

Self-informant Agreement for Personality and Evaluative Person Descriptors: Comparing Methods for Creating Informant Measures

LEONARD J. SIMMS^{1*}, KERRY ZELAZNY¹,
WERN HOW YAM¹ and DANIEL F. GROS²

¹University at Buffalo, The State University of New York, NY, USA

²Ralph H. Johnson Veterans Administration Medical Center and Medical University of South Carolina, USA

Abstract

Little attention typically is paid to the way self-report measures are translated for use in self-informant agreement studies. We studied two possible methods for creating informant measures: (a) the traditional method in which self-report items were translated from the first- to the third-person and (b) an alternative meta-perceptual method in which informants were directed to rate their perception of the targets' self-perception. We hypothesized that the latter method would yield stronger self-informant agreement for evaluative personality dimensions measured by indirect item markers. We studied these methods in a sample of 303 undergraduate friendship dyads. Results revealed mean-level differences between methods, similar self-informant agreement across methods, stronger agreement for Big Five dimensions than for evaluative dimensions, and incremental validity for meta-perceptual informant rating methods. Limited power reduced the interpretability of several sparse acquaintanceship effects. We conclude that traditional informant methods are appropriate for most personality traits, but meta-perceptual methods may be more appropriate when personality questionnaire items reflect indirect indicators of the trait being measured, which is particularly likely for evaluative traits. Copyright © 2010 John Wiley & Sons, Ltd.

Key words: self-informant agreement; Big Five; Big Seven; evaluative traits; informant measures

INTRODUCTION

Assessing the nature and magnitude of consensual agreement between self- and informant ratings of personality traits represents an important area of inquiry in the personality

*Correspondence to: Leonard J. Simms, Department of Psychology, Park Hall 218, University at Buffalo, State University of New York, Buffalo, NY 14260, USA. E-mail: ljsimms@buffalo.edu

Received 16 November 2009

Revised 25 January 2010

Accepted 25 January 2010

literature. In particular, self-informant agreement (also known as 'self-other' or 'self-peer' agreement) often has been used to support the construct validity of personality trait models and measures (e.g. Funder & Colvin, 1988; Funder, Kolar, & Blackman, 1995; John & Robins, 1993; Ready, Clark, Watson, & Westerhouse, 2000; Watson & Clark, 1991; Watson, Hubbard, & Wiese, 2000). Moreover, in the context of the person-situation debate, evidence of self-informant agreement played an instrumental role in re-establishing the validity of personality traits more generally (e.g. Kenrick & Funder, 1988). Finally, applied studies of personality and personality disorder have highlighted the incremental importance of collecting informant ratings, especially when assessing traits for which the target may have little insight (Clifton, Turkheimer, & Oltmanns, 2005; Klein, 2003; Oltmanns & Turkheimer, 2006).

Despite the importance of self-informant agreement studies in the applied and basic personality literature, surprisingly little attention has been paid to how informant versions of self-rated personality questionnaires are translated from their self-report counterparts. A review of the literature revealed that many self-informant studies fail to specify the manner in which the informant measure was created or adapted (e.g. DeYoung, 2006; Ready, Watson, & Clark, 2002; Spain, Eaton, & Funder, 2000; Vazire, 2006; Wagerman & Funder, 2007). In cases in which the conversion method is specified, researchers typically report using or creating informant versions by simply rephrasing the pronouns and verbs from the first- to the third-person (e.g. 'I am a talkative person' becomes 'He/she is a talkative person'; Duckworth & Quinn, 2009; Gros, Simms, & Antony, in press; Kurtz & Putnam, 2006; Rammstedt, Riemann, Angleitner, & Borkenau, 2004; Ready & Clark, 2005).

The allure of such methods is their apparent simplicity: It is a relatively straightforward process to translate self-report items to informant-report items by substituting third-person pronouns and verbs. However, although traditional third-person translations are simple, they make the assumption that an item's meaning is unaltered in the process. Such an assumption may be reasonable for certain self-report item types, such as those that *directly* assess an observable characteristic of the target. Examples include: (a) 'I am a sociable person,' (b) 'I have moments when I worry too much,' or (c) 'I sometimes have difficulty controlling my behaviour.' When translated, the traditional informant versions—(a') 'She is a sociable person,' (b') 'He has moments when he worries too much,' or (c') 'She sometimes has difficulty controlling her behaviour,' respectively—can be observed by informants and appear to accurately tap the same personality traits as their self-report counterparts.

However, for items that are meant to *indirectly* tap a particular trait or that are more evaluative in nature, such direct translations may not yield adequate informant items. Consider, for example, a different set of self-report items that were designed to tap a range of evaluative personality characteristics (Simms, Yufik, Thomas, & Simms, 2008): (x) 'I have a sharp mind,' (y) 'My life lacks meaning and purpose' and (z) 'I am a beautiful person.' In this case, traditional informant translations—(x') 'He has a sharp mind,' (y') 'Her life lacks meaning and purpose' and (z') 'She is a beautiful person,' respectively—do not appear to tap the self-perceptual processes that were implicit in the self-report items. Rather, translations focusing on the informant's perception of the target's self-perception—(x'') 'He thinks that he has a sharp mind,' (y'') 'She believes that her life lacks meaning and purpose' and (z'') 'She sees herself as a beautiful person,' respectively—may more directly tap the personality-relevant variance intended by the self-report items and their parent constructs.

A robust literature has established that self-informant agreement for traditional personality markers using traditional informant wording (TIW) ranges from .40 to .60 depending on factors such as trait visibility, acquaintanceship length and type, and trait ratability (e.g. Ready et al., 2000; Watson et al., 2000). However, much less is known about the nature of self-informant agreement for dimensions with a larger evaluative component. In one of the few studies addressing this issue, John and Robins (1993) showed that the more evaluative markers of the Big Five demonstrated lower self-informant agreement compared to evaluatively neutral markers, and they suggested that self-perceptions might become distorted for personality traits that are affectively charged.

Recent lexical studies have extended our understanding of the influence of evaluation in personality trait structure. In particular, a number of studies have suggested that the use of a less restrictive set of person descriptors—including evaluative terms such as those sorted into a separate, non-personality category in the work of Allport and Odbert (1936) and Norman (1967)—results in a broader model of higher-order personality trait dimensions. Tellegen and Waller (1987; Waller, 1999; Waller & Zavala, 1993) were the first to suggest the potential importance of evaluative person descriptors to the study of personality. They sampled 400 personality descriptors from the dictionary without imposing the exclusionary criteria that characterized previous natural language studies of personality. Factor analyses of self-ratings on the sampled terms revealed a Big Seven structure, five of which were relatively isomorphic with the Big Five. The remaining two factors—labelled *Positive Valence* (PV) and *Negative Valence* (NV) by Tellegen and Waller (1987)—represented highly valenced dimensions of extremely positive (e.g. describing oneself as exceptional, important, smart) and negative (e.g. describing oneself as evil, immoral, disgusting) person descriptors, respectively. Notably, significant debate exists in the literature regarding the meaning and status of PV and NV as personality constructs (e.g. Ashton & Lee, 2001; McCrae & Costa, 1995; Widiger, 1993; Widiger & Trull, 1992). For example, Ashton and Lee (2001) contend that the PV and NV factors largely represent indirect *signs* of self-esteem processes, rather than direct *descriptors* of personality characteristics as found on most traditional personality measures. Regardless of the outcome of this unresolved debate, Tellegen and Waller's findings have been supported in new samples, languages and cultures (e.g. Almagor, Tellegen, & Waller, 1995; Benet & Waller, 1995; Benet-Martinez & Waller, 1997; Benet-Martinez & Waller, 1998; Church, Katigbak, & Reyes, 1998; Saucier, 1997; Waller 1999).

Simms et al. (2008) studied the nature of PV and NV and compared the self-informant agreement of such evaluative dimensions to those of the Big Five, finding mixed results. Whereas the agreement coefficients for negatively valenced dimensions (Depravity and Oddity) were within the same range as those typically observed for the Big Five domains, PV and its facets (distinction, intellect, attractiveness, and self-worth) demonstrated substantially lower levels of self-informant agreement. Such mixed results suggest several possible conclusions. One possibility is that PV characteristics are not personality traits at all, but rather reflect individual differences in other human attributes, such as intelligence, success and appearance. However, Simms et al. found significant correlations between PV and traditional personality traits that were much higher than would be expected if the PV scales were tapping constructs completely independent of personality. Thus, a second possibility, relevant to the current study, is that the PV scale and its facets tap self-perceptual processes that are not directly ratable by traditional informant translations of self-report measures.

Current study

In summary, little attention typically is paid to the way self-report measures are translated for use in self-informant agreement studies. Traditional translations from the first- to the third-person make intuitive sense for personality questionnaire items that tap directly observable characteristics but may be less informative for indirect or evaluative items whose personality-relevant meaning is embedded in the self-perceptual processes of the individual being rated. In the present study, we propose an alternative translation method—one focused on the informant's perception of the target's self-perception—that may be more appropriate for personality constructs and items that are best tapped by indirectly worded self-report items. We term this alternative method *meta-perceptual informant wording* (MPIW) and contrast it with TIW as described above. Notably, for some multi-scale personality measures, the MPIW approach has been used to create informant versions of certain items and scales, such as certain items of the Modesty scale of the NEO-PI-R (see McCrae & Costa, 1995) and the Modesty and Social Self-Esteem scales of the HEXACO-PI-R (see Ashton & Lee, 2010). However, these uses of the MPIW approach were very specific in their application. No studies have systematically examined the MPIW approach across a broad range of scales varying in content and evaluativeness.

In the present study, we assessed self-informant agreement in undergraduate friendship dyads, each member of which completed self-report measures tapping the Big Five traits and the evaluative dimensions of the Big Seven model, as well as two informant versions of each: One translated using TIW methods and one translated using the alternative MPIW methods. We focused on several primary questions. First, do alternative MPIW methods that focus on the informant's estimate of the target's self-view result in an increase in self-other agreement for evaluative dimensions compared to traditional personality traits? Second, are any of these self-informant effects moderated by acquaintanceship length? For traditional personality dimensions, evidence for acquaintanceship length effects are mixed (e.g. Watson et al., 2000). However, little is known about the acquaintanceship effect for evaluative dimensions. In the present study, we examined whether self-informant agreement varies as a function of friendship duration (1 year or less vs. more than 1 year). Finally, are alternative MPIW methods redundant with TIW, or do they yield incremental predictive power with respect to target self-ratings?

METHOD

Participants and procedures

Participants were 606 undergraduate students (303 dyads) at the University at Buffalo. One person in each dyad was a student in Introductory Psychology and was compensated with course research credit. The students were asked to come to the laboratory with a friend or close acquaintance. Friends were compensated in one of two ways: (a) course research credit for those also admitted in the same course or (b) entry into a drawing for one of several \$25 gift certificates to a local mall. Participants were 51% female with a *M* age of 19.2 (range = 18–36). The sample self-reported their ethnic background as 59.1% White, 24.9% Asian American, 6.6% African American, 3.8% Hispanic/Latino(a) and 5.6% 'other/multiple ethnicities'. Dyads were comprised of both same-sex and opposite-sex friendship pairings: 37.0% were female–female dyads, 35.0% were male–male dyads and 28.0% were male–female dyads. The *Mdn* friendship duration across dyads was 12 months

(range = <6 months to >48 months). Dyad members predominantly described their relationship as a friendship (90.1% of dyads); a smaller number described their relationship as romantic (5.8%) and the remaining dyads listed 'other'.

Participants attended a 1-hour session and were asked to complete a series of questionnaires using a computer interface in groups of 10 or less. Questionnaires included the Big Five Inventory (BFI; John & Srivastava, 1999) to measure the Big Five personality traits, and the Evaluative Person Descriptors Questionnaire (EPDQ; Simms et al., 2008) to measure evaluative traits inspired by the Big Seven model of personality (Tellegen & Waller, 1987). To meet the goals of this study, both members of each dyad completed the BFI and EPDQ in three different ways: (a) self-report with standard instructions, (b) TIW in which the self-report version was converted to an informant-report version by making simple pronoun and verb changes from the first- to the third-person (e.g. 'I am a sociable person' became 'X is a sociable person') and (c) MPIW in which the informant-report version was created to tap the informant's perceptions of the target's self-view (e.g. 'I am a sociable person' became 'X sees himself/herself as a sociable person').

The following instructions were provided to participants: 'There are three parts to this study. In each part, you will be completing a series of questionnaires on the computer. In the first part, you will be rating your own personality. In the second part, you will be rating the personality of the friend who accompanied you today. Finally, in the third part, I would like you to rate what you think your friend would say about him-/herself for each item. However, I do not want you to talk with your friend about these questions as you are going through the study'. To ensure privacy within and across dyads, barriers were erected to shield computer screens and participants were not permitted to speak to one another during data collection sessions. The Social and Behavioral Sciences Institutional Review Board at the University at Buffalo approved all procedures.

Measures

Big Five Inventory (BFI; John & Srivastava, 1999)

The BFI is a 44-item instrument that uses a 5-point Likert rating scale (1 = *strongly disagree*; 5 = *strongly agree*) and provides scores for the domains of the Big Five model of personality (Neuroticism, Extraversion, Conscientiousness, Agreeableness and Openness). Benet-Martinez and John (1998) reported α coefficients of .84, .88, .82, .79 and .81, respectively, for the traits listed above in a sample of 711 English-speaking participants. They also reported good convergence with two other measures of the Big Five model. John and Srivastava (1999) further summarized that internal consistency reliabilities of BFI scales typically range from .75 to .90 in North American samples and 3-month test-retest reliabilities typically range from .80 to .90.

EPDQ

The EPDQ (Simms et al., 2008) is a 48-item measure designed to measure and elaborate the evaluative dimensions of the Big Seven model. It uses a 5-point Likert scale (1 = *disagree strongly*; 5 = *agree strongly*) and includes broad scales tapping PV, Depravity (akin to the NV dimension of the Big Seven model) and Oddity, as well as several lower-order facets of PV: Distinction, Intellect, Attractiveness, and Self-worth. Simms et al. reported evidence of internal consistency (α s ranged from .80 to .92 across scales), test-retest reliability over a 2-week interval (r s ranged from .73 to .82) and good convergent and discriminant validity with respect to alternative measures of PV and NV.

Data structure and analyses

Given the dyadic nature of the study, the indistinguishable nature of dyad members, and the goals of the present investigation, data were organized using a *pairwise* or *double-entry* structure in which all participants were represented twice in the data set, once as a target and once as an informant (Kenny, Kashy, & Cook, 2006). All analyses were conducted with the target/individual as the unit of analysis, so the effective sample size for all analyses was 606 participants. Pearson correlations and hierarchical regressions were the primary analytic methods. Correlations between dyad members' self-ratings of the BFI and EPDQ variables generally were small in magnitude (Median $r = .12$; range = $-.02-.23$), and most were statistically non-significant ($ps > .05$, based on an effective sample size of 303 dyads). Thus, statistical dependency among dyad members on the variables of interest in this study was minimal or non-existent, which permitted us to analyse the data using standard Pearson correlations and regression methods. Where appropriate, statistical tests were Bonferroni-corrected to Protect against Type I errors.

RESULTS

Descriptive statistics for the primary variables appear in Table 1. Effect size calculations revealed that 6 of 7 EPDQ scales (Positive Valence Composite, Intellect, Attractiveness, Self-worth, Depravity and Oddity) and 3 of 5 BFI scales (Neuroticism, Extraversion and Openness) demonstrated small but statistically significant differences between self-ratings and TIW-ratings, all $ts(605) \geq 3.94$, $ps < .01$. In all but one case, the significant differences showed TIW ratings that were more positive or less negative than self-ratings, suggesting that informants using traditionally translated measures tend to be more positive with their attributions of their friends' traits than the targets themselves. The lone exception to this pattern was Openness, for which self-ratings were higher than TIW-ratings. In contrast, the self-MPIW effect sizes revealed a very different pattern, especially for the EPDQ trait dimensions. Only one EPDQ scale (Oddity) and two BFI scales (Neuroticism and Openness) yielded significant differences between self-ratings and alternative MPIW-ratings, all $ts(605) \geq 3.94$, $ps < .01$, and the directions of these effects were the same as those for the corresponding self-TIW differences. Taken together, these results suggest that self-ratings and informant-ratings are more similar at the mean-level when informant ratings are directed to the target's self-view, especially for the relatively evaluative dimensions of the EPDQ.

Question #1: Is self-informant agreement moderated by informant wording type

Self-informant agreement correlations for TIW and MPIW methods appear in Table 2. Self-TIW correlations averaged .29 and .40 for the EPDQ and BFI scales, respectively. Self-MPIW correlations averaged .33 and .40 for the EPDQ and BFI scales, respectively. Only two scales—EPDQ Intellect and Attractiveness—demonstrated significant differences across informant wording methods, with MPIW showing a stronger self-informant correlation than TIW in both cases, $zs \geq 1.96$, $p < .05$. Thus, for most EPDQ scales and all BFI scales, self-informant agreement does not appear to vary as a function of informant wording type. Also, consistent with previously reported data (Simms et al., 2008), the BFI scales generally yielded stronger self-informant correlations than did the

Table 1. Descriptive statistics for self- and informant-ratings

Variable	Self <i>M (SD)</i>	TIW <i>M (SD)</i>	MPIW <i>M (SD)</i>	Effect sizes		
				Self-TIW	Self-MPIW	TIW-MPIW
EPDQ scales						
Positive valence	3.77 (0.57)	3.96 (0.54)	3.76 (0.59)	−0.35**	0.02	0.36**
Distinction	3.58 (0.70)	3.61 (0.66)	3.52 (0.67)	−0.04	0.09	0.14**
Intellect	3.92 (0.68)	4.12 (0.61)	3.86 (0.68)	−0.31**	0.09	0.41**
Attractiveness	3.51 (0.79)	3.80 (0.77)	3.60 (0.82)	−0.37**	−0.12	0.25**
Self-worth	4.05 (0.75)	4.31 (0.67)	4.04 (0.82)	−0.36**	0.01	0.36**
Depravity	2.03 (0.72)	1.86 (0.75)	1.95 (0.79)	0.23**	0.11	−0.11**
Oddity	2.92 (0.89)	2.65 (0.87)	2.63 (0.88)	0.31**	0.33**	0.02
BFI scales						
Neuroticism	23.45 (5.52)	21.00 (5.78)	20.77 (5.35)	0.43**	0.49**	0.04
Extraversion	28.17 (5.58)	29.21 (6.27)	28.66 (5.90)	−0.18**	−0.08	0.09**
Conscientiousness	30.61 (5.21)	31.23 (5.78)	31.11 (5.09)	−0.11	−0.10	0.02
Agreeableness	34.10 (5.11)	34.36 (5.98)	34.64 (5.16)	−0.05	−0.11	−0.05
Openness	35.86 (5.50)	34.37 (5.35)	34.59 (5.27)	0.27**	0.23**	−0.04

Note: *N* = 606 (303 dyads). TIW, traditional informant wording; MPIW, meta-perception informant wording; EPDQ, Evaluative Person Descriptors Questionnaire; BFI, Big Five Inventory.

**Significant difference between paired means (all $t_s(605) > 3.94$; $p < .01$, Bonferroni-corrected).

Table 2. Self-informant agreement by informant wording type

Scale	Self-informant correlation		
	TIW	MPIW	$r_{\text{TIW:MPIW}}$
EPDQ scales			
Positive valence composite	.34	.39	.61
Distinction	.27	.26	.63
Intellect	.25	.32*	.57
Attractiveness	.34	.41*	.54
Self-worth	.29	.29	.64
Depravity	.27	.31	.78
Oddity	.28	.32	.80
<i>M</i>	.29	.33	.65
BFI scales			
Neuroticism	.36	.34	.76
Extraversion	.62	.59	.85
Conscientiousness	.43	.41	.73
Agreeableness	.36	.36	.76
Openness	.34	.38	.74
<i>M</i>	.40	.40	.75

Note: $N = 606$ (303 dyads). TIW, traditional informant wording; MPIW, meta-perception of the target wording; EPDQ, Evaluative Person Descriptors Questionnaire; BFI, Big Five Inventory. All $rs \geq .15$ are significant ($p < .01$).

*Higher self-informant agreement ($p < .05$).

EPDQ scales. Finally, it is notable that the within-informant correlations between TIW and MPIW varied across scales: PV and its subscales (rs ranged from .54 to .64) yielded generally lower TIW–MPIW correlations than did Depravity, Oddity, and all BFI scales (rs ranged from .74 to .85). Thus, informant-rating methods appear to be more weakly related for dimensions with stronger evaluative components.

Question #2: Is self-informant agreement moderated by acquaintanceship length?

To test the possible influence of acquaintanceship length on self-informant agreement across informant wording methods, we split the sample at the median of friendship duration, yielding two groups: (a) friendships lasting 1 year or less and (b) longer friendships lasting more than 1 year. A summary of self-informant agreement correlations, separately by acquaintanceship length and informant wording type, appears in Table 3. These results revealed a sparse pattern of significant differences, but several notable patterns. First, averaging across all EPDQ scales, self-informant agreement did not differ appreciably across acquaintanceship levels or informant wording types (*Mean* rs ranged from .28 to .32). However, two scales—Intellect and the PV composite scale—yielded significantly higher self-informant correlations for MPIW ratings than for TIW ratings, but only for shorter friendships, $zs \geq 1.96$, $p < .05$.

Second, for the BFI scales, longer friendships generally were associated with higher agreement coefficients across the TIW and MPIW informant wording types: *Mean* rs were .46 and .45 for the longer acquaintanceship groups and .39 and .38 for the shorter acquaintanceship group. However, at the scale-level, only two BFI scales demonstrated significant acquaintanceship effects, but these effects were not consistent across informant wording methods. BFI Neuroticism showed stronger agreement as a function of friendship

Table 3. Self-informant agreement by friendship duration and informant wording type

	Friendship duration				Duration effect		Wording effect	
	≤1 Y		>1 Y					
	TIW	MPIW	TIW	MPIW	TIW	MPIW	≤1 Y	>1 Y
EPDQ scales								
Positive valence composite	.34	.43	.34	.34			*	
Distinction	.24	.27	.30	.25				
Intellect	.23	.34	.25	.28			*	
Attractiveness	.34	.40	.34	.43				
Self-worth	.26	.31	.32	.26				
Depravity	.27	.31	.27	.31				
Oddity	.25	.27	.32	.37				
<i>M</i>	.28	.33	.31	.32				
BFI scales								
Neuroticism	.27	.28	.46	.41	**			
Extraversion	.62	.61	.61	.56				
Conscientiousness	.43	.41	.45	.40				
Agreeableness	.32	.27	.42	.46		**		
Openness	.29	.35	.38	.41				
<i>M</i>	.39	.38	.46	.45				

Note: *N*s = 320 and 286 for <1 Y (less than or equal to 1 year friendship duration) and ≥1 Y (greater than 1 year friendship duration) groups, respectively. TIW, traditional informant wording; MPIW, meta-perception of the target wording; EPDQ, Evaluative Person Descriptors Questionnaire; BFI, Big Five Inventory. All *r*s ≥ .15 are significant ($p < .01$).

* $p < .05$; ** $p < .01$. Higher correlations within a comparison are presented in boldface.

duration, but only for the TIW ratings: *Mean r*s = .46 and .27 for longer and shorter acquaintanceship groups, respectively, z s ≥ 2.58, $p < .01$. BFI Agreeableness, on the other hand, showed stronger agreement as a function of acquaintanceship length, but only for the MPIW ratings: *Mean r*s = .46 and .27 for longer and shorter acquaintanceship groups, respectively, z s ≥ 2.58, $p < .01$.

Question #3: Does MPIW show incremental validity of over TIW methods?

Finally, we tested whether MPIW ratings were incrementally valid predictors of self-ratings after accounting for traditional informant rating methods. A series of hierarchical regressions were conducted for each scale, with target self-ratings entered as the dependent variable and the TIW and MPIW informant rating methods entered sequentially as the predictors. A summary of these results appears in Table 4. In the first step of each analysis, the TIW ratings significantly predicted target self-ratings, accounting for an average of 8.7 and 18.7% of the variance in target ratings for the EPDQ and BFI scales, respectively, all F s(1, 604) > 39.03, $p < .01$. When MPIW ratings were added in the next regression step, all incremental effects were significant, accounting for an additional average of 3.7 and 2.1% of the variance in target ratings for the EPDQ and BFI scales, respectively, all F s(1, 603) ranged from 8.37 to 55.66, $p < .05$. Taken together, the regressions show that TIW ratings yielded stronger connections with self-ratings for the BFI scales, and MPIW ratings resulted in stronger incremental effects for the evaluative scales of the EPDQ.

Table 4. Hierarchical regressions predicting self-ratings by TW and MPW informant wordings

	Step 1: TIW		Step 2: MPIW		
	<i>R</i>	<i>R</i> ²	<i>R</i>	<i>R</i> ²	ΔR^2
EPDQ scale					
Positive valence composite	.344	.118**	.413	.170	.052**
Distinction	.273	.075**	.296	.088	.013*
Intellect	.246	.061**	.327	.107	.047**
Attractiveness	.341	.116**	.437	.191	.075**
Self-worth	.286	.082**	.320	.102	.020**
Depravity	.271	.073**	.315	.099	.026**
Oddity	.284	.081**	.322	.104	.023**
<i>M</i>		.087		.123	.037
BFI scale					
Neuroticism	.359	.129**	.375	.141	.012*
Extraversion	.617	.380**	.631	.398	.017**
Conscientiousness	.430	.185**	.452	.204	.019**
Agreeableness	.362	.131**	.383	.146	.016*
Openness	.335	.113**	.388	.151	.038**
<i>M</i>		.187		.208	.021

Note: *N* = 606 (303 dyads). TIW, traditional informant wording; MPIW, meta-perception of the target wording; EPDQ, Evaluative Person Descriptors Questionnaire; BFI, Big Five Inventory.

p* < .05; *p* < .01 (Bonferroni-corrected).

DISCUSSION

The primary goal of the present study was to compare two methods for translating traditional self-report personality measures into forms that can be rated by informants. We compared traditionally worded informant translations (the TIW method)—in which the verb and pronouns of each item were simply translated from the first- to the third-person—to an alternative method in which the translated item was focused on the informant's perception of the target's self-perception (the meta-perceptual wording [MPIW] approach). We hypothesized that the MPIW method would be a more appropriate method for constructs and items that are more evaluative in nature and indirectly assessed. The results suggested that (a) mean self and MPIW ratings generally were more similar than were mean self and TIW ratings, (b) self-informant agreement was largely similar across informant methods, (c) Big Five personality traits generally demonstrated stronger self-informant agreement, using both informant methods, than did evaluative personality dimensions and (d) MPIW ratings incrementally predicted the targets' self-ratings, especially for traits with stronger evaluative components. We also studied whether any effects were moderated by friendship duration: Although the BFI agreement correlations appeared to demonstrate an acquaintanceship effect when averaged across scales, the result at the individual scale-level revealed a sparse and inconsistent pattern, likely due to the reduced statistical power associated with splitting the sample at the median of friendship duration.

Mean-level differences across informant rating methods

The mean-level agreement analyses revealed an interesting but unexpected pattern of results, in several respects. First, contrary to previous work showing that self-ratings tend to

display a positive or self-enhancement bias relative to the ratings of others (e.g. Colvin, Block, & Funder, 1995; Funder & Colvin, 1997; Taylor & Brown, 1988), we identified an opposite pattern in which the traditionally worded informant ratings tended to be more positive and less negative than were self-ratings, especially for relatively evaluative dimensions. The reasons for this finding are unclear. Surprisingly little research has looked at mean-level differences between self and informant ratings of personality; most work has been strictly correlational. The present findings could be due to the effects of selection biases, as participants were permitted to bring a friend of their own choosing to the study session. Given the nature of the task and the minimal compensation offered, it is likely that only acquaintances who were relatively fond of the invitee would agree to participate in the study. If true, then the present findings suggest that, in the context of a friendship, informant ratings may be subject to an *other-enhancement bias*, which would be a relatively novel finding in the personality literature requiring replication in future studies. Interestingly, Gros et al. (in press) recently found similar evidence of an other-enhancement bias in a sample of friendships dyads rating cognitive and somatic symptoms of anxiety. In the present study, the other-enhancement bias was most robust for the evaluative traits, which suggests that the level of trait evaluativeness may serve as a moderator of these mean-level agreement differences.

A second interesting mean-level pattern was that self-informant differences largely disappeared, especially for the evaluative dimensions, when meta-perceptual wording was used for the informant measures. This finding suggests that any possible self- or other-enhancement biases in reporting are ameliorated when the targets and informants both are focused on targets' self-perceptions. Of course, given the absence of a clear gold standard, the exact nature and direction of any possible self- or other-enhancement biases is difficult to determine with data like these. One possibility, for example, is that both members of friendship dyads are biased; in such a case, the mean-level differences approach is powerless to determine the 'true' level of the targets' traits or whether any biases exist in self- or informant-ratings. Future self-informant studies of TIW and MPIW may be improved to the extent that relevant criteria external to the dyad members are included to help objectively determine whose report, self or informant, better predicts such criteria (e.g. laboratory tasks or behavioural ratings of personality traits, etc.).

Incremental validity of meta-perceptual informant wording

Self-informant agreement was largely similar across informant rating methods—with the exception of several PV traits that showed stronger agreement with the MPIW method—but the hierarchical regression analyses suggested that the methods account for overlapping, but not identical, portions of the variance in self-ratings. Across all traits, meta-perceptual informant ratings significantly predicted self-ratings, even after accounting for traditional informant ratings. These results suggest that meta-perceptual wording taps something distinctive about the target that is not captured with traditional informant methods. The strongest incremental effects were found for the EPDQ dimensions, which suggest that prediction of traits with stronger evaluative components can be improved if the informants' focus is placed on targets' self-perceptions rather than on direct estimates of the targets' behaviour or features. Of course, an alternative explanation of our findings is that any second set of slightly differently worded informant ratings would yield some incremental effects beyond a first set. However, our finding that the size of the incremental effects varied substantially across scales would tend to argue against this interpretation. Regardless, future studies examining this possibility more directly are important.

Thus, although future studies are needed to replicate the present findings, we suggest that more consideration is needed when translating measures for use in self-informant personality agreement studies. Care must be taken to determine the most appropriate target of the translated item. Traditional translations appear to make the most sense for personality questionnaire items that target traits directly, such as (a) 'I am a sociable person' for Extraversion, (b) 'I have moments when I worry too much' for Neuroticism or (c) 'I sometimes have difficulty controlling my behaviour' for Conscientiousness or Constraint. However, the present results suggest that evaluative personality items whose trait inferences are more indirect—such as (x) 'I have a sharp mind' for PV Intellect, (y) 'My life lacks meaning and purpose' for PV Self-worth and (z) 'I am a beautiful person' for PV Attractiveness—may be best translated using MPIW.

Implications for the status of evaluative traits

Consistent with previous work, the present findings revealed that evaluative traits show significant but generally lower self-informant agreement than Big Five personality traits (Simms et al., 2008). The present findings extend this result to a new informant rating method focused on the targets' self-perceptions. As described above, a number of researchers and theorists have questioned the status of evaluative dimensions such as PV and NV as personality constructs (e.g. Ashton & Lee, 2001; McCrae & Costa, 1995; Widiger, 1993; Widiger & Trull, 1992). Although definitions vary widely, most conceptualizations of personality highlight the stable, enduring aspects of personality (e.g. Roberts & DelVecchio, 2000; Vaidya, Gray, Haig, & Watson, 2002), the importance of predicting behaviour (e.g. Cronbach & Meehl, 1955; Wu & Clark, 2003), and its ability to be consensually agreed upon across raters (John & Robins, 1993; Ready et al., 2000; Watson & Clark, 1991; Watson et al., 2000). The present study provides evidence that is most relevant to the last feature: Evaluative dimensions can be consensually rated in friendship dyads, regardless of the informant rating method employed.

But what do the present results say about *what* is being consensually agreed upon by members of these friendship dyads? As described above, Ashton and Lee (2001) suggested that evaluative dimensions such as those measured by the EPDQ reflect indirect signs of positive and negative self-esteem processes reflecting some combination of substantive and evaluative variance. This is an intriguing proposition, especially since the meta-perceptual informant rating methods are likely to more directly tap the substantive self-esteem processes implied by Ashton and Lee's suggestion. However, it is likely that all personality dimensions can be subdivided into substantive and evaluative components, and our data do not permit us to directly disentangle and directly study the substantive and evaluative components of the BFI and EPDQ scales (for one promising approach to disentangling these components factor analytically, see Ashton & Lee, 2010). Regardless, self-esteem personality processes are a reasonable hypothesis for what evaluative personality scales measure. Future studies are needed to study this issue further, especially in light of other recent work linking PV and NV to other substantive personality traits, such as extraversion and agreeableness (Durrett & Trull, 2005; Simms, 2007; Simms et al., 2008).

Limitations and Conclusions

This study is not without limitations that may serve as the impetus for future studies to clarify or extend our results. Several features of the friendship dyad participants limit the

generalizability of our results. Although the participants were reasonably diverse in terms of sex and ethnicity, they were an undergraduate sample of convenience and included a restricted range of ages. Also, we asked participants to self-select a single friend to bring to the session, which could lead to unintended informant rating biases with the potential to influence studies like this. Finally, our sample included a small number of dyads that were romantic or familial in nature, but these subsamples were not large enough to permit moderation analyses by dyad type. Thus, future studies may wish to extend the present work by (a) collecting similar responses from friendship dyads of different ages and in settings other than universities, (b) using multiple informants or, perhaps, single informants randomly selected from lists provided by the targets and (c) studying whether the findings extend to different types of relationships, such as dating partners, married couples, family members and coworkers.

Despite these limitations, our work suggests that personality researchers should pay closer attention to the way that self-report questionnaires are translated for use in self-informant agreement studies or in applied settings in which informant reports are used to guide decision-making about clients, patients, coworkers, etc. The findings suggest that traditional informant methods are appropriate for most personality traits in which the questionnaire items permit direct judgments of the trait being measured. However, alternative methods based on the informants' perception of the targets' self-perception should be considered when questionnaire items reflect indirect indicators of the trait being measured, which is particularly likely for strongly evaluative traits.

ACKNOWLEDGEMENTS

We appreciate the contributions of the research assistants who helped with data collection and the participants who provided the data included in this study. Some of the data contained herein were presented at the annual meeting of the Society for Personality and Social Psychology, January 2007. The preparation of this paper was supported by a research grant to the first author: National Institute of Mental Health #1R01MH080086.

REFERENCES

- Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs*, 47(211) p. 171.
- Almagor, M., Tellegen, A., & Waller, N. G. (1995). The big seven model: A cross-cultural replication and further exploration of the basic dimensions of natural language trait descriptors. *Journal of Personality and Social Psychology*, 69, 300–307.
- Ashton, M. C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality*, 15, 327–353.
- Ashton, M. C., & Lee, K. (2010). Trait and source factors in HEXACO-PI-R self- and observer reports. *European Journal of Personality*, 24, 278–289. DOI: 10.1002/per.759
- Benet-Martinez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75, 729–750.
- Benet-Martinez, V., & Waller, N. G. (1997). Further evidence for the cross-cultural generality of the Big Seven Factor model: Indigenous and imported Spanish personality constructs. *Journal of Personality*, 65, 567–598.

- Benet-Martinez, V., & Waller, N. G. (2002). From adorable to worthless: Implicit and self-report structure of highly evaluative personality descriptors. *European Journal of Personality*, 16, 1–41.
- Benet, V., & Waller, N. G. (1995). The 'Big Seven' model of personality description: Evidence for its cross-cultural generality in a Spanish sample. *Journal of Personality and Social Psychology*, 69, 701–718.
- Church, A. T., Katigbak, M. S., & Reyes, J. A. S. (1998). Further exploration of Filipino personality structure using the lexical approach: Do the Big-Five or Big-Seven dimensions emerge? *European Journal of Personality*, 12, 249–269.
- Clifton, A., Turkheimer, E., & Oltmanns, T. F. (2005). Self- and peer perspectives on pathological personality traits and interpersonal problems. *Psychological Assessment*, 17, 123–131.
- Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology*, 68, 1152–1162.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, 91, 1138–1151.
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT-S). *Journal of Personality Assessment*, 91, 166–174.
- Durrett, C., & Trull, T. J. (2005). An evaluation of evaluative personality terms: A comparison of the Big Seven and Five-Factor Model in predicting psychopathology. *Psychological Assessment*, 17, 359–368.
- 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 selfjudgments 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., Kolar, D. C., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology*, 69, 656–672.
- Gros, D. F., Simms, L. J., & Antony, M. M. (in press) The psychometric properties of the State-Trait Inventory for Cognitive and Somatic Anxiety (STICSA) in friendship dyads. *Behavior Therapy*.
- 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, 521–531.
- John, O. P., & Srivastava, S. (1999). The Big Five taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin, & O. P. John (Eds.), *Handbook of Personality* (2nd ed., pp. 102–138). New York: Guilford Press.
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford Press.
- Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, 43, 23–34.
- 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.
- Kurtz, J. E., & Putnam, S. H. (2006). Patient-informant agreement on personality ratings and self-awareness after head injury. *The Clinical Neuropsychologist*, 20, 453–468.
- McCrae, R. R., & Costa, P. T. Jr., (1995). Positive and negative valence within the five-factor model. *Journal of Research in Personality*, 29, 443–460.
- Norman, W. T. (1967). *2800 personality trait descriptors: Normative operating characteristics for a university population*. University of Michigan, Department of Psychology. Ann Arbor, MI
- Oltmanns, T. F., & Turkheimer, E. (2006). Perceptions of self and others regarding pathological personality traits. In R. F. Krueger, & J. L. Tackett (Eds.), *Personality and psychopathology* (pp. 71–111). New York, NY: Guilford Press.
- Rammstedt, B., Riemann, R., Angleitner, A., & Borkenau, P. (2004). Resilients, Overcontrollers, and Undercontrollers: The replicability of the three personality prototypes across informants. *European Journal of Personality*, 18, 1–14.

- Ready, R. E., & Clark, L. A. (2005). Psychiatric patient and informant reports of patient behavior. *Journal of Personality*, 73, 1–21.
- Ready, R. E., Clark, L. A., Watson, D., & Westerhouse, K. (2000). Self- and peer-related personality: Agreement, trait ratability, and the 'self-based heuristic.' *Journal of Research in Personality*, 34, 208–224.
- Ready, R. E., Watson, D., & Clark, L. A. (2002). Psychiatric patient- and informant-reported personality: Predicting concurrent and future behavior. *Assessment*, 9, 361–371.
- 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.
- Saucier, G. (1997). Effect of variable selection on the factor structure of person descriptors. *Journal of Personality and Social Psychology*, 73, 1296–1312.
- Simms, L. J. (2007). The Big Seven model of personality and its relevance to personality pathology. *Journal of Personality*, 75, 65–94.
- Simms, L. J., Yufik, T., Thomas, J. P., & Simms, E. N. (2008). Exploring evaluative person descriptors through scale development. *Journal of Research in Personality*, 42, 1271–1284.
- Spain, J. S., Eaton, L. G., & Funder, D. C. (2000). Perspectives on personality: The relative accuracy of self versus others for the prediction of emotion and behavior. *Journal of Personality*, 68, 837–867.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 193–210.
- Tellegen, A., & Waller, N. G. (1987). Reexamining Basic Dimensions of Natural Language Trait Descriptors. *Paper presented at the 95th Annual Meeting of the American Psychological Association*, New York, NY.
- Vaidya, J. G., Gray, E. K., Haig, J., & Watson, D. (2002). On the temporal stability of personality: Evidence for differential stability and the role of life experiences. *Journal of Personality and Social Psychology*, 83, 1469–1484.
- Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, 40, 472–481.
- Wagerman, S. A., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality*, 41, 221–229.
- Waller, N. G. (1999). Evaluating the structure of personality. In C. R. Cloninger (Ed.), *Personality and psychopathology* (pp. 155–197). Washington, DC: American Psychiatric Press.
- Waller, N. G., & Zavala, J. (1993). Evaluating the Big Five. *Psychological Inquiry*, 4, 131–134.
- Watson, D., & Clark, L. A. (1991). Self- versus peer-ratings of specific emotional traits: Evidence of convergent and discriminant validity. *Journal of Personality and Social Psychology*, 60, 927–940.
- Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, 78, 546–558.
- Widiger, T. A. (1993). The DSM-III-R categorical personality disorder diagnoses: A critique and an alternative. *Psychological Inquiry*, 4, 75–90.
- Widiger, T. A., & Trull, T. (1992). Personality and psychopathology: An application of the five-factor model. *Journal of Personality*, 60, 363–393.
- Wu, K. D., & Clark, L. A. (2003). Relations between personality traits and self-reports of daily behavior. *Journal of Research in Personality*, 37, 231–256.