



Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Generalizability of self–other agreement from one personality trait to another

Jüri Allik^{a,b,*}, Anu Realo^{a,b}, René Mõttus^{a,b}, Peter Kuppens^{c,d}

^a Department of Psychology, University of Tartu, Estonia

^b The Estonian Centre of Behavioural and Health Sciences, Estonia

^c Department of Psychology, Katholieke Universiteit Leuven, Leuven, Belgium

^d School of Behavioural Science, University of Melbourne, Australia

ARTICLE INFO

Article history:

Received 22 February 2009

Received in revised form 11 September 2009

Accepted 14 September 2009

Available online xxx

Keywords:

Generalized consistency

Cross-observer agreement

Personality traits

ABSTRACT

If you are an accurate judge of your friends' openness, are you also good at rating their conscientiousness? In this paper we examined how well self–other agreement on one personality trait accords with self–other agreement on other personality traits. Data from four Estonian and Belgian samples containing 818 targets and 1281 knowledgeable raters were analyzed. Results demonstrated that self–other agreement only moderately generalizes from one personality trait to another suggesting that the predictability of an individual can vary for different personality traits. When trait agreement was decomposed into the contributions of individual pairs of raters, these were only moderately correlated with different coefficients of profile agreement, suggesting that these two forms of agreement are far from being identical.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

People regularly make personality judgments of themselves and other people. Typically, the accuracy of such judgments is estimated as a degree of agreement between ratings made by different raters. For example, self–ratings are compared to those of well-informed observers. The agreement of personality ratings can be assessed in two principal ways either estimating the consistency between raters in the rank ordering on one particular trait or estimating congruence over a set of personality attributes (personality profile) for a single target–rater pair (Funder & Colvin, 1997). These two methods – trait and person centered approaches – look at the agreement from two different angles: the first deals with ranking of individual on some traits and the second with ranking of traits within individual.

Generally, for most personality traits people tend to achieve at least moderate cross-observer agreement (Connolly, Kavanagh, & Viswesvaran, 2007; Funder & Colvin, 1997). For example, the median cross-observer trait agreement in a number of studies using measures of the Five-Factor model was .40 or higher on all the Big Five personality dimensions (McCrae et al., 2004). The mean profile agreement across all target–rater pairs was shown to be in the same range (Kenny & Winquist, 2001; McCrae, 2008). This level of agreement is remarkable considering a complicated chain of events required for an accurate personality judgment: the target

of judgment must display behaviors and cues that are relevant to the trait being judged and the judge must detect these cues and correctly use them to make judgments (Funder & Colvin, 1997). Yet there seem to be considerable individual differences in how well external raters' and target's personality judgments correspond, and several moderators of such differences have been investigated. One of these moderators that has received much attention is "judgability": for example, psychologically better adjusted individuals are easier to judge than less adjusted people (Colvin, 1993; Furr, Dougherty, Marsh, & Mathias, 2007). Another equally powerful moderator is the ability to make correct personality judgments from available information: some individuals are believed to be better judges of personality than are others (Letzring, Wells, & Funder, 2006; Taft, 1955).

However, the search for stable individual differences in cross-observer agreement would benefit from knowing about the generalizability of cross-observer agreement over different personality traits. It would be important to know whether cross-observer agreement, for example, on extraversion is related to the cross-observer agreement on conscientiousness because it is difficult to speak of generally good target or good raters if the agreement does not generalize from one personality trait to at least some other personality traits. To our knowledge no studies have articulated or answered this question so far mainly because it was not clear how to compute agreement over different personality traits.

In this article we show that the problem of generalizability of self–other agreement from one personality trait to another can be solved by decomposing the self–other correlation on a trait into the individual pairs' contributions to this correlation. Asendorpf

* Corresponding author. Address: Department of Psychology, University of Tartu, Tiigi 78, Tartu 50140, Estonia. Tel.: +372 5184277.

E-mail address: juri.allik@ut.ee (J. Allik).

(1992) proposed a simple method of how to partial the rank-order correlation between two variables into the individual's contribution to overall correlation. This is based on a simple idea that these correlated pairs who occupy approximately the same position in their respective rankings contribute positively to the correlation whereas those who occupy very different positions in their respective rankings contribute negatively. Evidently, the inequality between positions of any pair members is inversely proportional to the level of the overall correlation. Knowing the individual contribution of each pair of raters to the trait agreement it is possible to ask to what extent the contribution of each pairs of informants to self–other agreement is generalizable from one personality trait to another. Are there “better predictable” pairs of raters who consistently occupy similar positions on nearly all trait rankings and “less predictable” ones whose positions are erratic across personality traits? Most researches would probably take it for granted that people can be divided on more and less predictable ones.

A second issue we want to address concerns the relationship between the two approaches to cross-observer agreement. It has been suggested that the two ways of estimating cross-observer agreement yield essentially similar results (Kenny & Winquist, 2001). However, as a method is now available to calculate the contribution of each pair of raters to the overall trait agreement, we can use these contributions to formally test the relationship between trait and profile agreement. It must be noted that the second research issue is closely related to the first. If cross-observer trait agreement generalizes at least to some degree from one trait to others, we can compare average trait agreement to the profile agreement. As well, we can examine how well contributions of individual pairs of raters to the agreement on single traits predict overall profile agreement.

In the present study we address the two issues. First, we will examine the generalizability of self–other agreement from one trait to another and, secondly, we will investigate the relationships between the trait and person centered approaches to cross-observer agreement.

2. Method

2.1. Samples and instruments

There were four samples from two different countries, Estonia and Belgium. Personality was assessed using different NEO-PI family instruments. The total number of participants was 2109 from whom 818 were targets and 1291 were external raters who evaluated personality of the targets.

2.1.1. Flemish sample

The Flemish sample (FLEM) consisted of 345 target participants (270 women and 75 men) who were psychology students of the Katholieke Universiteit Leuven and who as a course requirement rated their own personality with the Dutch version of the NEO-PI-R (Hoekstra, Ormel, & DeFruyt, 1996). They also recruited a well acquainted person ($n = 345$; 190 women, 112 men, and 43 did not specify their sex), either relative or friend, who rated their personality using the other-report form of the same instrument. The mean age of targets was 18.4 ($SD = 3.0$) years. The mean age of external raters was 29.5 ($SD = 13.7$) years.

2.1.2. Estonian sample no. 1

This sample (EST1) was assembled for the study of self–other congruence in mind-reading ratings (Realo et al., 2003). One hundred and one individuals (81 women and 20 men) living in Estonia served as “target-persons”. Their mean age was 21.9 years ($SD = 4.1$) ranging from 17 to 41 years. Each target recruited two

external raters. The mean age of the 202 raters (153 women and 49 men) was 26.0 years ($SD = 10.0$). All individuals ($n = 303$) volunteered to participate in this study and received no compensation for their involvement. All participants completed the 60-item self-report forms of the Estonian NEO-FFI (Allik, Laidra, Realo, & Pullmann, 2004). Since the NEO-FFI does not have subscales to measure the five dimensions, 12 items measuring each dimension were artificially assigned to 4 pseudo-subscales 3 items in each. The resultant factor structure derived from those 20 pseudo-subscales was exemplary: all subscales loaded most on the intended factors and there were no complementary high loadings ($>.40$) on the “wrong” factors.

2.1.3. Estonian sample no. 2

In the second Estonian sample (EST2), the EE.PIP-NEO personality inventory (Mõttus, Pullmann, & Allik, 2006) was administered to 154 participants (53 men and 101 women) with a mean age of 43.9 ($SD = 17.9$) years, ranging from 16 to 83. The sample was representative of different age groups containing 23% of participants who were older than 60 years. Participants had various educational backgrounds: only approximately 43% had secondary or higher education (6% did not report their educational attainment). The participants were volunteers who were reached through the personal contacts of the collaborating graduate students and psychologists. In addition to self-reported personality ratings, each participant (target) was rated by two well-informed external raters. The raters were found by targets themselves. The sample of raters ($n = 308$) included 203 women, 67 men, and 38 participants who did not report their gender. The mean age of the raters was 38.2 ($SD = 15.9$) years, ranging from 16 to 81 years. On average, the raters were about five years younger and more highly educated than the targets. In general, about 52% of raters were close relatives or partners, 25% were friends, and 12% colleagues.

2.1.4. Estonian sample no. 3

The third Estonian sample (EST3) included 218 students of social sciences (37 men, 180 women, one did not specify their gender) who answered questionnaires accompanied by a standard instruction to describe themselves honestly and accurately (Konstabel, Aavik, & Allik, 2006). They were also asked to provide two peer-reports ($n = 436$) from their acquaintances, relatives or close friends. The questionnaires were completed voluntarily; some students studying psychology received an extra credit toward fulfillment of their course requirements. The Big Five dimensions were measured with the Estonian version (Kallasmaa, Allik, Realo, & McCrae, 2000) of the NEO-PI-R (Costa & McCrae, 1992).

2.2. Self-other agreement

2.2.1. Trait agreement

The Pearson product moment correlation between self and other ratings was used to evaluate self–other agreement on a specific trait. The individual target-rater contribution r_A to the self–other agreement was computed according to the Asendorpf (1992) formula: $r_A = 1 - (z_s - z_o)^2 / 2$, where z_s and z_o are z-scores for the self and other ratings standardized across the sample. The mean of these r_A coefficients across all respondent pairs is equal to the Pearson product moment correlation between self and other ratings.

2.2.2. Profile agreement

The profile agreement was computed by correlating the scores of self-rated personality traits across all subscales with observers' ratings on the same set of subscales. There are several ways to compute the degree of similarity between two profiles and some

of these similarity coefficients are shown to be superior to others (McCrae, 2008).

The first coefficient we used was the Pearson correlation r which is the most widely used profile agreement index. Its main advantage is invariance under any linear transformations of both variables and therefore it is insensitive to profile elevation and scatter (Cronbach, 1955).

The second index was an intraclass correlation (double entry), ICC_{DE} , which was shown superior to several other agreement coefficients (McCrae, 2008). In the double entry ICC_{DE} each element in the paired profile is entered twice but in reversed order across all pairs of comparison. Unlike the Pearson r , ICC_{DE} is not only sensitive to the shape but to the profile elevation as well.

Finally, we used a profile similarity coefficient r_c , which is invariant over variable reflection (Cohen, 1969). The previous two measures of agreement suffer from the imperfection that their values varies with arbitrary decisions in which direction variables are coded, for example, whether neuroticism or emotional stability is scored high. For computing r_c each element in the paired profile is entered twice in original and reflected form across all pairs of comparison. The reflected score X' is computed from the original score X as $X' = 2m - X$, where m is the midpoint of the scale. Thus, r_c can be understood as an index of proportionality of paired deviations from the neutral point of the scale (Cohen, 1969). These three profile agreement indices were selected since they are all correlation coefficients. Other measures of the profile similarity, particularly based on the sum of squared differences, were avoided due to their incompatibility with the correlation coefficient that was used to estimate self–other trait agreement.

All three profile agreement indices – r , ICC_{DE} , and r_c – were computed before and after standardization of the scores within each trait across the entire sample. The standardization removes the normative pattern common to the whole sample which is known as *stereotypic accuracy* (Cronbach, 1955; Furr, 2008). Stereotypic accuracy constitutes the part of profile agreement which is there even regardless of any individuating information on the target. Due to similarity of normative scores all profiles tend to be inter-correlated and cross-observer profile agreement is artificially inflated. Standardization of raw scores, which results in each trait having a mean of 0 and a standard deviation of 1, the normative pattern is removed and the profile agreement estimates only the *differential accuracy*.

In all cases when there were two external raters who rated the same target their ratings were averaged before computing either trait or profile agreement.

3. Results

Analysis showed a good self–other congruence for both trait and profile agreement (see Table 1). The trait agreement values were in the range that is typically reported in previous studies (Connolly et al., 2007; Grucza & Goldberg, 2007; McCrae et al., 2004). Like many previous studies (Funder & Colvin, 1997), Extraversion showed better self–other congruence (.67) than did Neuroticism (.49) (Table 1, the last column).

Like Kenny and Winquist (2001), we found that the mean self–other trait agreement across the Big Five dimensions (.56) was approximately similar to the average profile agreement (Pearson r) before (.59) and somewhat lower after normalizing data (.44). As expected, the mean stereotypic profile agreement (.59, .53, and .70, respectively for r , ICC_{DE} , and r_c) was higher than the mean differential profile agreement (.44, .40, and .47) suggesting that a certain amount of accuracy in the raw self–other agreement is based on assumed similarity to the average person. The largest drop was in the Flemish sample where stereotypic agreement was substantially larger than the differential agreement for all three profile agreement coefficients. Nevertheless, in all samples a considerable amount of accuracy remained even after removing the normative profile. The profile agreement values measured in terms of the Pearson correlation r were comparable to the mean correlation between self and observer ratings obtained with the NEO-PI-3 which was .49 in an adult sample of Americans (McCrae, Costa, & Martin, 2005). Also the mean ICC_{DE} value (.40 for normalized data) was comparable to the mean self–other profile agreement of 635 American adults which was equal to .43 (McCrae, 2008).

In most samples the three different measures of the self–other agreement – r , ICC_{DE} , and r_c – were highly correlated, typically in the range from .91 to .98. Only in the normalized Flemish data did the correlation between r and ICC_{DE} and between r and r_c slightly drop to .73 and .74, respectively. However, the agreement between three different coefficients of the profile similarity was still high and the advantage of one of them over others could be only marginal.

In response to our first research question, we analyzed the consistency of self–other agreement across different traits. Table 2 demonstrates intercorrelations between individual contributions to agreement on different personality traits in four samples. From 40 intercorrelations all 19 statistically significant ($p < .05$) intercorrelations were positive. On average, cross-trait consistency of self–other agreement was rather moderate, however. Measured in terms of Cronbach alpha consistency indices were .41, .34, .38,

Table 1

The mean values of self–other trait agreement and self–other profile agreement.

	Trait/type of data	Index	Sample, number of targets, instrument				Mean
			FLEM N = 343 NEO-PI-R	EST1 N = 101 NEO-FFI	EST2 N = 154 EE.PIP-NEO	EST3 N = 212 NEO-PI-R	
Trait agreement	Neuroticism	r_A	.45	.58	.52	.41	.49
	Extraversion	r_A	.61	.67	.71	.67	.67
	Openness	r_A	.44	.61	.65	.68	.60
	Agreeableness	r_A	.48	.63	.50	.53	.54
	Conscientiousness	r_A	.48	.61	.48	.56	.53
	Mean	r_A	.49	.62	.57	.57	.56
Profile agreement	Raw data	r	.53	.56	.67	.58	.59
		ICC_{DE}	.48	.50	.62	.53	.53
		r_c	.81	.65	.70	.63	.70
	Normalized data	r	.30	.50	.50	.46	.44
		ICC_{DE}	.31	.43	.44	.41	.40
		r_c	.38	.51	.50	.47	.47

Note: r_A = Asendorpf index of the individual contribution to trait agreement; r = Pearson product moment correlation; ICC_{DE} = intraclass correlation (double entry); r_c = Cohen's profile similarity coefficient.

and .51 for FLEM, EST1, EST2, and EST3 samples respectively. The average intercorrelation across all trait pairs and samples was .13. Only intercorrelations between Agreeableness and Conscientiousness were statistically significant ($p < .05$) in all four samples and the highest mean correlation (.24) was reached between Extraversion and Openness to Experience. On average, these results seem to suggest that target-rater pairs who contribute significantly to self–other agreement on one trait do not necessarily contribute significantly to self–other agreement on another trait.

Although the generalizability of agreement from one trait to another was relatively low, mainly positive intercorrelations between the trait agreements across all traits allowed us to compute the overall self–other trait agreement by taking the average across all five agreement coefficients (“Average” columns in Table 3).

What is the relationship between trait agreement and profile agreement? Table 3 demonstrates correlations between individual contributions to the self–other trait agreement (r_A) and the three different profile agreement indices before and after the standardization of raw scores. Except few small discrepancies, all three profile agreement indices behaved in a similar manner supporting the previous conclusion that r , ICC_{DE} , and r_c are comparable coefficients of the self–other profile agreement. Expectedly the averaged contributions to agreements on different traits were the best predictors of profile agreement, meaning that the profile agreement

depends most on the summary agreement on all of the traits constituting the profile. At the specific level, the strongest trait–profile agreement was reached for Neuroticism and Conscientiousness both before and after normalizing data. Surprisingly, although Neuroticism yielded the weakest self–other trait agreement (.49, see the last columns of Table 2) it turned out to be the best in predicting the self–other profile agreement. Also interestingly, the generally relatively high self–other agreement on the Openness dimension was a very poor predictor of profile agreement: the trait–profile correlations were very low and almost identical either before or after the data normalization.

The relationship between trait agreement and profile agreement was lowest in the Flemish (correlation between r_A and ICC_{DE} was .28) and highest in the second Estonian (EST2) sample (.59). A considerable scatter and variation between these two forms of agreement suggest that in spite of a generally positive association the relationship between them is neither fixed across samples nor even close to being strictly functional.

4. Discussion

We were able to demonstrate that the average intraindividual consistency of self–other agreement across all possible trait pairs and samples was as low as .13. Thus, differences between positions in self and other ratings (i.e. cross-observer disagreement) appear predominantly to be caused by unsystematic factors such as measurement error or unique dyadic effects. It is very likely that some people are more predictable on some traits, whereas other persons may be better predictable on other traits. What these results demonstrate, however, is that the predictability in terms of self–other agreement is not a uniform dimension of individual differences.

This finding implies that the view according to which some people are more predictable than others may be inaccurate. Although the suggestion that there are certain individuals whose personality traits can be more accurately estimated by knowledgeable others appears intuitively almost irresistible (Bem & Allen, 1974; Bem & Funder, 1978), this intuition may prove to be wrong or at least of little practical value. Findings of this study provide a likely explanation for the fact that researchers have not been very successful in separating more consistent people from less consistent ones (Bem & Allen, 1974; Biesanz & West, 2000; Chaplin, 1991; Chaplin & Goldberg, 1984; Zuckerman et al., 1988): more consistent indi-

Table 2
Intercorrelations between individual contributions to agreement (r_A) on different traits.

Pairs of personality traits	Sample				
	FLEM	EST1	EST2	EST3	Mean
N–E	.09	.09	.03	.17	.10
N–O	.02	.03	.21	–.06	.05
N–A	.20	.03	.02	.26	.13
N–C	.23	.07	.23	.24	.19
E–O	.18	.11	.26	.40	.24
E–A	.11	.28	–.07	.18	.13
E–C	.04	.09	.10	.41	.16
O–A	.14	–.02	.08	–.09	.03
O–C	.02	.16	–.01	.04	.05
A–C	.18	.20	.25	.24	.22
Mean	.12	.10	.11	.18	.13

Note: Significant correlations ($p < .05$) are shown in boldface type.

Table 3
Correlations between self–other trait agreement (r_A) and different profile agreement indices for raw and normalized data.

Sample	Index of profile agreement	Raw data					Normalized data						
		N	E	O	A	C	Average	N	E	O	A	C	Average
<i>Asendorpf trait agreement index (r_A)</i>													
FLE	r	.25	.21	.19	.19	.15	.27	.15	.14	.10	.12	.06	.15
	ICC_{DE}	.24	.27	.25	.24	.19	.35	.24	.20	.19	.19	.15	.28
	r_c	.33	.31	.32	.32	.30	.49	.20	.19	.14	.18	.11	.22
EST1	r	.40	.10	.12	.05	.25	.37	.23	.15	.07	.05	.25	.29
	ICC_{DE}	.34	.18	.21	.17	.33	.48	.18	.23	.17	.19	.33	.42
	r_c	.31	.26	.21	.24	.49	.58	.20	.20	.21	.14	.37	.43
EST2	r	.39	.18	.04	.32	.52	.48	.46	.16	.21	.33	.41	.50
	ICC_{DE}	.38	.26	.13	.33	.55	.55	.43	.23	.30	.37	.45	.59
	r_c	.34	.21	.10	.34	.56	.55	.41	.18	.27	.35	.42	.55
EST3	r	.49	.37	.03	.30	.29	.53	.46	.27	–.03	.31	.30	.49
	ICC_{DE}	.47	.41	.09	.32	.32	.57	.44	.33	.05	.34	.33	.54
	r_c	.44	.41	.11	.31	.33	.56	.42	.33	.04	.33	.33	.52
Mean	r	.38	.21	.10	.21	.30	.42	.33	.18	.09	.20	.25	.36
	ICC_{DE}	.36	.28	.17	.26	.35	.49	.33	.25	.18	.27	.32	.46
	r_c	.35	.30	.18	.30	.42	.54	.31	.23	.16	.25	.31	.43

Note: r = pearson product moment correlation; ICC_{DE} = intraclass correlation (double entry); r_c = Cohen's profile similarity coefficient; N = neuroticism; E = extraversion; O = openness; A = agreeableness; C = conscientiousness; Average = average Asendorpf index (r_A) across the Big-Five personality dimensions. Significant correlations ($p < .05$) are shown in boldface.

viduals cannot be separated from less consistent ones because there is no generalizable cross-observer consistency across different personality traits. It seems to be impossible to make one universal ranking of “bad” or “good” raters that remains the same for all personality traits. Targets and raters from the same target-rater pairs may occupy identical or nearly identical positions in their respective rankings on one personality trait but can have a considerable disparity in their rankings on another personality trait.

The view that trait and profile agreement characterize essentially the same phenomenon (e.g. Kenny & Winquist, 2001) is difficult to maintain. Although the average values of the trait and profile agreement were almost equally high, the individual contributions to these two forms of agreement were much less coordinated. When the correlation between targets' and external raters' standings on personality traits was decomposed into individual contributions of target-rater pairs to the correlation, the association between the contributions and profile agreements was rather moderate. The average trait-profile correlations across personality dimensions for normalized data were in the range from .36 to .46. Although even these values may look relatively high by usual standards of personality research, it is relevant to bear in mind that we are not talking about two logically independent variables between which the correlation is computed. These coefficients characterize two schemes how to estimate the self-observer agreement on exactly the same set of data. An obvious implication of these relatively moderate correlations is that many target-rater pairs can have a good profile agreement without good agreement on all traits and vice versa. Therefore it is reasonable to conclude that the trait and profile agreement are two related but still sufficiently distinctive forms of self-other congruence.

Acknowledgements

This paper was started during the stay of the first author at the Katholieke Universiteit Leuven, Leuven, Belgium (Fall semester 2005/2006). This research was supported in part by the grants from the Estonian Science Foundation (#7020) and the Estonian Ministry of Science and Education (SF0182585s03 and SF0180029s08). The last author is a postdoctoral research fellow of the Fund for Scientific Research-Flanders.

References

- Allik, J., Laidra, K., Realo, A., & Pullmann, H. (2004). Personality development from 12 to 18 years of age: Changes in mean levels and structure of traits. *European Journal of Personality*, 18(6), 445–462.
- Asendorpf, J. B. (1992). Beyond stability – predicting interindividual differences in intraindividual change. *European Journal of Personality*, 6(2), 103–117.
- Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistency in behavior. *Psychological Review*, 81, 506–520.
- Bem, D. J., & Funder, D. C. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review*, 85, 485–501.
- Biesanz, J. C., & West, S. G. (2000). Personality coherence: Moderating self-other profile agreement and profile consensus. *Journal of Personality and Social Psychology*, 79(3), 425–437.
- Chaplin, W. F. (1991). The next generation of moderator research in personality psychology. *Journal of Personality*, 59, 143–178.
- Chaplin, W. F., & Goldberg, L. R. (1984). A failure to replicate the Bem and Allen study of individual differences in cross-situational consistency. *Journal of Personality and Social Psychology*, 47(5), 1074–1090.
- Cohen, J. (1969). R_c : A profile similarity coefficient invariant over variable reflection. *Psychological Bulletin*, 71, 281–284.
- Colvin, C. R. (1993). “Judgable” people: Personality, behavior, and competing explanations. *Journal of Personality and Social Psychology*, 64(5), 861–873.
- Connolly, J. J., Kavanagh, E. J., & Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment*, 15(1), 110–117.
- Costa, P. T. J., & 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.
- Cronbach, L. J. (1955). Process affecting scores on “understanding others” and “assumed similarity”. *Psychological Bulletin*, 52, 177–193.
- Funder, D. C., & Colvin, C. R. (1997). Congruence of others' and self-judgements on personality. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 617–647). San Diego: Academic Press.
- Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality*, 76(5), 1267–1316.
- Furr, R. M., Dougherty, D. M., Marsh, D. M., & Mathias, C. W. (2007). Personality judgment and personality pathology: Self-other agreement in adolescents with conduct disorder. *Journal of Personality*, 75(3), 629–662.
- Gruca, R. A., & Goldberg, L. R. (2007). The comparative validity of 11 modern personality inventories: Prediction of behavioral acts, informants reports, and clinical indicators. *Journal of Personality Assessment*, 89, 167–187.
- Hoekstra, H. A., Ormel, J., & DeFruyt, F. (1996). *NEO Persoonlijkheidsvragenlijsten NEO-PI-R en NEO-FFI. Handleiding*. Lisse, The Netherlands: Swets & Zeitlinger.
- Kallasmaa, T., Allik, J., Realo, A., & McCrae, R. R. (2000). The Estonian version of the NEO-PI-R: An examination of universal and culture-specific aspects of the five-factor model. *European Journal of Personality*, 14, 265–278.
- Kenny, D. A., & Winquist, L. A. (2001). The measurement of interpersonal sensitivity: Consideration of design, components, and unit of analysis. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 265–302). Englewood Cliffs, NJ: Erlbaum.
- Konstabel, K., Aavik, T., & Allik, J. (2006). Social desirability and consensual validity of personality traits. *European Journal of Personality*, 20(7), 549–566.
- Letzring, T. D., Wells, S. A., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*, 91(1), 111–123.
- McCrae, R. R. (2008). A note on some measures of profile agreement. *Journal of Personality Assessment*, 90, 105–109.
- McCrae, R. R., Costa, P. T., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised neo personality inventory. *Journal of Personality Assessment*, 84(3), 261–270.
- McCrae, R. R., Costa, P. T., Martin, T. A., Oryol, V. E., Rukavishnikov, A. A., Senin, I. G., et al. (2004). Consensual validation of personality traits across cultures. *Journal of Research in Personality*, 38(2), 179–201.
- Möttus, R., Pullmann, H., & Allik, J. (2006). Toward more readable Big-Five personality inventories. *European Journal of Psychological Assessment*, 22(3), 149–157.
- Realo, A., Allik, J., Nõlvak, A., Valk, R., Ruus, T., Schmidt, M., et al. (2003). Mind-reading ability: Beliefs and performance. *Journal of Research in Personality*, 38, 420–445.
- Taft, R. (1955). The ability to judge people. *Psychological Bulletin*, 52, 1–23.
- Zuckerman, M., Koestner, R., DeBoy, T., Garcia, T., Maresca, B. C., & Sartoris, J. M. (1988). To predict some of the people some of the time: A reexamination of the moderator variable approach in personality theory. *Journal of Personality and Social Psychology*, 54, 1006–1019.