can be found in the concept of *authentic leadership*. Luthans and Avolio (2003) defined authentic leadership "as a process that draws from both positive psychological capacities and a highly developed organizational context, which results in both greater self-awareness and self-regulated positive behaviors on the part of leaders and associates, fostering positive self-development" (p. 243). More recent definitions of authenticity include the four core components: self-awareness, relational transparency, internalized moral perspective and balanced processing (Walumbwa, Avolio, Gardner, Wernsing, & Peterson, 2008), while "self-awareness includes being aware of one's strengths and weaknesses as well as understanding one's emotions and personality" (Ilies et al., 2005, p. 378). It has been suggested that these four components influence both leader's *and* follower's well-being (Ilies, et al., 2005), Therefore, it is hypothesized that SOA is related to one core component of authentic leadership and consequently, to well-being.

The main purpose of the current study is to examine the relationship between leader personality and leadership outcomes by using both self- and other-ratings of personality. This was obtained by examining SOA and the direction of the discrepancy between the two types of ratings as predictors of the occupational well-being of subordinates. Based

on the idea that SOA on personality ratings can predict leadership outcomes, it is proposed that the direction of the discrepancy (overestimation vs. underestimation of traits) between ratings is likely to offer additional insights on leadership behavior.

Rating discrepancy and managerial performance

A leader who overestimates his/her skills is perceived very differently compared to a leader who underestimates his performance. Atwater and Yammarino (1997) distinguish in their model the two situations in which an individual perceives him/herself differently from the way others perceive him/her. Individuals in these discrepant categories were labeled either as *overestimators* (the individual's self-ratings are higher compared to others' ratings) or as *underestimators* (the individual's self-ratings are lower compared to others' ratings). Bass and Yammarino (1991) demonstrated that high self-ratings, when coupled with lower other-ratings (overestimation), can lead to arrogance, characterized by ignoring criticism and denying failures. This overestimation may be associated with poor performance. In turn, the performance of those whose self-ratings are lower than other-ratings (underestimation) could show a mix of positive and negative outcomes (Atwater, Waldman, Ostroff, Robie, & Johnson, 2005).

The literature has used a variety of metrics to assess SOA. The direction of discrepancy is particularly evident when adopting evaluative measures, such as 360-degree feedback[3] or performance ratings, where the higher pole is interpreted as more desirable than the lower pole of the scale. Interestingly, several studies have shown that similar desirability effects can be related to (seemingly neutral) personality traits (Nederström

---

3 The 360 measurement is a multi-source assessment feedback that comes from members of an employee's immediate circle in the workplace, e.g., superiors, colleagues and subordinates. This kind of feedback is utilized when there is a need to increase individuals' self-awareness and to develop their leadership behavior.

& Salmela-aro, 2014). John and Robins (1993) point out that the *evaluativeness* of the trait or personality attribute (i.e., whether a trait is considered socially desirable or undesirable: e.g., "empathic" versus "arrogant") varies from trait to trait. Similarly, some traits (e.g., Extraversion vs. Introversion) are considered more *leaderlike* or otherwise closely related to leadership emergence (Colbert et al., 2012). Therefore, it is plausible to assume that similar effects of over- and underestimation, which have already been established with other SOA measures, will also be found from personality ratings.

Although SOA in relation to personality is a widely-explored topic, the number of SOA studies on personality and its relations to behavioral outcomes is limited (Colbert et al., 2012; Connelly & Ones, 2010). More insight into the interrelations among these concepts is not only scientifically interesting but also has practical implications with regard to the design of workplace interventions.

Constructs of personality and rating context

One of the most widely used frameworks in SOA studies has been the Big Five framework (cf. Funder, 1999), including five global personality traits (Extraversion, Openness to Experience, Conscientiousness, Neuroticism, and Agreeableness) and their subscales (e.g., Costa & McCrae, 1992). Studies on the Big Five and leadership behavior have shown a consistent relation between leader personality and leadership style (Judge et al., 2002; Judge & Bono, 2000). However, most scholars would concede that the general Big Five framework does not exhaust the concept of personality (e.g., Block, 1995; Paunonen & Jackson, 2000). Furthermore, all personality judgments take place in a specific context (Bing, Whanger, Davison, & VanHook (2004). Antonakis, Avolio, and Sivasubramaniam (2003) demonstrated that the psychometric properties of leadership

instruments are affected by the context in which leadership is observed and evaluated. Taking into account this frame-of-reference effect has several benefits in predicting organizational outcomes. Use of a frame-of-reference that is conceptually relevant to the criterion reduces within-person inconsistency (Lievens, Corte, & Schollaert, 2008). This effect on personality scales may yield incremental validity above and beyond the more well-established, non-contextual personality measures (Bing, et al., 2004). Given the fundamental issue of context dependency in personality judgments, it is plausible to hypothesize that certain areas of personality are more exposed to observation in organizational settings than laboratory situations.

For these objectives, we used the Work Personality Inventory (WOPI) measuring personality from a specifically organizational angle (Leung & Zedeck, 2016; Nederström & Niitamo, 2010). This measure has been developed specifically for working life settings. The 14 WOPI scales are consistent with the Big Five framework, as described in detail in the measures section. The WOPI scales cover all Big Five traits: 1) Extraversion (Competitive Achievement, Leadership, and Inspiration), 2) Openness to Experience (Orientation, Perception, Thinking), 3) Conscientiousness (Ambiguity-change, Focused Achievement, Decision Making), 4) Emotional Stability (Optimism, Self-reflection) and 5) Agreeableness (Sociability, Empathy, Reliance).

Occupational well-being

Van Horn, Taris, Schaufeli, and Schreurs (2004) described a multidimensional model of occupational well-being in which well-being is conceptualized as a positive evaluation of aspects related to an individual's work. These researchers highlight the importance of studying multidimensional well-being instead of focusing only on affective dimensions.

This approach helps to gain a broader understanding of the nature, causes and consequences of occupational well-being. Consequently, we aimed at a multidimensional model of occupational well-being in this study.

Our model covered three areas of well-being: 1) superior support, 2) trust in management and 3) job satisfaction. Even not conceptually exhaustive, this three-factor model is compatible with earlier theoretical multidimensional frameworks and empirical findings. Superiors' support may be interpreted as an aspect of *social* well-being, whereas job satisfaction is considered to be as an aspect of *affective* well-being (Van Horn et al., 2004; Brough & Pears, 2004). Furthermore, trust in management as such has shown to be associated to several well-being outcomes: job engagement (Dirks & Ferrin, 2002), job satisfaction (Butler, Cantrell, & Flick, 1999; Dirks & Ferrin, 2002) and a level of employee turnover (Davis, Schoorman, Mayer, & Tan, 2000). Interestingly, authentic leadership has also been shown to increase team success by promoting trust in the group or organization (Clapp-Smith, Vogelgesang, & Avey, 2009). Therefore, this aspect of well-being is conceptually relevant to this study.

Aims and hypotheses

To the best of our knowledge, no research has been yet published between SOA on leader personality and occupational well-being of subordinates. In this context, subordinates' well-being and its relationships to 1) the overall magnitude of SOA, 2) the incremental validity of SOA over and beyond managers' self-reported personality, and 3) the direction of the discrepancy, when self- and other-ratings are not in agreement (overestimators vs. underestimators) will be investigated.

Atwater and Yammarino (1997) posit that when managers' self-ratings are in alignment with others' ratings it is more likely to be linked to more positive individual and organizational outcomes, such as occupational well-being of subordinates. Furnham and Stringfield (1994) found that managers rated by their superiors as more successful had fewer discrepancies between their self- and subordinates' ratings. Furthermore, SOA carries a conceptual relationship to self-awareness as an important component of authentic leadership promoting well-being (Ilies, et al., 2005). Therefore, we hypothesize that the overall magnitude of SOA is associated with the occupational well-being of subordinates.

Hypothesis 1: SOA in ratings is associated with occupational well-being: higher SOA leads to higher occupational well-being.

The magnitude of the relationship between self-reported personality and work-related outcomes has been moderate at best (Oh, Wang, & Mount, 2011). Some studies have suggested that the modest relationships between personality and leadership outcomes may be due to the still limited use of others' ratings of personality (Morgeson et al., 2007). Therefore, we propose that using SOA as a predictor[4] of well-being will explain more variance in well-being outcomes than managers' self-reports of personality.

Hypothesis 2: SOA explains more of the variance in subordinate occupational well-being than using managers' self-reported personality alone.

---

4 In the present study, we use the term "predictor" in a purely statistical sense. Therefore, this term does not imply causality in either direction.

We also propose that the direction of the discrepancy significantly predicts subordinate well-being: overestimating one's personality traits, particularly those related to "leader-like" behavior (Colbert et al., 2012) as well as Empathy (Brutus, Fleenor & McCauley, 1999) decreases occupational well-being outcomes. This effect will be strongest on Extraversion-related traits (Competitive Achievement, Leadership, Inspiration) and Empathy, as these traits have been demonstrated to associate with positive other-ratings (Brutus et al., 1999; Judge & Bono, 2000; Skakon, Nielsen, Borg, & Guzman, 2010) and desired leader traits (Nichols & Cottrell, 2014). Therefore, we expect overestimation of these traits to be more problematic than underestimation of the same traits.

Hypothesis 3: The direction of the discrepancy matters most for traits related to Extraversion (Competitive Achievement, Leadership, Inspiration) and Empathy. Managers' overestimation (in contrast to underestimation) of these traits leads to subordinates' lower occupational well-being.

METHODS

Participants and procedure

The target managers (N=180) were individuals with the job titles of team leaders (80%) and technology managers (20%), all of whom had a supervisory responsibility for leading over their teams. More than half of the managers were males (70 % males, 30 % females). All the managers filled in the WOPI self-report questionnaire (Leung & Zedeck, 2016) online. Subordinates rated their managers individually and ratings of managers with several subordinates were averaged across the responding subordinates for better reliability. Team sizes varied from 1 to 16 subordinates (median 4, while 80 % of the teams had 2-6 subordinates). Overall, for the 180 managers, a total of 1 951 indi-

vidual personality ratings, expressed as 180 team values, formed the other-ratings data set.

Data were collected by a Finnish HR consulting company (Psycon Corp.) in a major government funded technology research organization. The present data collection was carried out as a part of a broader personnel survey in which personnel received an online invitation to participate. Participants represented different organizational levels and units. Although total sampling was used, participation in the study was voluntary. Of the 2 780 persons working in the organization, 2 067 participated in the personnel survey, yielding a response rate of 74.4%.

The research setting is described in more detail in Figure 1. First, both managers and subordinates filled in the personnel survey, although the present study concerns was limited to the subordinates' responses. In conjunction with the survey, the managers were asked to complete a work-related personality inventory, WOPI. The subordinates were in turn asked to evaluate their manager's personality by rating the same set of personality dimensions included in WOPI inventory.

[INSERT FIGURE 1 ABOUT HERE]

Figure 1. The research setting.

Measures

Self-reported personality. WOPI (Work Personality Inventory) is a standardized self-report questionnaire on personality with 14 scales deemed important for work settings (Leung & Zedeck, 2016). The 224 respondent-descriptive item statements are answered

on a dichotomous (true-false) scale, yielding a trait score range of 0 to 16 (see Appendix 1 for example items). The validation study (Nederström and Niitamo, 2010) reports scale reliabilities collected from an occupational sample. All the scale reliabilities are comparable to those found in established personality inventories: (KR-20/re-test reliability in parenthesis): Focused Achievement (.76/.71), Competitive Achievement (.78/.83), Leadership (.80/.89), Inspiration (.74/.86), Sociability (.74/.84), Empathy (.75/.85), Reliance (.69/.76), Orientation (.77/.89), Perception (.78/.80), Thinking (.81/.80), Decision making (.77/.91), Ambiguity-change (.71/.85), Optimism (.81/.85), and Self-reflection (.78/.73). Concurrent and predictive validation has been shown in studies investigating the convergence of peer and spousal ratings of personality, and construct validity in studies on the relations between WOPI and more established personality inventories as well as miscellaneous other measures, such as values, learning styles, organizational culture, etc. (Leung & Zedeck, 2016; Nederström & Niitamo, 2010). Factor analysis demonstrates that the altogether 14 scales are readily interpretable in Big Five terms, connecting the present findings to prior studies on SOA. See Appendix 2, Table 1 for the confirmatory factor analysis (CFA) matrix.

Others' ratings. The same set of 14 work-related personality dimensions were rated by the managers' direct subordinates (N=1 951). The subordinates rated their managers' personality on an observer-rating form with a 1-10 Likert scale. For each dimension, short verbal descriptions of the behavioral markers of low- and high-scale points were given on the rating form. (e.g., Leadership: 1 = "Not willing to lead, withdrawing, unwilling to take the initiative, not actively in charge", 10 = "Willing to lead, directive, leading the way, initiating). An interrater agreement analysis was conducted employing intraclass correlation coefficients (ICC) to explore the reliability of team level aggre-

gates. The ICC2 analysis demonstrated values from .57 (mean) to .66 (median). These values are regarded as “fair” (.40-.59) or “good” (.60-.75) (Cicchetti, 1994). Thus, it seems that aggregating teams’ personality judgments was statistically justified.

Personnel survey and occupational well-being. The occupational well-being of subordinates was assessed through the 77-item personnel survey. The survey themes ranged from evaluating one’s own job-related aspects and the functioning of the team to assessing the management style of one’s own manager. All the items were answered on a five-point Likert scale where the opposite ends of the scales were described verbally (e.g., “Team spirit in our team is 1 = bad … 5 = good”; “My workload is 1 = excessive … 5 = reasonable”; “I feel enthusiasm and joy at work 1 = very rarely … 5 = almost always”). The respondents were instructed to use the whole range of the response scale. The answers to the 77 questions were analyzed by explorative factor analysis using Principle Axis Factoring as an extraction method and Promax with Kaiser Normalization rotation to explore the underlying factors.

For the present purposes, only the theoretically relevant factors were examined and other irrelevant (such as organizational practices, strategy and customer relations, etc.) omitted. This analysis helped us to construct three reliable, interpretable and relatively independent (correlation between variables $r < .49^{**}$) outcome scales for occupational well-being:

1) Satisfaction with one’s direct superior and team spirit (Superior Support, sample item: “Superior's time to listen to subordinates”);

2) Trust in the management, sharing information and involving employees in the decision-making (Trust in Management, sample item: "Trust in the management of the function");

3) Satisfaction with one's work and development opportunities (Job Satisfaction, sample item: "Enjoyment of the work").

The internal consistencies (Cronbach's alpha) of the scales were: Supervisor Support .94 (12 items), Trust in Management .92 (10 items) and Job Satisfaction .88 (6 items). The CFA matrix is presented in Appendix 2, Table 2.

Statistical methods

We used polynomial regression analysis and response surface tests to test our research hypotheses on SOA and its outcomes. SOA may be operationalized in several ways, and the one used in the present study is considered a self-insight approach (Fleenor et al., 2010). In this method, the self-ratings of a target individual are compared to the relevant others' ratings of the target (Kwan, John, Kenny, Bond, & Robins, 2004). This is considered as the best method for analyzing this kind of data as it overcomes methodological problems (e.g., those related to using difference scores) found in earlier research on SOA (Fleenor et al., 2010). The procedure consisted of polynomial regression analysis followed by response surface analysis, which graphs the polynomial regression outcomes in three-dimensional space (e.g., Edwards, 1994; Edwards & Parry, 1993). Such a procedure allows investigating how the two predictor variables (self- and other-ratings of personality) and especially, the discrepancy between them relate to the outcome variable (the occupational well-being of subordinates) (Shanock, Baran, Gentry, Pattison, &

Heggestad, 2010). When the two ratings are kept separate, computation of higher-order terms enables examination of linear and non-liner relations. Because the predictor variables were not measured on the same scale (WOPI self-report scores 0-16 vs. other-ratings 1-10), we transformed the variables to a standardized scale which places them on a common metric (Shanock et al., 2010).

Polynomial regression. In polynomial regression analysis, the dependent variable (Z) is regressed on each independent variable (X and Y), the interaction between the independent variables (X × Y) and the squared terms for the independent variables (X2 and Y2). Thus, three new variables were created for each analysis: the square of the subordinates' rating, the cross-product of the subordinates' and manager's rating, and the square of the managers' rating. Next, polynomial regression analyses were conducted by regressing the dependent variables on the independent variables, the product of the independent variables, the subordinate rating squared, and the manager rating squared terms.

Response surface analysis. In response surface analysis, the slope and the curvature of two lines graphically illustrate the phenomenon (Shanock et al., 2010; see Figure 2). First, the slope of the line of perfect agreement (X = Y) illustrates how agreement between the independent variables (subordinates' rating and manager's rating) relates to the dependent variable (i.e., occupational well-being of subordinates). In addition, a curvature along the line of perfect agreement shows whether or not the relationship between the in-agreement ratings and the dependent variable is nonlinear. The line of incongruence (X = -Y) represents a situation where X and Y are not in agreement. A significant curvature along the line shows how the degree of discrepancy between the in-

dependent variables relates to the dependent variable. Furthermore, the slope along the line of incongruence indicates the direction of the discrepancy, which shows that the dependent variable can be influenced more when the discrepancy is in a given direction (X > Y or X < Y). This helps to compare discrepancies between two variables in predicting outcomes and the degree to which one type of discrepancy (X > Y) better predicts that outcome than another type (Y > X). We examined all combinations with three response surface test values (a1, a3, a4). The results allow three questions to be formulated: 1) How does the agreement between the independent variables relate to the dependent variable? 2) How does the degree of discrepancy between the independent variables relate to the dependent variable? and 3) How does the direction of the discrepancy between the independent variables relate to the dependent variable (Shanock et al., 2010)?

RESULTS

Preliminary analyses

The relationships between subordinates' occupational well-being, managers' self-reported personality and subordinates' ratings of the managers' personality are presented in Table 1. First, we examined the general magnitude of SOA on all WOPI dimensions. Self-other agreement was statistically significant on 10 of the 14 personality traits assessed. As shown in Table 1, SOA between managers and their subordinates ranged from r = .47 to r < .10. The difference in judgability between traits was obvious, as previous research has shown (Funder, 1999), while the traits easiest to judge (Decision Making and Sociability, r = .47 and r = .34) yielded substantially greater correlations than Reliance, Focused achievement and Self-reflection (r < .10). Contrary to our expectations and previous research (Funder, 1999; Funder & Dobroth, 1987), Extraversion

related traits (Competitive Achievement, Leadership, and Inspiration) were not the most observable traits but instead were of average magnitude.

[INSERT TABLE 1 ABOUT HERE]

Second, we estimated the relationship between managers' self-reported personality and subordinates occupational well-being. As can be seen from Table 1, only two statistically significant correlations were observed between the occupational well-being outcomes and managers' personality. Self-reported Leadership predicted both Trust in Management and Job Satisfaction. Although statistically significant, the association between Leadership and Job Satisfaction was only moderate ($r = .15$, $p<.01$). The other-ratings showed several significant relationships with a greater effect on subordinate well-being. The correlations between well-being outcomes and the other-rated personality traits reached .75 at their highest (Superior Support and Empathy/Inspiration).

Primary analyses

We then examined the relationships associated with Hypotheses 1–4. For each of the 14 traits and 3 outcomes, we performed regressions using the following equation:

$$Z = b0 + b1X + b2Y + b3X2 + b4XY + b5Y2 + e$$

where Z is the occupational well-being of subordinates, X stands for the superior's self-reported trait and Y represents other-ratings of the trait. We also calculated and tested the relevant slopes/curves following procedures discussed in Shanock et al. (2010).

First, we examined the slope of the line of agreement (X = Y). In light of our hypotheses, this value is expected to be positive, indicating a perfect line of agreement. In addition, the curvature along the line of incongruence (X=-Y) shows the effect of the discrepancy. If this value is negative, it indicates that as the discrepancy between self-reports and other-ratings increases, the outcome decreases. Thus, the value for a1 should be positive and the value for a4 negative.

As illustrated in Table 2, the results differed depending on the trait and the outcome under consideration. Hypothesis 1 suggested that the outcome is highest when self-reported and the other-ratings are in agreement and lowest when these ratings are discrepant. Consistent with our H1, higher SOA predicted higher well-being, while a higher discrepancy predicted lower well-being.

[INSERT TABLE 2 ABOUT HERE]

To estimate the relative importance of self-reports of personality and SOA as predictors of well-being, we defined the "perfect" SOA conditions (Shanock et al., 2010) for each trait: a positive a1 combined with a negative a4. Table 2 shows that SOA offered a significantly wider perspective on well-being outcomes than self-reported personality. While only one self-reported trait (Leadership) reached two significant correlations with the well-being outcomes (Trust in Management and Job Satisfaction; see Table 1), the surface response analysis of SOA revealed that several traits had a significant positive a1 combined with a significant and negative a4. Superior support was significantly predicted by SOA on seven traits (Competitive Achievement, Leadership, Inspiration, Thinking, Ambiguity-change, Optimism and Sociability), Trust in Management on two

traits (Competitive Achievement, Inspiration) and Job Satisfaction on three traits (Competitive Achievement, Perception, Sociability). This indicates the strong predictive value of SOA compared to self-reports on several well-being variables, thereby supporting H2.

Finally, we examined the direction of the discrepancy, indicated by the slope of the line of incongruence. A statistically significant slope indicates that the two types of discrepancy (i.e., X > Y, Y > X) differentially predict the outcome variable. A significant negative a3 indicates that well-being is higher when the other-rating is higher than the self-report. In the present results, all the significant slopes were negative. This finding suggests that overestimators of the WOPI dimensions are more fatal for the well-being of their subordinates than underestimators. This effect appears particularly for one facet of Extroversion (Inspiration) and for Empathy, which is consistent with H3. Other facets of Extraversion were also significant, as predicted, but the effect sizes were considerably lower than those for Inspiration and Empathy. Openness to Experience and Conscientiousness revealed similar effects, although the effect sizes were also lower.

DISCUSSION

Our findings extend existing literature in three important ways: 1) they demonstrate that SOA on leaders personality is an important variable when investigating employee occupational well-being, 2) SOA can bring incremental validity above and beyond self-reported personality in predicting well-being outcomes and 3) the direction of discrepancy (overestimators vs. underestimators) has an independent effect on employee well-being. To the best of our knowledge, no previous SOA research has investigated the connection between leaders' personality and subordinates' occupational well-being.

The main results of the present study seemed to be consistent with the previous findings on SOA and its beneficial effects on leadership outcomes (e.g., Fleenor et al., 2010). As hypothesized, the results revealed that as the discrepancy between self-reports and other ratings increased, well-being decreased. Although the evidence suggests that SOA has a substantial relationship with well-being, the exact mediating process by which the effect occurs remains somewhat unclear. We offer some explanations for underlying mechanisms of SOA.

One suggested mediator for the SOA has been self-awareness. This explanation gives support to the effect of authentic leadership, because authenticity, in part, refers to awareness of one's strengths and weaknesses, ambiguities, inconsistencies and limits of self-knowledge and acting in tune with one's true self (Luthans & Avolio, 2003). According to Atwater et al. (2005), the theoretical rationale underlying self-awareness and its leadership outcomes stems from four aspects of leadership behavior. First, a leader who does not recognize the discrepancy between self-reports and other-ratings will not see the need to alter his/her behavior to meet with others' expectations. Thus, the leader will continue to take on assignments that are beyond his/her competence and will eventually fail. Second, agreement in ratings shows a stronger relationship with positive outcomes since a realistic self-view helps a leader to correct mistakes and align his/her behavior with the demands of the organization. Third, ratings that are in agreement are preferable because they help to promote mutual understanding between, for example, leader and subordinates. Fourth, self-awareness can be used as an indicator of emotional intelligence, which, in turn, has been associated with effective leadership.

Which mechanisms mediate the relationships between discrepancy and lower well-being? We offer three plausible explanations. First, subordinates may have a fairly accurate judgement of their superior, yet the superior views him/herself differently. In this case judgability of traits may be present, allowing for valid subordinate judgements. Yet, due to a low level of self-knowledge (understanding one's own strengths and weaknesses) and self-awareness (capacity to reflect/introspect), the superior perceives him/herself differently. This, in turn, brings about low levels of employee work satisfaction, as the theory of authentic leadership predicts (Luthans & Avolio, 2003).

Second, the superior may have a fairly accurate judgement of him/herself, yet the subordinates view him/her differently. This may be the case if, for example, the leader is not accessible or visible and high discrepancy scores are a result of this. The negative impact of low levels of visibility has been well established in previous literature (Furnham, Race & Rosen, 2014). In the current study we did not control for frequency of interaction, which appears to be an important covariate to consider.

Third, there may be a lack of self-consistency in behavior. A leader can be self-aware but is missing components of authenticity, namely authentic behavioral and relational orientation (Ilies, et al., 2005), which is shown in lack of SOA. Although leaders may know themselves rather well, it is vital for them to show ''consistency between their values, beliefs, and actions'' (Walumbwa et al., 2008, p. 93). Subordinates need to be certain about their superior's behavior, i.e., they need a sufficient level of predictability. This enables subordinates to create a ''realistic social relationship'' (Avolio, Gardner, Walumbwa, Luthans, & May, 2004, p. 810) which is positively linked to job satisfaction and commitment (Peus, Wesche, Streicher, Braun, & Frey, 2012). As Kouzes and

Posner (2002) point out, alignment between a superior's values and his/her behaviors produces credibility and authenticity. This explanation gives a theoretical rationale for differentiating self-awareness from SOA, even these concepts have been sometimes used interchangeably (Sosik, 2001).

The results demonstrated that SOA can bring a significant incremental validity above and beyond self-reported personality. While only one self-reported trait (Leadership) showed two significant correlations with well-being outcomes (Trust in Management and Job Satisfaction), the other-ratings showed several significant relationships with the well-being outcomes. This result is consistent with previous findings (Colbert et al., 2012; Oh et al., 2011), where observer ratings of personality were more strongly related to leadership outcomes than self-reports. Colbert, Judge, Choi, and Wang (2012) suggest that there are two reasons why observer ratings can be expected to contribute to the prediction of work-related outcomes. First, self-reports and observer ratings are able to capture unique information on the target. Second, both sources of information are likely to predict unique variance in the outcomes as both personality and outcomes are based on observer ratings of trait expression.

In terms of the personality traits, SOA with regard to Extraversion (Competitive Achievement, Leadership, Inspiration), was most strongly related to the occupational well-being of subordinates. This outcome was not expected and it seems that personality traits are not equal as SOA predictors. It was interesting to note that the Extraversion-related traits were not the most observable traits but instead were of average magnitude. Nonetheless, they were the best SOA predictor of well-being. This supports the previous findings that Extraversion is an important trait in leadership behavior, emergence

and leaderlikeness (Colbert et al., 2012; Offermann, Kennedy, & Wirtz, 1994) *an sich*, and this importance cannot be accounted solely by its visibility.

It proved to be beneficial for subordinates' well-being that their assessments of their managers' personality clustered towards the right pole of the assessment scale for certain traits, even if the manager's self-assessment tended towards the left pole of the assessment scale. More specifically, when subordinates see the manager as "leaderlike" (facets of Extraversion in WOPI) but the manager disagrees, the fact that the manager might underestimate his/her leadership traits does not endanger the well-being of subordinates. This effect appears in particular for one facet of Extroversion (Inspiration) and Empathy, as predicted, when using Supervisor Support as a criterion. This finding suggests that underestimating one's own Empathy or Extraversion is less fatal for well-being outcomes than overestimating these traits (i.e., humility is better than hubris and overestimation of Empathy may turn out to be "knowing better" the needs of others). In addition, traits related to extremely high self-reported Extraversion combined with a low level of Empathy (or Agreeableness) can lead to narcissism and, thus to lack of self-awareness (Nederström & Furnham, 2012). In line with this, Bass and Yammarino (1991) demonstrated that high self-ratings coupled with lower other-ratings may lead to arrogance and narcissism, manifested by ignoring criticism and denying failures.

The methods chosen for the statistical analyses (polynomial regression and response surface tests) can be considered as a technical strength of the present study, as these methods are widely accepted and possess sophistication in investigating the complex topic of SOA and its relation to different outcomes (Colbert, et al., 2012; Fleenor et al., 2010; Shanock et al., 2010). Had we conducted traditional moderated regression analy-

sis alone, we might have concluded that the relationship between SOA and well-being depends in part on self-reports and that others' ratings are more important for high well-being than self-reports. However, the response surface analysis yielded more detailed and convincing information on how combinations of the personality variables can affect well-being outcomes.

Limitations and future research

We used a cross-sectional research design, which calls for caution when drawing conclusions about causality. Many previous studies (e.g., Nichols & Cottrell, 2014) have implied that high SOA may be expected to lead to positive outcomes. Sosik and Megerian (1999) suggested that managerial SOA moderates relationships between aspects of emotional intelligence, transformational leadership behavior and managerial performance. Yet it is theoretically plausible to assume that a positive relationship between a leader and a subordinate will lead to increased interaction, which, in turn, will lead to a better knowledge and predictability of the other person and evidenced as higher SOA. Therefore, the direction of causality and the effect of mediators should be viewed as tentative.

In the present sample the outcome variable (well-being of subordinates) and the other-ratings were not obtained from independent sources. Thus, it is possible that *common method variance* at least partially explains the strong associations found between the observer ratings and the leadership outcomes. The limitations of our sample prevented us from using relative weight analysis (Johnson, 2000) or similar analytical methods, which may have revealed the proportionate contribution of each rating type in explaining the total variance of the well-being outcomes (Connelly & Ones, 2010). It would be

useful to adopt less subjective measures of well-being and performance that focus on outcomes of competent leadership behavior. This would help obtaining data from truly independent sources.

We do not know of studies that have used longitudinal research designs to investigate this topic. However, a longitudinal design could be useful in further research, because managerial implications may vary depending on the various mediators of SOA. For example, possible changes in SOA as a result of training could be examined. More specifically, it would be of great interest to see if training (for example, trying to reach a consensus on a manager's personality) implemented among the manager and his/her subordinates' increases agreement on work-related personality and the level of self-awareness, and if it has associations with the occupational well-being of subordinates. After all, knowing one's strengths and shortfalls is fundamental to any adaptation or development.

Given that our analysis demonstrates evidence of significant relationships between SOA on certain personality traits and hypothesized outcomes, we suggest that personality (in addition to 360 degree or performance measures) should be recognized as an important variable regarding SOA implications. Moreover, for the concept of self-awareness to be more useful, additional research is needed to clarify the distinctions between different types of discrepancy and personality traits. For example, both over- and underestimators of Extraversion are lacking self-awareness, but the organizational outcomes may be completely different, as our results demonstrated.

REFERENCES

Antonakis, J., Avolio, B.J., & Sivasubramaniam, N. (2003). Context and leadership: An examination of the nine-factor full-range leadership theory using the Multifactor Leadership Questionnaire. *The Leadership Quarterly*, *14*, 261-295.

Atwater, L., & Yammarino, F. (1997). Self-other rating agreement: A review and model. Research in *Personnel and Human Resources Management, 15*, 121–174.

Atwater, L. E., Waldman, D. A., Robie, C., Ostroff, C., & Johnson, J. (2005). Self-other agreement: Comparing its relationship with performance in U.S. and Europe. *International Journal of Selection and Assessment, 13*, 25-40.

Avolio, B. J., Gardner, W. L., Walumbwa, F. O., Luthans, F., & May, D. R. (2004). Unlocking the mask: A look at the process by which authentic leaders impact follower attitudes and behaviors. *The Leadership Quarterly*, *15*, 801-823.

Bass, B.M., & Yammarino, F.J. (1991). Congruence of self and others' leadership ratings of Naval Officers for understanding successful performance. *Applied Psychology: An International Review, 40,* 437-454.

Bing M.N., Whanger J.C., Davison, H.K., & VanHook, J.B (2004). Incremental validity of the frame-of-reference effect in personality scale scores: a replication and extension. *Journal of Applied Psychology, 89*, 150-157.

Block, J. (1995). A Contrarian View of the Five-Factor Model. *Psychological Bulletin, 117*, 187–215.

Brough, P., & Pears, J. (2004). Evaluating the influence of the type of social support on job satisfaction and work related psychological well-being. *International Journal of Organisational Behaviour, 8*, 472-485.

Brutus, S., Fleenor, J.W., & McCauley, C.D. (1999). Demographic and personality predictors of congruence in multi-source ratings. *Journal of Management Development,* 18, 417-435.

Butler, J.K., Cantrell, R.S., & Flick, RJ. (1999). Transformational leadership behaviors, upward trust, and satisfaction in self-managed work teams. *Organization Development Journal*, 17, 13-28.

Church, A. H. (1997). Managerial self-awareness in high-performing individuals in organizations. *Journal of Applied Psychology*, *82*, 281.

Clapp-Smith, R., Vogelgesang, G. R., & Avey, J. B. (2009). Authentic leadership and positive psychological capital: The mediating role of trust at the group level of analysis. *Journal of Leadership and Organizational Studies, 15*, 227–240.

Colbert, A.E, Judge, T.A., Choi, D., & Wang, G. (2012). Assessing the trait theory of leadership using self and observer ratings of personality: The mediating role of contributions to group success. *The Leadership Quarterly*, *23*, 670-685.

Connelly, B. S., & Ones, D. S. (2010). Another perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*, 1092–1122.

Costa, P. T., & McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources.

Davis, J.H., Schoorman, F.D., Mayer, R.C, & Tan, H.H. (2000). The trusted general manager and business unit performance: Empirical evidence of a competitive advantage. *Strategic Management Journal, 21*, 562-576.

Dirks, K.T., & Ferrin, D.L. (2002). Trust in leadership: Meta-analytic findings and implications for research and practice. *Journal of Applied Psychology, 87*, 611-628.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes, 58*, 51–100.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36*, 1577–1613.

Fleenor, J. W., Smither, J. W., Atwater, L. E., Braddy, P. W., & Sturm, R. E. (2010). Self-other rating agreement in leadership: A review. *The Leadership Quarterly, 21,* 1005–1034.

Fletcher, C., & Baldry, C. (2000). A study of individual differences and self-awareness in the context of multi-source feedback. *Journal of Occupational and Organizational Psychology, 73,* 303-319.

Fletcher, C., & Bailey, C. (2003). Assessing self-awareness. *Journal of Managerial Psychology, 18,* 395-404.

Funder, D.C. (1999). Personality Judgment: A Realistic Approach to Person Perception. London: Academic Press.

Funder, D. C., & Dobroth, K. M. (1987) Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52*, 409-418.

Furnham, A., & Stringfield, P. (1998). Congruence in job-performance ratings: A study of 360 feedback examining self, manager, peers, and consultant ratings. *Human Relations, 51*, 517–530.

Furnham, A., Race, M-C., & Rosen, A. (2014). Emotional intelligence and the Occupational Personality Questionnaire (OPQ). *Frontiers in Psychology, 5,* 1-9.

Ilies, R., Morgeson, F. P., & Nahrgang, J. D. (2005). Authentic leadership and eudaemonic well-being: Understanding leader–follower outcomes. *The Leadership Quarterly, 16*, 373–394.

John, O.P., & Robins, R.W. (1993). Determinants of interjudge agreement on personality traits: the Big Five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61*, 521-551.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research, 35*, 1–19.

Judge, T. A., & Bono, J. E. (2000). Five-factor model of personality and transformational leadership. *Journal of Applied Psychology, 85*, 751–765.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology, 87*, 765–780.

Kouzes, J. M., & Posner, B. J. (2002). *Leadership challenge (3rd ed.).* San Francisco: Jossey-Bass.

Kwan, V. S., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: an interpersonal approach. *Psychological Review, 111*, 94–110.

Leung, S.A., & Zedeck, S. (2016). Reviews of the Work Personality Inventory. In Carlson, J.F., Geisinger, K.F & Jonson, J.L. (Eds.) The Twentieth Mental Measurements Yearbook. Buros: University of Nebraska Press, Nebraska.

Lievens F., De Corte W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology, 93*, 268-279.

Luthans, F., & Avolio, B. (2003). Authentic leadership: A positive development approach. In K. S. Cameron, J. E. Dutton, & R. E. Quinn (Eds.), *Positive organizational scholarship* (pp. 241–258). San Francisco: Berrett-Koehler.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60,* 683–729.

Nederström, M., & Furnham, A. (2012). The relationship between the FFM and personality disorders in a personnel selection sample. *Scandinavian Journal of Psychology, 53*, 5, 421–429.

Nederström M., & Niitamo P. (2010). Construction and validation of a work personality inventory. Helsinki University of Technology, *Department of Industrial Engineering and Management*, Report 2010/1.

Nederström, M., & Salmela-Aro, K. (2014). Self-other agreement of personality judgements in job interviews: exploring the effects of trait, gender, age and social desirability. *Scandinavian Journal of Psychology, 55*, 520-526.

Nichols, A.L., & Cottrell, C.A. (2014). What do people desire in their leaders? The role of leadership level on trait desirability. *The Leadership Quarterly, 25*, 711-729.

Offermann, L. R., Kennedy, J. K., & Wirtz, P. W. (1994). Implicit leadership theories: Content, structure, and generalizability. *The Leadership Quarterly, 5*, 43–58.

Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five-factor model of personality traits: a meta-analysis. *Journal of Applied Psychology, 96*, 762-773.

Paunonen, S. V., & Jackson, D. N. (2000) What Is Beyond the Big Five? Plenty! *Journal of Personality, 68*, 821-835.

Peus, C., Wesche, J., Streicher, B., Braun, S., & Frey, D. (2012). Authentic leadership: An empirical test of its antecedents, consequences, and mediating mechanisms. *Journal of Business Ethics*, *107*, 331-348

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology, 25,* 543–554.

Skakon, J., Nielsen, K., Borg, V., & Guzman, J. (2010). Are leaders' well-being, behaviours and style associated with the affective well-being of their employees? A systematic review of three decades of research. *Work & Stress, 24*, 107–139.

Smither, J., London, M., & Reilly, R. (2005). Does performance improve following multisource feedback? A theoretical model, meta-analysis and review of empirical findings. *Personnel Psychology, 58,* 33-66

Sosik, J. J. (2001). Self-Other Agreement on Charismatic Leadership Relationships with Work Attitudes and Managerial Performance. *Group & Organization Management*, *26*, 484–511.

Sosik, J. J., & Megerian, L. E. (1999). Understanding leader emotional intelligence and performance the role of self-other agreement on transformational leadership perceptions. *Group & Organization Management, 24*, 367-390.

Van Horn, J. E., Taris, T. W., Schaufeli, W. B., & Schreurs, P. J (2004). The structure of occupational well-being: A study among Dutch teachers. *Journal of Occupational and Organizational Psychology, 77,* 365–375.

Van Vugt M., Hogan R., & Kaiser R.B. (2008). Leadership, followership, and evolution: Some lessons from the past. *American Psychologist, 63*, 182–196.

Walumbwa, F., Avolio, B., Gardner, W., Wernsing, T., & Peterson, S. (2008). Authentic Leadership: Development and Validation of a Theory-Based Measure. *Journal of Management, 34*, 89-126.

Wicklund, R.A. (1975). Objective self-awareness. *Advances in Experimental Social Psychology, 8,* 233–275.

Table 1. Descriptive statistics and correlations among study variables (N=180).

	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1. Competition - S	7,58	3,94																														
2. Competition - O	6,43	1,28	.21**																													
3. Leadership - S	9,99	3,41	.39**	.25**																												
4. Leadership - O	6,33	1,41	-.04	.48**	.26**																											
5. Inspiration - S	8,43	3,53	.43**	.24**	.59**	.14																										
6. Inspiration - O	6,82	1,62	.14	.41**	.16*	.43**	.19*																									
7. Orientation - S	6,17	3,27	.09	.10	.00	-.04	.16*	.10																								
8. Orientation - O	5,60	1,32	.23**	.23**	.01	.07	.19*	.42**	.24**																							
9. Perception - S	8,74	3,75	.14*	-.06	-.16*	-.07	-.03	.04	.44**	.18*																						
10. Perception - O	5,64	1,19	.25**	.35**	.11	.12*	.20**	.29**	.25**	.35**	.21**																					
11. Thinking - S	7,43	3,32	-.01	.07	-.11	-.03	.07	.17*	.53**	.21**	.07	.12																				
12. Thinking - O	7,43	3,32	.22**	.18**	.12	.07	.26**	.49**	.22**	.64**	.17*	.23**	.30**																			
13. Ambiguity-change - S	8,60	3,54	.00	.17*	.26**	0.10	.31**	.16*	.30**	.15*	.05	.16*	.17**	.30**																		
14. Ambiguity-change - O	6,05	1,49	.19*	.41**	.18*	.31**	.28**	.64**	.13	.54**	-.04	.31**	.19*	.58**	.26**																	
15. Focused Achievement - S	6,94	3,34	.18**	-.05	-.08	-.14	-.16*	-.15*	.00	-.14	.21**	-.10	-.16*	-.20*'	-.41*'	-.17*																
16. Focused Achievement - O	7,28	1,26	.07	.57**	0.14	.51**	.05	.45**	-.01	.08	-.07	.15**	-.01	.06	.02	.32**	.01															
17. Decision Making - S	8,66	3,59	.15*	.33**	.42**	.23**	.36**	.23**	.25**	.23**	-.18*'	.22**	.32**	.34**	.45**	.36**	-.35*'	.11														
18. Decision Making - O	5,85	1,75	.15*	.47**	.23**	.45**	.30**	.48**	.15*	.45**	-.03	.24**	.19**	.52**	.15*	.65**	-.24*'	.28**	.47**													
19. Optimism - S	8,80	3,57	.15*	.21**	.46**	.19**	.44**	.13	.01	.14	-.07	.10	-.13	.12	.41**	.26**	-.21*'	.18*	.36**	.28**												
20. Optimism - O	6,68	1,40	.11	.50**	.15*	.38**	.28**	.54**	.18*	.53**	-.10	.33**	.14	.51**	.16*	.73**	-.20*'	.33**	.34**	.66**	.27**											
21. Self-reflection - S	8,55	3,13	-.14*	-.04	-.10	-.09	-.12	-.13	-.07	-.11	-.03	-.11	-.22*'	-.18*	-.04	-.19*	.22**	.10	-.13	-.21*'	.29**	-.17*										
22. Self-reflection - O	4,57	1,17	.17*	.19**	-.04	-.06	.19*	-.11	.09	.24**	-.03	.14*	.07	.27**	.06	.16**	-.05	-.14*	.20**	.26**	.10	.31**	-0.14									
23. Sociability - S	8,09	3,31	.16*	.17*	.30**	.20**	.45**	.23**	.11	.21**	-.18*'	.11	.15*	.27**	.24**	.25**	-.13	.10	.23**	.24**	.27**	.25**	.03	.11								
24. Sociability - O	6,82	1,83	.15*	.34**	.15	.37**	.17*	.70**	.14	.40**	.02	.15**	.23**	.47**	.20**	.57**	-.16*	.36**	.20**	.48**	.18*	.52**	-.12	.05	.34**							
25. Empathy - S	8,80	3,41	-.13*	-.01	-.09	-.09	.03	.17*	.25**	.16*	.04	.11	.32**	.17*	.01	.16	.07	.00	.10	.04	-.01	.10	.19**	-.10	.39**	.23**						
26. Empathy - O	6,82	1,63	.06	.13*	-.03	.26**	.00	.74**	.10	.27**	.04	.13*	.15*	.34**	.05	.47**	-.07	.36**	.03	.25**	-.04	.39**	-.13	-.21*'	.14	.66**	.20**					
27. Reliance - S	8,03	3,25	-.07	-.01	-.21*'	.02	-.03	.01	-.01	.00	-.09	.00	.20**	-.04	-.27*'	.02	.05	.02	-.28*'	-.10	-.40*'	-.03	-.11	-.10	.30**	.01	.28**	.10				
28. Reliance - O	5,21	1,37	-.06	-.17*'	-.06	-.12*	-.03	.12*	-.07	.08	.02	.01	-.10	.01	.05	.04	.00	.00	-.20*'	-.17*'	-.10	-.04	.00	-.23*'	.03	.12*	-.02	.29**	.07			
29. Supervisor support	3,79	0,55	.03	.24**	.11	.42**	-.03	.75**	-.04	.21**	.01	.10	-.05	.24**	.05	.40**	-.10	.51**	.01	.25**	.10	.35**	.02	-.30*'	.11	.57**	.10	.75**	-.05	.18**		
30. Trust in management	2,76	0,55	.06	.25**	.15*	.29**	.04	.32**	-.05	.10	-.03	.16**	-.11	.15**	.13	.20**	-.10	.24**	.03	.15**	.05	.19**	.00	.04	.11	.30**	.00	.25**	-.10	.10	.40**	
31. Job satisfaction	3,68	0,50	.11	.23**	.22**	.23**	.13	.36**	.00	.08	.10	.18**	-.12	.11	.13	.22**	-.03	.22**	.00	.13*	.13	.23**	.10	-.16*'	.01	.22**	.02	.27**	-.13	.13*	.44**	.49**

S = Self; O = Observer

**p <.05. **p <.01. ***p<.001.*

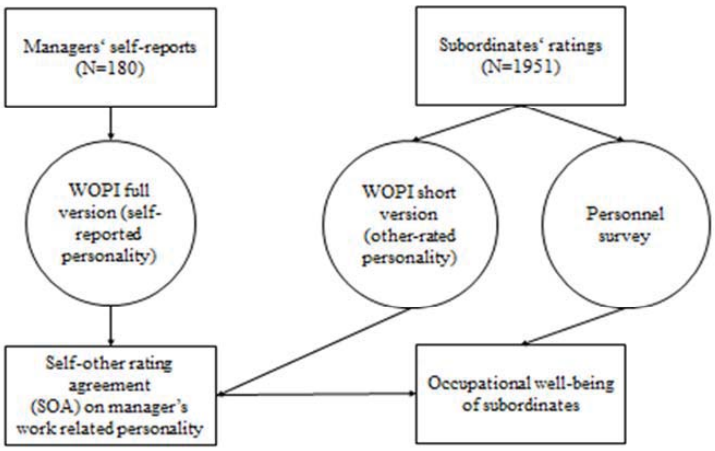
Table 2. Polynomial regression results combined with surface tests across WOPI traits and three dimensions of occupational well-being (N=180).

	Occupational well-being content								
	Supervisor support			General Management			Job satisfaction		
WOPI scale	a1	a3	a4	a1	a3	a4	a1	a3	a4
Extroversion									
Competition	.13 (.05)*	-.17 (.05)***	-.17 (.06)**	.15 (.06)*	-.15 (.06)**	-.17 (.06)**	.13 (.05)*		-.15 (.06)*
Leadership	.22 (.05)***	-.19 (.05)***	-.17 (.05)**	.13 (.05)*			.13 (.05)*		
Inspiration	.30 (.04)***	-.47 (.04)***	-.08 (.04)*	.16 (.06)**	-.19 (.06)**	-.18 (.06)**	.20 (.05)***	-.13 (.05)*	
Openness to Experience									
Orientation		-.17 (.05)**	-.21 (.06)***		-.21 (.06)***				-.14 (.06)*
Perception		-.17 (.05)**	-.20 (.06)***		-.19 (.06)**		.14 (.05)*	-.13 (.05)*	-.15 (.06)*
Thinking	.14 (.06)*	-.28 (.06)*	-.18 (.06)**	.10 (.04)*	-.31 (.08)***				-.25 (.06)***
Concientiousness									
Ambiguity-change (reversed)	.14 (.05)**	-.20 (.05)***	-.15 (.04)***	.15 (.06)*			.14 (.05)*		
Focused Achievement	.21 (.05)***	-.32 (.05)***							
Decision Making (reversed)	.11 (.04)**	-.27 (.07)***			-.17 (.08)*			-.20 (.07)**	
Emotional Stability									
Optimism	.14 (.06)*		-.17 (.06)**						.13 (.04)**
Self-reflection	-.15 (.06)**								
Agreaableness									
Sociability	.21 (.05)***	-.28 (.05)***	-.12 (.05)*	.15 (.05)***			.14 (.06)*	-.28 (.06)*	-.18 (.06)**
Empathy	.37 (.04)***	-.41 (.04)***							-.16 (.06)**
Reliance		-.14 (.07)*						-.16 (.06)**	

Notes: a1 = (b1 + b2), where b1 is beta coefficient for self-report (SR) and b2 is beta coefficient for other-rating (OR), b4 is beta coefficient for the cross-product of the SR and OR, and b5 is beta coefficient for OR squared. a3 = (b1 − b2) and a4 = (b3 − b4 + b5). Standard errors in parenthesis.

**p <.05. **p <.01. ***p<.001.*

Figure 1. The research setting.



Appendix 1. WOPI scales and example items

(fo) FOCUSED ACHIEVEMENT

I like jobs where quality is never compromised because of pressing deadlines.
I always want to work thoroughly, even if it wasn't necessary.
I like the kind of work that allows me to concentrate on particular matters.

(co) COMPETITIVE ACHIEVEMENT

I strive for top results in everything I do.
In my ideal job I would compete against my own performance and that of others.
I want to win, and I hate losing.

(le) LEADERSHIP

I like to give orders and get things going.
In the company of another person, I usually make the decisions.
I'm very demanding towards others.

(is) INSPIRATION

I'd much rather work in a well-known organization than an unknown one.
What impression I make on other people is usually important to me.
I want to be noticed when I'm with other people.

(so) SOCIABILITY

I'd never want to miss an opportunity to be with other people.
I make friends easily; I quickly feel at home even in the company of strangers.
I always prefer to work with other people rather than by myself.

(em) EMPATHY

I'm happy to put my own things aside to do someone else a favor.
I'm always willing to give and lend things to people who need them more than I do.
In my ideal job, I would work for the benefit of other people.

(re) RELIANCE

I carefully sound out others' opinions before I make a decision.
I find it natural to follow stronger individuals.
I often ask other people for advice

(or) ORIENTATION

I'm more interested in ideas than facts.
I get the most peculiar ideas.
I'm more of a radical innovator than a stable builder.

(pc) PERCEPTION

I enjoy interpreting complex and difficult concepts.
Theories help me enormously to understand things.
I often question the usefulness of traditional ways of thinking and doing things.

(th) THINKING

I let my feelings influence my decision making a great deal.
When I start doing something, I don't often know what the end result will look like.
My thinking is very organized and logical. (R)

(dc) DECISION MAKING

I'm able to get many projects going at the same time because I don't plan too much ahead.
I usually make decisions quickly, without delay.
I enjoy taking risks.

(am) AMBIGUITY-CHANGE

I don't like unexpected situations.
It bothers me if my duties at work are unclear.
I prefer changes to happen gradually.

(op) OPTIMISM

I usually succeed in everything I do.
I'm always full of energy.
I always look at the positive side of things.

(sr) SELF-REFLECTION

I never try to cover up my mistakes.
I never cause inconvenience to my colleagues at work.
I can always make the right decisions, even in difficult situations.

Appendix 2. Confirmatory factor analyses

Appendix 2, Table 1. Factor loadings based on a maximun likelihood analysis with promax rotation for 14 scales of WOPI (N=235)

	Extraversion	Conscientiousness	Agreeableness	Openness	Emotional stability
Inspiration	.74				
Leadership	.70				
Competition	.66				
Ambiguity-change (reversed CON)		-.65			
Focused Achievement		.65			
Decision making (reversed CON)		-.65			
Sociability			.64		
Empathy			.63		
Reliance			.61		
Orientation				.85	
Perception				.60	
Thinking				.43	
Self-reflection					.73
Optimism	.42	-.45			.55

Note. Factor loadings < .4 are suppressed.

Appendix 2, Table 2. Factor loadings based on a maximun likelihood analysis with promax rotation for 28 items (N=297)

	Supervisor support	Trust in management	Job satisfaction
Providing circumstances to succeed in the work	.84		
Superior as a motivator	.81		
Support for subordinates' development efforts	.80		
Spirit of the team	.78		
Equality of treatment of subordinates	.78		
Consideration of subordinates' views	.77		
Clarification of subordinates' objectives	.75		
Encouraging co-operation among employees	.73		
Amount of feedback from superior	.72		
Sharing information and skills in the team	.69		
Usefulness of the latest development discussion	.66		
Superior's time to listen to subordinates	.65		
Trust in the management of the function		.88	
Dialogue between management and personnel		.85	
Consideration of staff in decision-making		.84	
Implementation of changes in the function		.82	
Processing of survey results		.75	
Support to co-operation by function management		.71	
Fairness of rewarding in the function		.65	
Decentralization of decision-making in the func		.62	
Lack of rumours circulating among staff		.61	
Information about strategies of the function		.59	
Level of the challenge provided by job			.88
Enjoyment of the work			.85
Degree of satisfaction from the work			.84
Opportunity to utilize competencies			.73
Opportunities for prof. development at VTT			.62
Importance of the job for the organization			.61

Note. Factor loadings < .4 are suppressed.