

# Self–other Disagreement in Personality Assessment: Significance and Prognostic Value

R. M. Mosterman<sup>1\*</sup> and A. A. Jolijn Hendriks<sup>2</sup>

<sup>1</sup>Psychologenpraktijk Elf, Zwolle

<sup>2</sup>University of Groningen

The feasibility of the use of multiple informant reports in clinical practice was examined in a sample of 105 psychiatric outpatients who provided self-ratings and (2–3) informants' reports on the Five Factor Personality Inventory. The response rate was 97%. The patients assessed themselves as less extraverted and more emotionally stable than their proxies did. In addition, the significance of self–other disagreement was investigated. Our first hypothesis, stating that self–other disagreement would correlate with (personality) pathology, was confirmed: self–other disagreement predominantly occurred in introverted, shy, hostile and depressed persons who tended to have more personality problems and co-morbidity. We found no support for our second hypothesis, stating that self–other disagreement would predict a diminished therapy effect. An important finding, however, was that self–other disagreement proved to be a strong predictor of dropout. Furthermore, a decrease in depression, hostility and shyness was positively correlated with a decrease in self–other disagreement. Copyright © 2010 John Wiley & Sons, Ltd.

## Key Practitioner Message:

- The use of multiple informants in personality assessment is a feasible enterprise in clinical practice.
- Self–other disagreement in personality assessment predominantly occurs in introverted, shy, hostile and depressed patients.
- Self–other disagreement is found more often in patients with personality problems and co-morbidity.
- Self–other disagreement, in particular on conscientiousness, proves to be a strong predictor of dropout.

**Keywords:** Personality, Big Five, Informant Report, Self–other Disagreement, Psychopathology

## INTRODUCTION

### *The Credibility of Self-Reports*

Assessment of personality relies heavily on self-report measures (e.g., Achenbach, Krukowski, Dumenci, & Ivanova, 2005; Kagan, 2007; Paulhus &

Vazire, 2007; Vazire, 2006). At first sight this makes sense: 'One would expect that the individual possessing the particular personality traits should be able to provide the most informative and accurate information about these constructs' (McDonald, 2008, p. 2). However, can people be trusted in what they say about themselves? Besides response biases such as socially desirable, acquiescent or extreme responding, the reliability and validity of self-ratings can be questioned because of distorted self-percep-

\*Correspondence to: R. M. Mosterman (Ineke), Bloemendalstraat 5, 8011 PJ Zwolle, the Netherlands.  
E-mail: i.mosterman@psychologenpraktijkelf.nl

tions due to memory failure, consistency seeking, self-enhancement or self-interests such as having the secondary benefits of being ill or the reduction of sanctions of criminal acts. It is also possible that as a consequence of (personality) pathology, self-knowledge is limited. For these reasons, it is preferable to maximize the validity by combining the self-report approach with other methods, such as informant reports and observational measures (McDonald, 2008; Paulhus & Vazire, 2007).

Hofstee (1994, 2009) goes even further: he *defines* personality in terms of intersubjective agreement. As judgment is subjective by definition, whether provided by the target person himself or herself, a third person, or the diagnostician, and objectivity is not within reach, the best available point of reference for the assessment of personality is the averaged judgments of multiple informants (Hofstee, 1994, p. 150). It yields a more reliable and in the case of personality pathology, potentially also more valid result because response tendencies and interpretation and measurement errors will have been averaged out. Hofstee recommends that those who know the target person best should be asked to complete an informant personality questionnaire. Self-reports add to the reliability of assessments to the extent that they agree with the assessments by others. In the words of Hofstee (2009): '*... to the extent that subjective personality deviates systematically from the consensus of others, such deviations may be found clinically relevant. [...] In clinical diagnosis and treatment, discrepancies are potentially much more relevant than would appear from the virtual absence of self-other comparisons in practice. At the very least, a client could benefit from a systematic confrontation between the two perspectives*' (p. 225).

### The Use of Informant Methods

Vazire (2006) disproved the preconceptions of many researchers that informant methods for personality assessment are time-consuming, expensive, ineffective and vulnerable to faking or invalid responses. According to her, informant reports can be '*a cheap, fast and easy method for personality assessment*'. Still, in daily clinical practice, systematically collecting information from others than the client is not common use. Clinicians generally use interview and observation techniques to determine personality and psychopathology. Incidentally, self-report questionnaires are used and occasionally also information from others by means of an unstructured interview is obtained to complete the diagnostic process. One might think that a

cheap, fast and easy method of gathering multiple informant information on client personality and psychopathology would be embraced by clinicians, but as long as benefits in terms of clinical significance and prognostic value will not be clear, it is not likely that this method will be applied. In the present, exploratory study, we therefore investigated both Vazire's (2006) claim of the feasibility of collecting multiple informant reports through questionnaires and Hofstee's (2009) claim that self-other deviations are clinically relevant.

### The Level of Self-Other Agreement

It is known from the literature that self-other agreement tends to be moderate. Achenbach et al. (2005) found in 108 articles published in the period 1993–2003 qualifying correlations between self-reports and informants' reports on adult psychopathology in both normal and non-normal populations (e.g., inpatients, outpatients, forensic, substance, general population, college, medical, elderly) with several types of informants. The ratings concerned, among others, personality problems, internalizing problems (anxiety, depression, neuroticism, suicidality), externalizing problems (aggression, sociopathy, antisocial behaviour) and substance use (alcohol, tobacco and other drugs). Correlations as high as 0.92 (kappa 0.99) and as low as 0.03 (kappa 0.04) were found. Meta-regression analyses revealed that when self-reports and informant reports were obtained with the same instrument, the mean cross-informant correlations were 0.68 for substance use, 0.43 for internalizing problems and 0.44 for externalizing problems. Klonsky, Oltmanns, and Turkheimer (2002), in their review of the literature on agreement between informant and self-reports of personality disorder, found a median correlation of 0.36 (kappa 0.14) regarding Diagnostic and Statistical Manual (DSM) personality disorders and  $r = 0.47$  in non-DSM domains of personality pathology. Self-informant agreement appeared to be roughly equivalent for questionnaires and interviews. They found a slight trend in the literature for informants to report more personality pathology than subjects. Also Meyer (2002) found cross-informant correlations ranging from 0.27 to 0.44 between self and informant ratings of personality.

### Moderators of Self-Other Agreement

Several possible moderators of this modest overall self-informant agreement and the wide range of

correlations have been investigated. The length of acquaintanceship between subject and informant is supposed to enhance the accuracy of ratings (Kurtz & Sherker, 2003). However, even when raters had been acquainted with the target for up to 70 years, a median correlation of only 0.41 between self-reports and informant reports on personality was found (McCrae & Weiss, 2007). McCrae, Stone, Fagan, and Costa (1998) investigated causes of disagreement between self-reports and spouse ratings of personality on the revised NEO Personality Inventory (NEO-PI-R) (Costa & McCrae, 1992). As they did not find any significant associations between self-spouse (dis)agreement and the hypothesized moderator variables (such as response styles, age, education, similarity and years married), they interviewed the couples about reasons for substantial disagreements. By content analysis of the couples' explanations, 16 reasons were identified from which idiosyncratic understanding of items, reference to different time frames or roles and unavailability for the spouse of covert experience of the subject appeared to be the main reasons.

Biesant, West, and Graziano (1998; see also Biesant & West, 2000) found that individuals with temporally stable response patterns had higher self-other agreement than those with less temporally stable patterns, independent of their relative trait position on the Big Five personality traits. Normatively based variables, such as the participant's variability of item scores (inter-item variability) or the differences between a person's response profile and the normative response profile (scalability and construct similarity), were not found to influence the extent of self-other agreement. In psychiatric samples, the consistency between self-assessment and assessment by significant others proved to vary systematically according to the observability of the personality disorders' criteria. Information from others added particular value when it came to predicting limitations and interpersonal problems (Miller, Pilkonis, & Clifton, 2005; Miller, Pilkonis, & Morse, 2004) or depressive symptoms and personality characteristics (Kamphuis, Emmelkamp, & De Vries, 2003).

In contradiction with these findings and counter to common perceptions are the results of a study of Ready and Clark (2002). They found no overall effect of psychopathology on self-ratings of personality, temperament and interpersonal problems in a non-psychotic psychiatric sample. In their study, self-informant correlations on a Big Five questionnaire ranged from 0.32 (for Agreeableness) to 0.53

(for Openness to Experience). The mean self-other agreement levels for patients with or without a personality disorder were, respectively, 0.47 and 0.40, and for patients with or without a depressive diagnosis, 0.42 and 0.42. Agreement for Openness, however, was significantly higher in depressive patients than in non-depressive patients and in outpatients than in inpatients ( $r = 0.67$  and  $r = 0.31$ , respectively).

It has been suggested that self-assessment in clinical research has the most validity (Dreessen, Hildebrand, & Arntz, 1998). Klein (2003), on the other hand, found that assessment by others yields the most reliable information. Other studies are based on the assumptions that each of the two sources of information yields unique information (e.g., Clark, Bosworth, Welsh-Bohmer, Dawson, & Siegler, 2000; Connolly, Kavanagh, & Viswesvaran, 2007; Klein, 2003; Klonsky et al., 2002; Kraemer et al.; Ready & Clark, 2002; Ready, Watson, & Clark, 2002). It was also found that aggregate ratings by multiple informants correlated higher with observed behaviour than did self-ratings, thus, yielding more accurate information (Kolar, Funder, & Colvin, 1996; Meyer, 2002; Oltmanns & Turkheimer, 2006). Instead of treating discrepancies between self-reports and informant reports as unwanted error to be minimized or to discuss which one is more valid, Clifton, Turkheimer, and Oltmanns (2004) recommended to examine the content of disagreement. Exploration of content is one of the foci of the current study.

### *Examining the Content of Self-Other Disagreement*

A first aim of our study was to investigate whether the collection of multiple informant reports by means of personality questionnaires is a feasible enterprise in daily clinical practice where the burden on clients and informants should be low and efforts and time spent by the diagnostician should be limited. A second aim was to investigate the clinical relevance of self-other disagreement on personality traits in relation to symptoms, diagnoses and therapy outcome. As far as we know, these relationships have not been empirically investigated yet. It is quite conceivable that any self-other disagreement between clients and informants is not restricted to the chosen informants and the behaviour under consideration but generalizes to other subjects and other persons in the client's life as well, including the views and the person of the therapist. Clients may (temporarily)

lack realistic insights due to psycho(patho)logical turmoil. Skilled therapists will be attentive on a non-specific factor such as therapeutic relationship to enhance therapy results. The therapeutic alliance early in therapy is a significant predictor of final treatment outcome (Bachelor & Horvath, 1999). It is important to know whether self–other disagreement should warn the therapist of possible threats of the therapeutic relationship. In line with the theoretical notions of Hofstee (2009) we expected that:

1. the larger the discrepancy between self-report and (averaged) informants' reports on personality, the more (personality) pathology in terms of symptoms and DSM-IV diagnoses is found; and
2. the larger the discrepancy between self-report and (averaged) informants' reports, the more a positive therapy outcome, defined as a reduction in symptoms and an improvement in functioning, will be limited.

## METHOD

### Participants

The participants in the study were 105 clients referred by a primary care physician at the first author's psychotherapy practice in the period from October 2006 to November 2007. The sample was composed of 25 men and 80 women ranging in age from 18 to 69 years, with a mean age of 36 (standard deviation [SD] = 12) years. Of the sample, 67% were married or cohabiting, 33% were unmarried or divorced. Educational levels were 13% elementary school or lower vocational level, 31% intermediate vocational education (USA: high school), 45% higher vocational education (USA: bachelor's degree) and 11% university education (USA: master's degree/PhD). More than two thirds of the clients were employed, 16% were unemployed, 12% were attending higher education and 4% were retired. A large percentage of the clients had received psychological help in the past: 34% first-line help (primary mental health care), 23% second-line help (specialized outpatient mental health care) and 5% third-line help (being hospitalized).

The participants provided self-ratings on a number of questionnaires (here, we report on those relevant for our research questions). In addition, each participant was asked to get personality ratings from three well-acquainted informants; in total, 303 informant reports were actually returned. Three participants

(3%) returned only one other report. These cases were excluded from further analyses. Six of the participants returned two informant reports and 96 participants provided three informant reports. The 300 informants in the analyses were 76 partners (25%), 105 family members (35%), 82 friends (27%), 29 (10%) other and eight (3%) unspecified acquaintances. So for  $n = 102$  participants with a minimum of two informant ratings, averaged informant reports on personality were available for the analyses, which is a response rate of 97%.

### Measures

#### *Five-Factor Personality Inventory (FFPI)*

The FFPI (Hendriks, 1997; Hendriks, Hofstee, & De Raad, 1999a, 1999b, 2002) assesses the Dutch psycholexically based Big Five personality dimensions Extraversion, Agreeableness, Conscientiousness, Emotional Stability and (intellectual) Autonomy. The latter construct is a measure of independent decision-making. The instrument is suitable for self-ratings and other ratings and has been designed to be broadly applicable because of its item format and wording. The response scale runs from 1 (*not at all applicable*) to 5 (*entirely applicable*). Strict guidelines were used for item construction, such as to phrase items as brief and simple as possible, to avoid negations and modifiers and to phrase items in observable terms. The 100 items satisfy the criterion of being comprehensible at lower levels of education. In the normal population, the FFPI has been found to be reliable and construct-valid, both nationally and internationally (Hendriks et al., 2002, 2003). In the present sample, the structure of the FFPI showed fair to good similarity with the structure in the normative sample (Hendriks et al., 1999b): congruence coefficients Tucker's phi ranged from 0.86 (for Autonomy) to 0.94 (for Extraversion and Emotional Stability) for the self-reports and 0.91 (for Agreeableness) to 0.96 (for Extraversion and Emotional Stability) for the averaged informant reports. The internal consistency reliability (stratified alpha) of the FFPI factors was good: it ranged from 0.81 (Autonomy) to 0.89 (Extraversion) for the self-reports and from 0.82 (Autonomy) to 0.91 (Extraversion) for the averaged informant reports.

#### *Nederlandse Verkorte Minnesota Multiphasic Personality Inventory (MMPI) (NVM: Dutch Short Form of the MMPI)*

The NVM (Luteijn & Kok, 1985) is composed of 83 items derived from the MMPI (Hathaway & McKinley, 1943). Five scales are differentiated:



Negativism, Somatization, Shyness, Serious Psychopathology and Extraversion. The internal consistency reliability (Cronbach's alpha) of these scales was satisfactory, ranging from 0.71 (Extraversion) to 0.89 (Shyness), except for Serious Psychopathology (0.64 at t1 and 0.21 at t2).

#### *Symptom Check List-90 (SCL-90)*

The SCL-90 (Arrindell & Ettema, 1986) is an inventory of physical and emotional symptoms. The questionnaire is composed of 90 descriptions comprising eight scales: Anxiety, Agoraphobia, Depression, Somatic complaints, Insufficiency of Cognition and Action, Suspiciousness and Interpersonal Sensitivity, Hostility, and Sleep problems. The sum across all items, labelled Psychoneuroticism, indicates the overall level of emotional and physical dysfunction. In the present sample, the Cronbach's alpha ranged from 0.68 for Agoraphobia to 0.97 for Psychoneuroticism.

#### *DSM of Mental Disorders 4th Edition (DSM-IV-TR)*

Diagnoses were classified according to the five axes of DSM-IV-TR (Dutch translation: Koster van Groos, 2001). Axis V, the Global Assessment of Functioning (GAF) Scale, was coded in 10-point intervals, starting at 1 (GAF = 40 or below) to 7 (GAF = 91–100).

#### *Treatment Characteristics*

The duration of treatment was established in terms of the number of 45-minute sessions as well as the number of weeks. A five-point scale (from 1 = poor to 5 = very good) was used to determine the quality of contact as experienced by the therapist. Criteria were the number of no-shows, the number of cancelled appointments, continuity, progression, cooperation, use of insight acquired, effort and motivation (see also: Mosterman, Eurelings-Bontekoe, & Hofstee, 2008). The conclusion of therapy was coded: therapy completed, further therapy recommended or therapy broken off by the client (dropout).

#### *Procedure*

Prior to the first appointment, the clients completed a registration form detailing socio-demographic data. Their problems, symptoms and the requested assistance were assessed in detail during the first interview. At the end of this session, the clients were given the self-report questionnaires to be completed at home. In addition, they received

three envelopes containing FFPI to be handed to well-acquainted informants who the client believed to know well. The instruction was not to discuss ratings. The informants were instructed to seal the envelope after completion of the questionnaire and return it to the client. The informants were assured that their ratings would remain confidential. They were explained that an average personality profile of their target person across raters would be calculated and that no individual answers would be discussed with the ratee.

The completed questionnaires were collected during the client's second interview session. Subsequently, the diagnostic procedure was continued by assessing the client's development and biography. The questionnaires were scored afterwards. Based on clinical interview and observation, supported by results of the self-report questionnaires, the DSM-IV classification was determined on five axes. Neither the information from the informant reports nor the self–other discrepancies were used in determining diagnoses or in carrying out therapy. An indication report that included the questionnaire interpretation, DSM-IV classification and treatment design was written. All data collected in this first phase were recorded as t1.

In the third session, the indication report was discussed with the client. The FFPI results for the self-reports and the averaged informant reports were fed back only globally, according to the nine verbally limited intervals as specified in the FFPI manual (extreme, clearly, fairly and somewhat at each of the two poles of the personality trait with a neutral category in the middle), with no interpretation of possible meanings of any discrepancies. For instance, the client was told that she rated herself as clearly extraverted, while her informants rated her, on the average, as less extraverted, but still, as fairly extraverted. Subsequently, therapy took place according to standard practice.

At the end of the agreed period of therapy, the therapist determined the treatment characteristics as well as the quality of functioning (Axis V). In addition, a number of clients were willing to complete the self-report questionnaires again and to ask their informants to complete the FFPI once more, resulting in a sample of  $n = 48$ . These data were recorded as t2.

#### *Analyses*

For each participant, absolute discrepancy scores between the self-reports and the averaged informant reports per FFPI factor were determined.

Table 1. Descriptive statistics of FFPI factor scores based on self-reports and averaged informant reports

Factor	FFPI factor scores					
	Self-report	Other report	Agreement		Disagreement	
	M (SD)	M (SD)	<i>r</i>	<i>e'</i>	Range	M (SD)
Extraversion	0.36 (1.15)	0.84** (1.05)	0.78	0.71	0.01–2.40	0.71 (0.53)
Agreeableness	2.21 (0.95)	2.09 (0.85)	0.40	0.40	0.02–3.03	0.78 (0.62)
Conscientiousness	1.08 (1.09)	1.10 (0.85)	0.69	0.67	0.00–2.03	0.60 (0.51)
Emotional stability	–0.11 (1.14)	–0.26* (0.84)	0.78	0.74	0.02–1.78	0.59 (0.41)
Autonomy	0.91 (1.02)	1.06 (0.79)	0.61	0.58	0.01–2.44	0.65 (0.52)

\* $p < 0.05$ . \*\* $p < 0.01$ .

Note.  $n = 102$ . Agreement:  $r$  = relative,  $e'$  = absolute. Disagreement: absolute difference between factor scores based on self-report and based on averaged informant reports.

Overall self–other disagreement across all five factors was established by calculating the Euclidean distance (the square root of the sum of the squared differences per factor). The Euclidean distance is a commonly used measure of dissimilarity between profiles based on self-reports and informant reports. Self–other differences on any element (e.g., personality factors) in a profile may each be below the threshold of exerting an effect, while together, an effect of the differences may turn out to be significant. The disagreement scores were used in further analyses. We reported one-sided  $p$ -values for the hypothesized relationships.

We also calculated the amount of agreement between the FFPI factor scores based on the self-ratings and based on the averaged informant reports. Relative agreement was established by the Pearson correlation coefficient. Absolute agreement was established by the corrected identity coefficient  $e'$  (Zegers, 1986; Zegers & Ten Berge, 1985).<sup>1</sup>

The Reliable Change Index (RCI, Jacobson & Truax, 1991) was calculated to investigate the degree of change on an individual level. The RCI is viewed as a means of measuring clinical sig-

nificance, as a more relevant tool to supplement statistical significance. In calculating the RCI, measurement errors in the instrument are discounted, resulting in a reliable individual measure of change.

## RESULTS

### Self–Other Disagreement

Table 1 presents the means, standard deviations and discrepancy scores of the five FFPI factors as well as the level of relative and absolute self–other agreement. Paired samples  $t$ -tests revealed that, on average, informants judged their ratee as statistically significantly more extraverted than the ratees judged themselves,  $t(101) = -6.58$ ,  $p < 0.001$ ,  $d = 0.43$ , and less emotionally stable,  $t(101) = 2.02$ ,  $p = 0.046$ ,  $d = 0.14$ . Relative agreement was highest for Extraversion and Emotional Stability (both:  $r = 0.78$ ) and lowest for Agreeableness ( $r = 0.40$ ). Absolute agreement was highest for Emotional Stability ( $e' = 0.74$ ) and lowest for Agreeableness ( $e' = 0.40$ ). Self–other disagreement in the values of the factor scores (absolute discrepancy scores) ranged from 0.00 (Conscientiousness) to 3.03 (Agreeableness). Overall disagreement across the five factors (Euclidean distance) ranged from 0.40 to 3.77 (Mean [ $M$ ] = 1.79,  $SD = 0.65$ ).

Self–other disagreement showed no relationship with gender, age, education, previous therapeutic help, work situation and work disability. The clients who were married or cohabiting ( $n = 68$ ), demonstrated significantly more self–other disagreement on Autonomy than those living alone (unmarried or divorced,  $n = 34$ ),  $t(87) = 2.22$ ,  $p = 0.029$ . Marital status was not associated with symptoms assessed with the SCL-90, DSM-IV-

<sup>1</sup> The identity coefficient is an association measure for absolute scales (like FFPI factor scores; e.g., Hendriks et al., 1999a), which respects information concerning differences in both level (means) and dispersion (variances). The value of  $e'$  is constrained to be between  $-1$  and  $+1$ . Whereas the Pearson correlation coefficient answers the question 'to what extent are the two variables linearly related?', the identity coefficient answers the question 'to what extent are the two variables identical?' (Fagot & Mazo, 1989). The corrected identity coefficient differs from the uncorrected one in the sense that the corrected coefficient assesses the amount of agreement in excess of what is to be expected by chance. It is therefore the more appropriate measure of the two (Fagot & Mazo, 1989).

Table 2 Spearman's rank correlations between self-other disagreement on the FFPI and Questionnaire Scale scores at t1

Scale	self-other discrepancies t1					
	E	A	C	ES	Au	Profile
FFPI extraversion (self)	-0.32**	0.09	-0.23*	0.09	-0.06	-0.21*
FFPI agreeableness (self)	-0.04	0.17	0.08	0.03	0.08	0.16
FFPI conscientiousness (self)	0.10	0.12	-0.17	-0.02	0.05	0.00
FFPI emotional stability (self)	0.02	-0.03	-0.09	0.16	0.04	-0.04
FFPI autonomy (self)	-0.10	0.00	0.05	0.02	-0.28**	-0.07
FFPI extraversion (others)	0.03	-0.01	-0.09	0.10	-0.01	-0.04
FFPI agreeableness (others)	0.14	-0.11	0.04	0.01	-0.01	-0.08
FFPI conscientiousness (others)	-0.09	0.11	-0.29**	-0.04	0.04	-0.09
FFPI emotional stability (others)	0.01	-0.06	0.00	0.14	-0.04	0.03
FFPI autonomy (others)	-0.08	-0.18	-0.03	0.01	-0.11	-0.08
NVM negativism	0.01	-0.05	0.15	-0.14	0.03	0.06
NVM somatization	-0.01	-0.12	0.06	-0.16	0.04	0.01
NVM shyness	0.15	0.02	0.02	-0.04	0.25*	0.24*
NVM psychopathology	0.17	0.09	0.00	-0.07	-0.11	0.03
NVM extraversion	-0.08	0.10	0.00	0.14	0.00	-0.01
SCL-90 anxiety	-0.01	-0.07	0.16	-0.01	-0.07	0.04
SCL-90 agoraphobia	-0.03	-0.17	0.19	-0.19	-0.08	-0.03
SCL-90 depression	0.10	0.01	0.16	0.01	0.15	0.27**
SCL-90 somatic complaints	-0.02	-0.13	0.06	-0.02	0.00	0.02
SCL-90 insufficiency	0.04	-0.21*	0.16	-0.02	0.09	0.03
SCL-90 sensitivity	0.02	0.01	0.14	-0.14	0.09	0.15
SCL-90 hostility	0.13	0.09	0.13	0.02	0.13	0.24*
SCL-90 sleep problems	-0.03	0.01	0.09	-0.01	0.01	0.07
SCL-90 total	0.05	-0.04	0.19	-0.03	0.10	0.18

\* $p < 0.05$ . \*\* $p < 0.01$ .

Note.  $n = 102$ . E = Extraversion. A = Agreeableness. C = Conscientiousness. ES = Emotional Stability. Au = Autonomy. Profile = entire FFPI profile.

classifications or treatment characteristics so there was no need to control for this variable in the further analyses.

### *Self-Other Disagreement in Relation to Symptoms and Structural Psychopathology*

In Table 2, it can be seen that larger Euclidean distances on the entire profile were associated with higher scores on shyness, depression and hostility and with lower scores on self-reported extraversion. Different FFPI factors were involved: self-other discrepancies on Extraversion showed an inverse relation with self-reported extraversion scores. Greater self-other discrepancies on Conscientiousness were associated with relatively low conscientiousness scores as reported by others and low extraversion scores as reported by self. Greater self-other discrepancies on Autonomy were associated with relatively low scores on self-reported autonomy and relatively high scores on shyness.

### *Self-Other Disagreement in Relation to DSM-IV classifications*

Because of the small numbers in many DSM-IV categories, no reliable analyses could be carried out. Therefore, a simplification of the categorization in axes I and II was undertaken. We clustered Axis I diagnoses in mood disorders ( $n = 37$ ), anxiety disorders ( $n = 22$ ) and other disorders ( $n = 43$ ). Axis II was recoded into a dichotomous personality problem variable. Fifty-two ( $n = 52$ ) of the participants showed personality problems (six of them were diagnosed with a personality disorder; 46 participants had pathological personality traits, but did not meet the criteria to diagnose a personality disorder). There were no personality problems for  $n = 50$ . In  $n = 72$  participants, multiple problems (co-morbidity) on one or more axes were observed, whereas  $n = 30$  had no co-morbidity. A one-way analysis of variance on the Euclidean distance as the dependent variable demonstrated significant differences between the three DSM-IV axis I groups,

$F(2, 99) = 3.81, p = 0.025$ . The responsible factor for these differences was Autonomy,  $F(2, 99) = 3.37, p = 0.038$ . *Post hoc* comparisons (Scheffé) showed that the participants with mood disorders (Euclidean distance  $M = 1.97$ ; Autonomy  $M = 0.79$ ) displayed significantly more self–other disagreement than the participants with anxiety disorders (Euclidean distance  $M = 1.51$ ; Autonomy  $M = 0.43$ ),  $M_{(\text{diff})} = 0.47, p = 0.026$  and  $M_{(\text{diff})} = 0.36, p = 0.039$ , respectively).

An independent samples *t*-test indicated that personality problems were associated with more overall self–other disagreement,  $t(100) = -2.25, p = 0.014$ , one-sided,  $d = 0.45$ . Self–other discrepancies on Extraversion and on Conscientiousness counted for the disagreement,  $t(100) = -2.55, p = 0.006$  and  $t(100) = -2.61, p = 0.005$ , respectively. Also comorbidity was statistically significantly associated with more overall self–other disagreement,  $t(100) = -1.89, p = 0.031$ , one-sided, due to discrepancies on Extraversion,  $t(76,41) = -2.08, p = 0.021$  and on Conscientiousness  $t(72,25) = -4.01, p < 0.001$ . The GAF scores varied from range 41–50 to range 81–90; the average GAF score corresponded with a range of 61–70. The GAF scores were not significantly associated with overall self–other disagreement.

### Self–Other Disagreement in Relation to Treatment Characteristics

The average number of sessions was 10 ( $SD = 6$ ) with a minimum of three and a maximum of 45 sessions. The sessions took place for an average of 22 weeks ( $SD = 12$ ) with a minimum of three and a maximum 76 weeks. Sixty-four of the 102 treatments were completed at t2; in 23 cases, further

therapy was recommended, and 15 times the contact was broken off by the client. The average quality of contact experienced by the therapist was 3.3 ( $SD = 1.3$ ) on a five-point scale. We found a significant negative correlation between the Euclidean distance and the quality of contact ( $r_s = -0.22, p = 0.014$ , one-sided): the more overall disagreement, the less positive the therapeutic contact proved to be. This association disappeared, however, when the dropouts (15%) were excluded from the analysis ( $n = 87, r_s = -0.09, p = 0.206$ ). In comparison, with the participants who continued treatment, the dropouts had significantly larger self–other discrepancies on Conscientiousness,  $t(100) = -3.37, p < 0.001, d = 0.90$ . No associations were found between overall self–other disagreement and the duration of therapy or the number of sessions.

### Self–Other Disagreement in Relation to Therapy Outcome

Because of the limited sample size at t2, we checked for a possible selection bias by comparing these 48 participants with the 54 participants for whom only t1 data were available. No significant differences were found on Euclidean distances, questionnaire scores and on DSM-IV classifications.

Paired samples *t*-tests demonstrated that at the group level, significant changes in symptoms and functioning took place between t1 and t2. The effect sizes for the decrease in symptoms on the SCL-90 scales, with the exception of agoraphobia, ranged from moderate ( $d > 0.5$ ) to large ( $d > 0.8$ ). The effect size of the change in functioning (GAF) was very large:  $d = 3.07$ . In Table 3, the results of these

Table 3. Therapy outcomes

	t1		t2		$M_{(\text{diff})}$	$d$	$r_{xx}$	RCI (in%)		
	M	SD	M	SD				–	0	+
SCL-90 anxiety	18.10	5.45	14.04	5.05	–4.06***	0.77	0.37	35	61	4
SCL-90 agoraphobia	8.85	2.48	8.46	3.71	–0.40	0.13	0.05	10	86	4
SCL-90 depression	34.79	10.46	24.21	8.70	–10.58***	1.10	0.44	56	40	4
SCL-90 somatic complaints	20.23	6.75	16.00	4.05	–4.23***	0.78	0.65	19	81	0
SCL-90 insufficiency	19.17	6.19	14.71	5.23	–4.46***	0.78	0.27	40	52	8
SCL-90 sensitivity	32.19	8.92	26.25	7.18	–5.94***	0.74	0.44	40	54	6
SCL-90 hostility	9.21	3.18	7.65	2.45	–1.56***	0.55	0.63	17	83	0
SCL-90 sleep problems	6.54	2.89	5.25	2.49	–1.29*	0.48	0.16	29	67	4
SCL-90 total	163.83	37.60	128.48	35.74	–35.35***	0.96	0.46	73	21	6
GAF score	3.71	0.58	4.85	0.17	1.15***	3.07	0.35	–	–	–

\* $p < 0.05$ . \*\* $p < 0.01$ . \*\*\* $p < 0.001$ .

Note.  $n = 48$ . RCI = Reliable change index.  $d$  = effect size:  $M_{(\text{diff})}/([SD_1 + SD_2]/2)$ .  $r_{xx}$  = test–retest correlation. GAF score: 3 = range 51–60; 4 = range 61–70; etc.



Table 4. Linear regressions of changes in scale scores on disagreement-related scales on changes in self-other disagreement

Dependent variable (t2-t1)	Independent variable (t2-t1)	B	SE	$\beta$	$t(46)$	$p$	$R^2$	$F(1,46)$
SCL-90 depression	Euclidean distance	4.39	2.04	0.30	2.15	0.037	0.09	4.63
	Autonomy distance	5.76	2.11	0.37	2.74	0.009	0.14	7.49
SCL-90 hostility	Euclidean distance	1.05	0.50	0.30	2.10	0.041	0.09	4.43
	Autonomy distance	1.23	0.53	0.33	2.34	0.024	0.11	5.48
NVM shyness	Euclidean distance	3.14	1.02	0.42	3.09	0.003	0.17	9.57
	Autonomy distance	3.51	1.07	0.44	3.29	0.002	0.19	10.83

Note.  $n = 48$ . Method for Euclidean distance: enter; method for FFPI factor distances: stepwise

changes as the main measures of therapy effect are displayed. The last three columns in Table 3 list the percentages of participants for whom there was a clinically significant increase or decrease per scale ( $RCI \leq -1.96$  c.q.  $RCI \geq 1.96$ ) and those for whom change was not found clinically significant.

At t2, no associations between self-other disagreement and NVM and SCL-90 scale scores were found, except for hostility, for which the correlation remained significant,  $r_s = 0.29$ ,  $p = 0.022$  (one-sided),  $n = 48$ . The disappearance of the associations at t1 between self-other disagreement and depression and shyness could not be attributed to changes in self-other disagreement between t1 and t2 at the group level: paired sample  $t$ -tests did not show any significant changes in overall self-other disagreement or factor discrepancies between t1 and t2. The mean Euclidean distance for  $n = 48$  was 1.74 (SD = 0.65) at t1 and 1.68 (0.58) at t2.

### Self-Other Disagreement and Its Predictive Value

In order to investigate whether self-other disagreement at t1 was related to therapy outcome, we conducted linear regression analyses of the changes in symptoms on self-other disagreement at t1 using only those symptom scales that were found to be associated with disagreement at t1: SCL-90 depression, SCL-90 hostility and NVM shyness. No significant predictions could be made from t1 self-other disagreement with respect to t1-t2 changes in these symptom scales. However, changes in self-other disagreement between t1 and t2 did significantly predict changes in those scale scores. Changes in the mean Euclidean distance appeared to be attributable mainly to changes in disagreement on Autonomy, see Table 4.

## DISCUSSION

### Vazire's Claim: Feasibility

In this exploratory study, we found support for the feasibility of the use of multiple informant reports by means of personality questionnaires in clinical practice. The response rate was 97%. Data collection and processing did not prove to be a very time-consuming or expensive enterprise.<sup>2</sup> Our learning is that both clients and informants were cooperative and were interested in and attached value to the findings, although this cooperation had faded at the repeated assessment at t2, which resulted in a participation decline at t2 of over 50%. The clients were not obliged or pressed to participate.

In our opinion, symptoms, interpersonal context and personality are inextricably related and should all be taken into account in assessing a client in order to establish an appropriate therapeutic relationship and to choose an optimal therapeutic method. In achieving this, diagnostics can be viewed as therapeutic interventions with which the motivation of the client can be increased and a positive therapeutic relationship can be advanced (see, e.g., Finn, 2007). The use of the supplementary diagnostic method described in this study can be seen as such an intervention, at the very least,

<sup>2</sup>It should be noted that our expenses could stay low because we had free use of the FFPI, which in the Netherlands, is a commercially published questionnaire. The costs for applied clinical use will be substantially higher if a personality questionnaire has to be commercially obtained, considering four test runs (one self and three informants' ratings) per client. It is worth mentioning, however, that there are other psychometrically sound personality inventories available in the public domain (i.e., at no cost) like the Big Five scales of the International Personality Item Pool (IPIP; Goldberg et al., 2006): <http://www.ipip.ori.org/ipip/>.

as Hofstee (2009) stated because the client could benefit from the confrontation between the perspectives of self and averaged others. One could argue that providing feedback biased the results at t2. It can also be taken into account as a part of therapy effect. We do not know yet whether or not any changes in self–other disagreement occurred spontaneously, are side effects of therapy, or have to be seen as an effect of the feedback. Fact is that a decrease of disagreement was mainly due to changes in clients' self-judgments.

### *Hofstee's Claim: Clinical Relevance*

We found self–other disagreement to be a relatively independent constant in terms of the personality characteristics determined with the instrument. At a group level, a significant discrepancy in reports between self and informants was found on Extraversion and Emotional Stability, where the participants viewed themselves, on average, as less extraverted and more emotionally stable than their proxies did. There proved to be great variation over the participants in the degree of disagreement. These individual differences did not appear to be random but seem to reflect, in view of their relationships with other diagnostic data, a structure with a meaningful content. Self–other disagreement occurred predominantly in introverted, shy, hostile and depressed persons, who tended to have more personality problems. These results are in line with Hofstee's claim that systematic deviations may be found clinically relevant, and also with the findings of Achenbach et al. (2005) that self-informant correlations on internalizing problems and personality problems were low.

At first sight, the results seem to contradict the findings of Ready and Clark (2002), who found only marginal effects of psychopathology on self-ratings. For instance, these authors found no differences in average self–other agreement on the Big Five personality dimensions between depressed and non-depressed patients (for both groups, mean  $r = 0.42$ ), while we did find a significant relationship between overall self–other discrepancy and depression (see Table 2). However, the results are not readily comparable. Had we followed Ready and Clark, the average self–other agreement across the Big Five between the depressed ( $r = 0.63$ ,  $n = 37$ ) and the non-depressed patients ( $r = 0.67$ ,  $n = 65$ ), and between patients with ( $r = 0.67$ ,  $n = 52$ ) and without personality problems ( $r = 0.63$ ,  $n = 50$ ) in our sample would also have turned out to be non-significantly different, respectively,  $z = -0.32$ ,  $p = 0.37$  and  $z = -0.34$ ,

$p = 0.37$ , one-sided. However, by investigating self–other disagreement instead of self–other agreement, we did find differences between the subgroups. Incidentally, the magnitudes of the agreement correlations—0.6 in our study compared with 0.4 in the Ready and Clark study, where (mostly) one informant was used—illustrate the benefit of using multiple informants: it maximizes reliability, which sets the boundary for validity.

In view of the lack of cohesion with the overall symptom level and the level of functioning at the start of therapy, self–other disagreement does not appear to reflect current pathological state but rather to represent a more structural characteristic of either the client's personality or his or her interpersonal relationships. This interpretation is supported by the finding that, on average, symptoms decreased and functioning increased but self–other disagreement did not significantly change between t1 and t2. All together, we tentatively conclude that the results confirm our first hypothesis, stating that self–other disagreement on personality reflects psychopathology.

### *Prediction of Therapy Outcome*

For our second hypothesis, which states that self–other disagreement on personality reduces therapy effect, we found less convincing support. Self–other disagreement was not significantly negatively associated with reduction of symptoms or increase in functioning. However, an interesting and important finding is that self–other disagreement strongly predicted dropout. Both self–other disagreement and dropout were significantly negatively associated with the quality of the therapeutic relationship, but a direct association between self–other disagreement and the therapeutic relationship was no longer demonstrated once dropout was statistically controlled for. Furthermore, we found that a decrease in depression, hostility and shyness goes together with a decrease in the overall self–other discrepancy on the client's personality profile. A decrease in disagreement on Autonomy appeared to play a key role in these changes, whereas disagreement on Conscientiousness was the main predictor of dropout.

### *Remaining Questions*

The results do not answer the question of causality: does self–other disagreement precede psychopathology or *vice versa*? Are there isolating and

alienating consequences for oneself and one's social environment when there is constant experience of self-other discrepancies, which may constitute a factor in the development of psychopathology? Or does psychopathology create self-other disagreement because of the isolating and alienating effects of being 'ill'. Does self-other disagreement maintain existing psychopathology, partially because the risk of dropout and a less optimal treatment relationship is greater, and therefore, the client remains cut off from the assistance he or she needs?

Other interesting questions that were beyond the scope of this study are: whose ratings predict therapy outcome better: those of clients or those of informants? Are some (type of) informants more accurate than others? How many informants do one need: Would two be sufficient in terms of reliability? Does the amount of inter-informant agreement play a role in predicting therapy outcome? We hope to address these questions in further papers.

### Limitations

There are some limitations to this study that should be mentioned. The informants may have been influenced by the lack of total confidentiality in the design. The fact that the clients were to receive feedback about the aggregate of their informants' ratings may have affected the informants' frankness. As the informants were explained that confidentiality was secured because no information on individual or item level would be reported back, this should not be a big problem. Furthermore, the amount of self-other agreement in this sample appeared to match the amount of self-other agreement in samples in the general population, where no feedback was given. Another possible bias concerns the fact that the therapist was not blind for the results of the self-report and informant-reports: they unintentionally may have influenced diagnoses or determining the quality of the relationship. To this, it can be said that the therapist did not know the level of overall disagreement nor the meaning or normality of the factor discrepancies in light of the full sample and the analyses. Even if her judgments were influenced, bias should be reasonably random.

With regard to the generalizability of the results, we compared our sample with data from a national professional association of primary health-care psychologists (LVE, 2007) who collected informa-

tion from about 84 000 clients over the year 2007. It appeared that the male-female ratio differed as it was 35:65 in the national sample versus 24:76 in our sample. No differences were found on other socio-demographic variables. The dropout percentages differed slightly: 18% in the national sample against 15% in our sample. As we did not find any gender effect in our results, generalizability should not be affected much.

Another constraint concerns the restricted sample size at t2. Although no significant differences were found between the t1-only group ( $n = 54$ ) and the t1 + t2 group ( $n = 48$ ) with regard to self-other disagreement, questionnaire scores and DSM-IV classifications, a possible restriction of range in variance in the scores may have obscured the results. A lack of variance sets limits to the magnitude of an association between variables. Based on the Levene test, a statistically significant ( $p < 0.01$ ) difference in variance was found between the t1 + t2 group and the t1-only group for the SCL-90 insufficiency scale ( $SD = 6.3$  versus  $SD = 8.6$ ) and the SCL-90 total score ( $SD = 37.9$  versus  $SD = 51.7$ ). In these two variables, the variance in the analyses proved to be limited. As these variables did not appear to play a key role in our study, the impact of restriction of range on the results should be limited.

### Conclusion

The results of the present study can only lead to a tentative conclusion. The clinical relevance of assessing self-other disagreement on the client's personality seems two-fold. First, information on the amount of disagreement might supplement the diagnostic process, in that greater self-other discrepancies can signal the presence of personality pathology as well as a greater risk of dropout. Second, information on the amount of self-other disagreement might supplement the therapy process in that the therapist could discuss the validity of the client's view in light of this client's average informant's view in order to improve specific therapy results. These discussions may help investigate whether specific therapeutic methods are needed to enhance, for instance, a client's levels of autonomy. Also a focus on the client's level of conscientiousness early in therapy could be considered as larger self-other discrepancies concerning conscientiousness proved to be strong predictors of dropout.

In this practice-based study, the contours of the interpretation of self-other disagreement on personality begin to appear; further analyses and rep-

lication studies in larger samples will be necessary in order to refine the content and structural validity of the present findings.

## REFERENCES

- Achenbach, T.M., Krukowski, R.A., Dumenci, L., & Ivanova, M.Y. (2005). Assessment of adult psychopathology: Meta-analysis and implications of cross-informant correlations. *Psychological Bulletin*, 131, 361–382.
- Arrindell, W.A., & Ettema, J.H.M. (1986). *SCL-90. Handleiding bij een multidimensionele psychopathologie-indicator [SCL-90 multidimensional psychopathology indicator professional manual]*. Lisse: Swets & Zeitlinger.
- Bachelor, A., & Horvath, A. (1999). The therapeutic relationship. In M.A. Hubble, B.L. Duncan, & S.D. Miller (Eds), *The heart and soul of change: What works in therapy* (pp. 133–178). Washington, DC: American Psychological Association.
- Biesantz, J.C., & West, S.G. (2000). Personality coherence: Moderating self–other profile agreement and profile consensus. *Journal of Personality and Social Psychology*, 79, 425–437.
- Biesantz, J.C., West, S.G., & Graziano, W.G. (1998). Moderators of self–other agreement: Reconsidering temporal stability in personality. *Journal of Personality and Social Psychology*, 75, 467–477.
- Clark, L.M., Bosworth, H.B., Welsh-Bohmer, K.A., Dawson, D.V., & Siegler, I.C. (2000). Relation between informant-rated personality and clinician-rated depression in patients with memory disorders. *Neuropsychiatry, Neuropsychology, and Behavioral Neurology*, 13, 39–47.
- Clifton, A., Turkheimer, E., & Oltmanns, T.F. (2004). Contrasting perspectives on personality problems: Descriptions from the self and others. *Personality and Individual Differences*, 36, 1499–1514.
- 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, 110–117.
- Costa, P.T. Jr, & McCrae, R.R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessments Resources.
- Dreessen, L., Hildebrand, M., & Arntz, A. (1998). Patient–informant concordance on the Structured Clinical Interview for DSM-III-R personality disorders (SCID-II). *Journal of Personality Disorders*, 12, 149–161.
- Fagot, R.F., & Mazo, R.M. (1989). Association coefficients of identity and proportionality for metric scales. *Psychometrika*, 54, 93–104.
- Finn, S.E. (2007). In *our clients' shoes: Theory and techniques of therapeutic assessment*. Mahwah, NJ: Erlbaum.
- Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R., & Gough, H.C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, 84–96.
- Hathaway, S.R., & McKinley, J.C. (1943). *The Minnesota multiphasic personality inventory*. New York: Psychological Corporation.
- Hendriks, A.A.J. (1997). *The construction of the Five-Factor Personality Inventory (FFPI)*. Unpublished doctoral dissertation, Groningen: University of Groningen.
- Hendriks, A.A.J., Hofstee, W.K.B., & De Raad, B. (1999a). The Five Factor Personality Inventory (FFPI). *Personality and Individual Differences*, 27, 307–325.
- Hendriks, A.A.J., Hofstee, W.K.B., & De Raad, B. (1999b). *Handleiding bij de Five-Factor Personality Inventory (FFPI) [The Five-Factor Personality Inventory professional manual]*. Lisse: Swets Test Publishers.
- Hendriks, A.A.J., Hofstee, W.K.B., & De Raad, B. (2002). The Five Factor Personality Inventory: Assessing the Big Five by means of brief and concrete statements. In B. De Raad & M. Perugini (Eds), *Big Five assessment* (pp. 79–108). Göttingen: Hogrefe & Huber.
- Hendriks, A.A.J., Perugini, M., Angleitner, A., Ostendorf, F., Johnson, J.A., De Fruyt, F., Hrebickova, M., Kreitler, S., Murakami, T., Bratko, M., Conner, M., Nagy, J., Rodriguez-Fornells, A., & Ruisel, I. (2003). The Five-Factor Personality Inventory: Cross-cultural generalizability across 13 countries. *European Journal of Personality*, 17, 347–373.
- Hofstee, W.K.B. (1994). Who should own the definition of personality? *European Journal of Personality*, 8, 149–162.
- Hofstee, W.K.B. (2009). Restyling personality assessments. In L.B. Palcroft & M. V. Lopez (Eds), *Personality assessment: New research* (pp. 223–235). Hauppauge, NY: Nova Science Publishers.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 50, 12–19.
- Kagan, J. (2007). A trio of concerns. *Perspectives on Psychological Science*, 2, 361–376.
- Kamphuis, J.H., Emmelkamp, P.M.G., & De Vries, V. (2003). Moderated validity of clinical informant assessment: Use in depression and personality. *Clinical Psychology & Psychotherapy*, 10, 102–107.
- Klein, D.N. (2003). Patients' versus informants' reports of personality disorders in predicting 7 1/2-year outcome in outpatients with depressive disorders. *Psychological Assessment*, 15, 216–222.
- Klonsky, E.D., Oltmanns, T.F., & Turkheimer, E. (2002). Informant-reports of personality disorder: Relation to self-reports and future research directions. *Clinical Psychology*, 9, 300–311.
- Kolar, D.W., Funder, D.C., & Colvin, C.R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality*, 64, 311–337.
- Koster van Groos, G.A.S. (2001). *Beknopte handleiding bij de Diagnostische Criteria van de DSM-IV-TR. [Short Manual of the DSM-IV-TR]*. Lisse: Swets & Zeitlinger.
- Kraemer, H.C., Measelle, J.R., Ablow, J.C., Essex, M.J., Boyce, W.T., & Kupfer, D.J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching



- contexts and perspectives, *American Journal of Psychiatry*, 160, 1566–1577.
- Kurtz, J.E., & Sherker, J.L. (2003). Relationship quality, trait similarity, and self-other agreement on personality traits in college roommates. *Journal of Personality*, 71, 21–48.
- Luteijn, F., & Kok, A.R. (1985). *NVM, Nederlandse verkorte MMPI, herziene handleiding* [NVM, Dutch short form of the MMPI professional manual revised]. Lisse: Swets & Zeitlinger.
- LVE. (2007). *Jaarbericht 2007*. [Annual Report 2007]. Amsterdam: Landelijke Vereniging van Eerstelijnspsychologen.
- McCrae, R.R., Stone, S.V., Fagan, P.J., & Costa, P.T. Jr. (1998). Identifying causes of disagreement between self-reports and spouse ratings of personality. *Journal of Personality*, 66, 285–313.
- McCrae, R.R., & Weiss, A. (2007). Observer ratings of personality. In R.W. Robins, R.C. Fraley, & R.F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 259–272). London: The Guilford Press.
- McDonald, J.D. (2008). Measuring personality constructs: The advantages and disadvantages of self-reports, informant reports and behavioural assessments. *Enquire*, 1, 1–18. Retrieved from <http://www.nottingham.ac.uk/Sociology/Prospective/Postgraduate/ENQUIRE/index.aspx>.
- Meyer, G.J. (2002). Implications of information gathering methods for a refined taxonomy of psychopathology. In L.E. Beutler & M.L. Malik (Eds.), *Rethinking the DSM: A psychological perspective* (pp. 69–105). Washington, DC: American Psychological Association.
- Miller, J.D., Pilkonis, P.A., & Clifton, A. (2005). Self and other-reports of traits from the five-factor model: Relations to personality disorder. *Journal of Personality Disorders*, 19, 400–419.
- Miller, J.D., Pilkonis, P.A., & Morse, J.Q. (2004). Five-factor model prototypes for personality disorders: The utility of self-reports and observer ratings. *Assessment*, 11, 127–138.
- Mosterman, R.M., Eurelings-Bontekoe, E.H.M., & Hofstee, W.K.B. (2008). Voorspellers van behandelresultaat. De samenhang tussen diagnostiek, indicatie, problemen en behandelresultaat [Predictors of treatment effect. The association between assessment, indication, problems, and treatment effect]. *Tijdschrift Klinische Psychologie*, 2, 98–118.
- Oltmanns, T.F., & Turkheimer, E. (2006). Perception of self and others regarding pathological personality traits. In R.F. Krueger & J. Tackett (Eds.), *Personality and psychopathology* (pp. 71–111). New York: Guilford.
- Paulhus, D.L., & Vazire, S. (2007). The self-report method. In R.W. Robins, R.C. Fraley, & R.F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). London: Guilford.
- Ready, R.E., & Clark, L.A. (2002). Correspondence of psychiatric patient and informant ratings of personality traits, temperament, and interpersonal problems. *Psychological Assessment*, 14, 39–49.
- Ready, R.E., Watson, D., & Clark, L.A. (2002). Psychiatric patient- and informant-reported personality: Predicting concurrent and future behaviour. *Assessment*, 9, 361–372.
- Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, 40, 472–481.
- Zegers, F.E. (1986). A family of chance corrected association coefficients for metric scales. *Psychometrika*, 51, 559–562.
- Zegers, F.E., & Ten Berge, J.M.F. (1985). A family of association coefficients for metric scales. *Psychometrika*, 50, 17–24.