

A narrower scope or a clearer lens for personality? Examining sources of observers' advantages over self-reports for predicting performance

Citation for published version (APA):

Connelly, B. S., & Hülshager, U. R. (2012). A narrower scope or a clearer lens for personality? Examining sources of observers' advantages over self-reports for predicting performance. *Journal of Personality, 80*(3), 603-631. <https://doi.org/10.1111/j.1467-6494.2011.00744.x>

Document status and date:

Published: 01/06/2012

DOI:

[10.1111/j.1467-6494.2011.00744.x](https://doi.org/10.1111/j.1467-6494.2011.00744.x)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 17 Apr. 2025

A Narrower Scope or a Clearer Lens for Personality? Examining Sources of Observers' Advantages Over Self-Reports for Predicting Performance

Brian S. Connelly¹ and Ute R. Hülshager²

¹University of Toronto

²Maastricht University

ABSTRACT Emerging studies have shown that observers' ratings of personality predict performance behaviors better than do self-ratings. However, it is unclear whether these predictive advantages stem from (a) use of observers who have a frame of reference more closely aligned with the criterion ("narrower scope") or (b) observers having greater accuracy than targets themselves ("clearer lens"). In a primary study of 291 raters of 97 targets, we found predictive advantages even when observers were personal acquaintances who knew targets only outside of the work context. Integrating these findings with previous meta-analyses showed that colleagues' unique perspectives did not predict incrementally beyond commonly held trait perceptions across all raters (except for Openness) and that self-raters who overestimate their Agreeableness and Conscientiousness perform worse on the job. Broadly, our results suggest that observers have clearer lenses for viewing targets' personality traits, and we discuss the theoretical implications of these findings for studying and measuring personality.

Though the bulk of research on personality traits has been built on the backs of self-report measures, trait descriptions from observers has reified much of what we know from self-reports. Emerging from an era when personality's prospects seemed dim (e.g., Guion & Gottier, 1965; Mischel, 1968), finding consensus across raters' perceptions of personality traits validated self-report measures and bolstered the notion that traits are cross-situationally stable (Kenrick

We thank Gabriele Schupp for her help in collecting data.

Correspondence concerning this article should be addressed to Brian S. Connelly, University of Toronto—Management, 1265 Military Trail, Toronto, Ontario M1C1A4, Canada. Email: brian.connelly@utoronto.ca.

Journal of Personality 80:3, June 2012

© 2011 The Authors

Journal of Personality © 2011, Wiley Periodicals, Inc.

DOI: 10.1111/j.1467-6494.2011.00744.x

& Funder, 1988). The longstanding use of observers has shown that observers' reports can be quite accurate (Funder, 1995). However, most of what we know about observers' accuracy comes from studies using observer consensus or self-observer correlations as accuracy criteria. In contrast, across the breadth of studies measuring personality via observer reports, relatively few studies have examined how strongly observer reports predict behavioral criteria, despite the prediction of behavior being highlighted as perhaps the ultimate accuracy criterion (Funder & West, 1993).

However, the handful of studies using observer reports to predict behaviors have shown fascinating and encouraging patterns. For instance, observer reports predict specific behaviors in lab settings (Connelly & Ones, 2010; Kolar, Funder, & Colvin, 1996; Vazire, 2010) and as recorded from daily interactions (Vazire & Mehl, 2008). In addition, observers' reports of personality disorders contribute unique but useful information for diagnosing clinical symptoms, particularly those that are behaviorally manifested (e.g., Oltmanns & Turkheimer, 2006). Relatively small meta-analyses also found that observers' ratings of personality traits were strong predictors of job performance and academic performance (Connelly & Ones, 2010; Oh, Wang, & Mount, in press).

Interestingly, these studies show that personality ratings from a single observer predict performance behaviors (at work and at school) more strongly than do self-ratings. For example, self-reports of Conscientiousness produce operational validity correlations (corrected for performance unreliability but not unreliability in the personality trait) of $\rho_{ov} = .20$ (Barrick, Mount, & Judge, 2001), whereas observer reports of Conscientiousness yield correlations of $\rho_{ov} = .29$ to $.32$ and non-overlapping confidence intervals (Connelly & Ones, 2010; Oh, Mount, & Wang, in press). The differences in prediction between self- and observer reports become even more pronounced when other Big Five traits are combined in multiple regression: a single observer's ratings of the Big Five would predict academic performance at $R_{ov} = .59$ (self-report $R_{ov} = .30$) and job performance at $R_{ov} = .38$ (self-report $R_{ov} = .25$).¹ These findings come at a critical

1. Based on operational validities described in Connelly & Ones (2010). Big Five meta-analytic intercorrelations were drawn for observer ratings from Chang, Connelly, and Geeza (2010) and for self-ratings from Mount, Barrick, Scullen, and Rounds (2005).

time when organizational measurement of personality has come under fire for producing "low" predictive validity (e.g., Morgeson et al., 2007).

Job performance is perhaps the most studied and most important criterion in industrial and organizational (I/O) psychology, but job performance is also an especially rich behavioral criterion for studying behavioral manifestations of personality. For most adults, few domains of life are expected to consume as much time and effort as the workplace. Thus, job performance represents a criterion for weighing the relative accuracy of self- and observer reports that (a) is behaviorally based, (b) is observed outside the lab, (c) is sustained over a period of time, and (d) holds great importance for one's success and well-being.

Finding a stronger prediction of job performance from observers' personality ratings than from self-ratings points to one critical question: what is the source of these differences? On one hand, the observers in job and academic performance studies derived perceptions of targets from the same context as the criterion. That is, job performance studies have always used observers from work (similarly, academic performance studies have always used observers from school). Thus, one plausible explanation is that the validity advantages stem from observers drawing on a narrower perceptual basis that is more closely aligned with the behavioral criterion (i.e., a narrower scope). On the other hand, however, it may be that observers simply view targets more accurately than they view themselves (i.e., a clearer lens). Though these two rival explanations have been acknowledged for a decade and a half (Mount, Barrick, & Strauss, 1994), empirical research has yet to disentangle the two.

The purpose of the present article is to critically examine and empirically test these two potential explanations. In the paragraphs that follow, we provide theoretical elaboration of the "narrower scope" and "clearer lens" explanations. In a primary study, we test whether observers' advantages for predicting job performance extend to ratings from observers who are acquainted with targets outside the workplace (i.e., observers who may have a clearer lens but not a narrower scope). Comparing the validities of personal acquaintances' ratings (observers outside the workplace) to those for self-ratings and colleagues' ratings will indicate whether the clearer lens or the narrower scope explanation accounts best for validity

advantages of observer ratings. Next, we test whether the unique perspectives of self, personal acquaintances, or coworkers contribute to prediction above and beyond the sources' common perspective. Lastly, we describe theoretical contributions and practical applications that may be afforded by using observer ratings.

Explanation 1: Coworkers Have a Narrower Scope Aligned With the Criterion

As noted, when observers' ratings of personality traits have been used to predict job performance, studies have always used work colleagues as observers. Whereas self-rating respondents consider trait information from a variety of contexts when describing themselves, work colleagues typically have only the work context from which to draw trait information. Thus, work colleagues view targets' trait expressions in the workplace without the potential "contamination" of other contexts. Adopting this narrower, uncontaminated, and performance-specific perspective may produce validity advantages for work colleagues over self-ratings.

The logic of the narrower scope explanation rests on satisfying three basic assumptions. First, it assumes that traits are expressed differently across context but consistently within context (i.e., there must be a meaningful work-specific expression of personality). Although context specificity has long been a point of contention in personality research, empirical research is increasingly suggesting that general, cross-situational approaches to studying personality can be complemented by considering situational variability. This notion is at the heart of Mischel's cognitive-affective personality system (Mischel & Shoda, 1995; Wright & Mischel, 1987), though even mainstream five-factor theorists acknowledge that individuals may have characteristic adaptations in trait expression (McCrae & Costa, 1995). Subsequent research has integrated general and situation-specific personality manifestations by emphasizing intraindividual variability in personality (e.g., Fleeson, 2001), behavioral profiles (Furr, 2009), and the development of contextualized identities (e.g., Swann, Johnson, & Bosson, 2009). Heller, Watson, Komar, Min, and Perunovic (2007) outline and evaluate four methodological approaches to measuring this contextualized personality expression: context-specific self-descriptions, diary reporting, experimental priming, and situational priming. However, the use of raters whose

observation is contextually limited represents a potential fifth method.

The second assumption of the narrow scope explanation is that within-context observers' ratings must be sensitive to context-specific expressions of personality. Historically, observer ratings from different contexts have produced substantial intercorrelations, which offered important support that personality traits are at least somewhat stable across situations. However, both primary studies (e.g., Carlson & Furr, 2009; Funder, Kolar, & Blackman, 1995) and a meta-analysis (Geeza, Connelly, & Chang, 2010) have shown that observers within the same context tend to rate personality more consistently than observers across contexts. In particular, context specificity is less pronounced for intimately acquainted observers (Geeza et al., 2010), a finding consistent with predictions from Kenny's (2004) PERSON model of interpersonal perception. Work colleagues (who tend to not be intimately acquainted with targets) may be especially affected by this context specificity.

Lastly, even if observers' perceptions are influenced by context-specific manifestations of personality, the viability of the narrow scope hypothesis ultimately depends on those context-specific manifestations contributing to the prediction of performance above and beyond cross-situational tendencies. Several lines of research offer indirect support of this third assumption. For example, strangers' perceptions of traits predict behaviors similarly or better than close acquaintances' ratings when based on that same behavioral set (Funder & Sneed, 1993). Frame-of-reference approaches to measuring personality have also offered indirect support for the importance of context specificity (Schmit, Ryan, Stierwalt, & Powell, 1995). In frame-of-reference measures, individual self-report items are adjusted by adding context-defining suffixes that direct the respondent to consider a particular context. For example, the item "Instead of procrastinating, I get to work on tasks quickly *at work*" would make respondents consider whether they meet work project deadlines but ignore how they make home improvements. Frame-of-reference approaches have demonstrated some validity advantages over general self-ratings in the prediction of both performance (Lievens, De Corte, & Schollaert, 2008) and domain-specific satisfaction (Heller, Ferris, Brown, & Watson, 2009), though general self-ratings and frame-of-reference measures tend to correlate near the limits of their reliability.

Explanation 2: Observers Have Clearer Lenses for Viewing Targets

In addition to the narrower scope explanation, an alternate explanation is that observers show stronger criterion-related validities than self-ratings because they view targets more clearly than targets view themselves. Existing theory and research suggest that self-reports may have cloudier lenses both because of the kind of information they have access to and because of differences in how this information is used to form perceptions.

First, unlike observers, self-raters have access to both external *and* internal trait expressions. According to the way accuracy has typically been conceptualized (e.g., Funder, 1995), self-raters' access to more information should presumably translate into self-reports being more accurate. However, Hogan and Shelton (Hogan, 1996; Hogan & Shelton, 1998) suggested that personality consists of two components: "inner" personality and "outer" personality. Inner personality depicts personality from the actor's perspective and describes an individual's internal motives and identity, whereas outer personality depicts personality from an observer's perspective and relates to the reputation individuals create in social interactions. Vazire (2010) discussed similar ideas in proposing that there are asymmetries in the knowledge observers and self-raters have about personality traits. Applying this logic to the work context, Hogan and Shelton (1998) argued that because organizations reflect social settings, job performance is likely to be more closely aligned with outer personality than inner personality. Thus, although self-reports' access to internal trait information may be an asset for predicting criteria aligned with "inner" personality, there are theoretical reasons to expect access to "inner" personality" to hamper prediction of criteria aligned with "outer" personality. Following Hogan's socioanalytic theory, one would thus expect self-ratings to show lower validities for job performance than observers' ratings.

There are theoretical grounds from three related research streams to expect that self- and observer raters differ in how they use information to form trait perceptions. First, self-assessments of personality likely suffer from the same errors underlying self-assessments of skills. Specifically, self-assessments are often overly optimistic and only weakly relate to skill and performance indicators (e.g., Falchikov & Boud, 1989; Mabe & West, 1982). Dunning, Heath, and

Suls (2004) argued that self-assessments are error-prone because individuals often fail to account for information that they do not have (e.g., others' general standing on the construct being rated) and because individuals neglect information that they do have (e.g., overlooking negative feedback). Observer assessments, however, tend to be less affected by these errors and are consequently better at predicting performance and life outcomes.

Second, a wealth of research in social psychology on "The Self" has explored processes through which individuals selectively attend to information about oneself (e.g., Baumeister & Newman, 1994; Kunda, 1990). This literature has shown that individuals' self-views are motivated by potentially opposing forces: the motivation to perceive accurately and the motivation to perceive in a particular way (e.g., maintaining a favorable or consistent impression of oneself). When these motivations are at odds, the motivation to perceive oneself in a particular way may cause individuals to actively avoid observing or recalling information disconfirming desired perceptions. Though similar processes could also affect observer ratings of personality traits (e.g., parents forgetting about the deviant behaviors of their little angels), the motivation to maintain consistent and positive perceptions is likely stronger for self-perceptions than for perceptions of others.

Lastly, classic concerns about socially desirable responding (e.g., Crowne & Marlowe, 1960; Furnham, 1986) present a third reason to expect poorer accuracy in self-ratings. Paulhus (1984) presented a theoretically useful distinction between two factors of individual differences in socially desirable responding, self-deceptive enhancement (unconscious inflation tendencies), and impression management (intentional misrepresentation). Even unintentionally, socially desirable responding would be a self-protective motivation more threatening to the accuracy of self-reports than to observer reports.

Purpose

The purpose of the present studies is to evaluate the relative merit of narrower scope and clearer lens explanations by examining the predictive validity of observers who are acquainted outside the work context. This is, to our knowledge, the first study to disentangle rating source from acquaintance context and thus allows a direct comparison of both explanations for predicting performance. In a primary study (Results Part 1), we correlate personality ratings from

personal acquaintances with ratings of job performance to examine whether observers outside the work context produce the same validity advantages as those inside the work context. The typical finding from previous organizational personality research has been that self-ratings of Conscientiousness predict performance with observed correlations of $r = .12$; when appropriate corrections for range restriction and unreliability in performance measures are accounted for (i.e., operational validities are estimated, denoted as ρ_{ov}), correlations increase to $\rho_{ov} = .20$ (Barrick, Mount, & Judge, 2001). Self-ratings of other Big Five traits tend to produce observed correlations with performance below $r = .10$ when estimated across occupations/criteria, though using more specific occupations or criteria produces stronger correlations for Emotional Stability and Agreeableness. Thus, Conscientiousness has stood out as the trait most predictive across jobs. Although Conscientiousness's operational validity of .20 is conventionally regarded as only a small to moderate-sized effect, correlations of such magnitude can have tremendous real-world implications when it comes to selecting job applicants. Similarly, even small increases in observed predictive power can produce substantial increases in the average performance of selected job applicants, such as the increase in operational validities from $\rho_{ov} = .20$ for self-ratings of Conscientiousness to $\rho_{ov} = .28$ for colleagues' ratings. Thus, we focus on Conscientiousness in particular because it has been the driver of most of personality's prediction of job performance as well as being the trait for which observer advantages have been most pronounced. We also examine predictive validity differences for the other Big Five traits, as well as examining multiple facets of overall job performance (in-role behavior and organizational citizenship behavior). In Part 2 of our results, we integrate meta-analytic validities from self-, coworker, and personal acquaintance ratings of personality.

We use these data to address four research questions. If the narrower scope hypothesis is correct, personal acquaintance ratings in the primary study should (1) yield predictive validities near zero, (2) not incrementally predict performance beyond self-reports, and (3) be less than meta-analytic estimates for colleagues' validities and be less than or equal to meta-analytic estimates of self-ratings' validity. In addition, (4) the unique information held by work colleagues should predict job performance above and beyond the information commonly held across self-, personal acquaintance, and colleague

ratings. Finding support for the narrower scope hypothesis would indicate that the context from which information is drawn is tremendously important for predicting job performance, and directing raters to focus only on the context of interest may be especially fruitful.

On the other hand, if the clearer lens hypothesis is correct, personal acquaintances' ratings in the primary study should (1) produce strong predictive validities, (2) predict performance incrementally beyond self-ratings, and (3) yield operational validities that are greater than meta-analytic estimates for self-ratings and greater than or equal to meta-analytic estimates for colleague ratings. In addition, (4) the unique information (residual variance) of a particular rating source would not contribute incrementally to predicting job performance beyond the trait perspective common across all rating sources. Finding such results would highlight the necessity for considering the differential perspectives of self- and observer raters.

PRIMARY STUDY METHODS

Sample

Participants were 319 adults living in Germany. One hundred eleven employees provided self-ratings and were in turn assessed by 106 personal acquaintances and 102 work colleagues. Employees were instructed to distribute observer rating forms to one personal acquaintance and one work colleague, who then independently returned the forms directly to the researchers via post. Employees who had filled out the self-rating form were not allowed to provide observer ratings for personal acquaintances or colleagues. Employees, personal acquaintances, and colleagues were primarily women (73.9%, 44.3%, and 68.3%, respectively) and had mean ages of 37.8 ($SD = 10.8$), 38.6 ($SD = 12.7$), and 37.5 ($SD = 11.0$) years, respectively. On average, employees had about 15 years of work experience. Thirty-three percent of personal acquaintances were friends, 24.1% were spouses, 23.2% were life partners, 4.5% were parents, 3.6% were siblings, 2.7% were children, and 3.6% were listed as "other" (relationship status was not reported for six personal acquaintances). Personal acquaintances knew targets on average for 14.8 years ($SD = 13.08$), reported to know targets very well ($M = 7.94$, $SD = .99$ on a scale from 1 to 9), reported having a good

relationship with targets ($M = 8.42$, $SD = .78$ on a scale from 1 to 9), and met targets frequently in their spare time ($M = 8.19$, $SD = 1.04$ on a scale from 1 to 9). Similarly, work colleagues knew targets on average for 7.97 years ($SD = 9.10$), knew targets well ($M = 6.55$, $SD = 1.32$), and reported having a good relationship with targets ($M = 7.34$, $SD = 1.24$). However, colleagues reported meeting targets only occasionally in their spare time ($M = 4.73$, $SD = 2.40$).

In total, complete data with (a) a self-rating of personality, (b) a non-colleague personal acquaintance's rating of the target's personality, (c) a work colleague's rating of personality, and (d) the same colleague's rating of job performance were available for 97 individuals. Thus, the same work colleagues who provided colleague ratings of personality also provided job performance ratings. Analyses were conducted for individuals with complete data ($N = 97$ targets rated in total by 291 raters).

Measures

NEO-PI-R (German version). Self-ratings, colleague ratings, and personal acquaintance ratings of personality were collected via the German translation of the self- and peer report forms of the NEO-PI-R (Ostendorf & Angleitner, 1994). The NEO-PI-R contains 240 items that measure five factors (Emotional Stability,² Extraversion, Openness to Experience, Agreeableness, and Conscientiousness) and 30 facet traits organized within the Five-Factor Model. The German translation of the NEO-PI-R has produced reliability coefficients and a factor structure comparable to its English counterpart for both self-report and peer report forms, and these German translations have been widely used in cross-cultural and behavioral genetics research on personality (e.g., Kandler et al., 2010). Cronbach's alphas for the NEO-PI-R domain scales ranged between .84 and .95.

Job performance criteria. Job performance criteria were assessed by colleagues' ratings. Both in-role behavior and organizational citizenship behavior (OCB) were assessed with a questionnaire provided by Staufienbiel and Hartz (2000). This questionnaire is a well-established set of scales that is widely used in German-speaking

2. Originally, this factor measured Neuroticism. To ease interpretation, we have reversed the signs of correlations with this scale such that positive correlations indicate that high Emotional Stability predicts higher criterion scores.

samples (e.g., Binnewies, Sonnentag, & Mojza, 2009). The five-item in-role behavior subscale measures task performance components of job performance ($\alpha = .85$) that represent performance in the job's core task functions. The other four OCB subscales measure organizational citizenship behaviors (altruism, general compliance, sportsmanship, and initiative) and reflect performance components that contribute to the work environment more than to one's own tasks. A composite OCB score was calculated by averaging standardized scores on the altruism, general compliance, sportsmanship, and initiative scales ($\alpha = .84$). Similarly, an overall job performance composite score was calculated by averaging standardized scores on the in-role behavior scale and the OCB composite ($\alpha = .82$).

RESULTS PART 1: PRIMARY STUDY FINDINGS

Table 1 presents means, standard deviations, reliabilities, and intercorrelations among the primary study performance criteria and personality scores from self-, personal acquaintance, and colleague ratings. Note that the convergent validities (i.e., mono-trait, hetero-method correlations) are relatively strong and are consistently larger than hetero-trait, mono-method correlations. These findings show that trait perceptions of self-, personal acquaintance, and colleague raters overlap considerably and are distinctive from perceptions of unrelated traits.

Research Question 1: Predicting Performance from Personality Ratings

Table 1 presents observed correlations (with no corrections for unreliability) between personality ratings and performance. Self-ratings of personality produced a similar pattern as that observed in previous research. That is, self-ratings of Conscientiousness correlated moderately with overall performance (uncorrected $r = .21$) and with in-role behaviors (uncorrected $r = .25$; correlations with organizational citizenship behaviors were more modest), but correlations between self-ratings of other Big Five traits and performance were generally lower.

Central to weighing the relative merits of the narrow scope and clearer lens hypotheses is whether personal acquaintances' ratings will produce strong correlations with performance. Compared to

Table 1
Means, Standard Deviations, Reliabilities, and Intercorrelations Among Primary Study Variables

	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
1. Perf.	.00	1.00	.82																						
2. In-role	22.49	2.74	.92	.85																					
3. OCB	.00	1.00	.92	.69	.84																				
4. Altruism	19.50	3.48	.67	.47	.77	.78																			
5. G. Comp	21.35	3.38	.70	.63	.65	.29	.77																		
6. Sports	19.70	3.07	.51	.33	.61	.30	.17	.66																	
7. Initiative	18.13	3.43	.67	.48	.75	.55	.33	.22	.80																
8. ES self	85.40	22.66	-.06	-.12	.01	-.12	-.14	.12	.17	.93															
9. ES pacq	82.09	21.61	.18	.16	.17	-.02	.04	.21	.23	.40	.92														
10. ES col	79.54	21.58	.33	.21	.41	.27	.09	.51	.26	.27	.32	.93													
11. Ex self	114.34	16.69	-.03	-.04	-.02	.05	-.29	-.01	.19	.25	.28	.19	.86												
12. Ex pacq	114.45	18.98	.06	.08	.03	.06	-.15	-.04	.20	.03	.39	.08	.64	.89											
13. Ex col	114.14	18.48	.30	.23	.32	.43	-.10	.17	.38	.03	.18	.34	.52	.47	.90										
14. O self	114.41	16.92	.05	.05	.03	.11	-.15	-.02	.14	.10	.12	.18	.34	.12	.16	.86									
15. O pacq	106.14	16.98	.06	.15	-.04	.15	-.12	-.18	.05	-.05	.13	.03	.30	.44	.25	.56	.87								
16. O col	105.53	16.37	.35	.26	.39	.50	.07	.16	.36	-.09	.02	.18	.15	.13	.55	.38	.31	.89							
17. A self	118.55	14.63	-.08	-.13	-.02	.13	-.17	-.06	.04	.14	-.03	-.09	-.09	.05	.00	.05	.02	-.06	.84						
18. A pacq	117.92	19.58	.15	.14	.13	.21	.19	.00	-.04	-.11	.11	.04	-.07	.01	.04	-.05	.18	.07	.41	.91					
19. A col	114.60	21.66	.35	.18	.45	.55	.26	.35	.11	-.15	-.17	.21	-.16	-.11	.10	-.10	-.17	.22	.31	.26	.93				
20. C self	120.97	17.79	.21	.25	.13	-.07	.26	-.01	.17	.40	.35	.09	.15	.05	-.02	.00	-.08	-.11	.09	.01	-.10	.89			
21. C pacq	125.85	24.38	.33	.41	.19	.03	.29	-.01	.20	.27	.53	.09	.14	.26	.03	.11	.32	.12	.00	.15	-.07	.51	.95		
22. C col	126.09	19.35	.61	.57	.54	.33	.58	.17	.43	.01	.20	.26	-.07	.02	.22	-.01	-.05	.33	-.08	.00	.32	.35	.39	.92	

Note. *N* = 97. Perf. = performance; OCB = organizational citizenship behavior; G. Comp = general compliance; Sports = sportsmanship; ES = Emotional Stability; Ex = Extraversion; O = Openness; A = Agreeableness; C = Conscientiousness; self = self-ratings; pacq = personal acquaintance ratings; col = colleague ratings. Internal consistency reliabilities (α) are italicized along the main diagonal of the matrix; mono-trait, hetero-method correlations are boldfaced. Because of potential method bias between performance and colleague personality ratings, correlations between these variables should be interpreted cautiously. Correlations whose absolute value is greater than $r = .20$ have 95% confidence intervals that do not include zero/ $p < .05$.

correlations from self-reports, personal acquaintances' ratings of Conscientiousness are strongly predictive of performance (for overall job performance, $r_{pacq} = .33$ vs. $r_{self} = .21$; for in-role behaviors, $r_{pacq} = .41$ vs. $r_{self} = .25$; for organizational citizenship behaviors, $r_{pacq} = .19$ vs. $r_{self} = .13$). These predictive validity advantages for personal acquaintance ratings become more pronounced when self and personal acquaintance validities are corrected for criterion unreliability: for overall job performance, $\rho_{ovpacq} = .46$ versus $\rho_{ovself} = .29$; for in-role behaviors, $\rho_{ovpacq} = .52$ versus $\rho_{ovself} = .31$; for organizational citizenship behaviors, $\rho_{ovpacq} = .26$ versus $\rho_{ovself} = .18$.³ As was the case with self-reports, personal acquaintance ratings of the other Big Five traits were more modest. More specific and conceptually related performance facets were related to personal acquaintances' ratings of Emotional Stability (sportsmanship and initiative) and Agreeableness (altruism), with these uncorrected correlations being larger than those typically observed for self-reports. These results offer encouraging support for predicting performance criteria from observer ratings of Conscientiousness (and for predicting more specific performance facets from Emotional Stability and Agreeableness), even when observers' perceptions of the target come from interactions outside the workplace.

Lastly, note that although Table 2 also provides correlations between colleague ratings of personality and performance criteria, these correlations are likely affected by common method bias because they were provided by the same rater. We therefore caution against comparing these validities for colleagues' trait ratings to those from sources independent of the criterion.

Research Question 2: Incremental Validity of Personal Acquaintance Ratings

Next, we examined the incremental prediction from combining self-ratings and personal acquaintance ratings in multiple regression compared to using either rating source individually. The incremental

3. Interrater reliabilities were drawn from Viswesvaran, Ones, and Schmidt (1996) to make performance unreliability corrections. For correlations with overall job performance, Viswesvaran and colleagues' reliability estimate for overall job performance ($r_{xx} = .52$) was used. For correlations with in-role behaviors, Viswesvaran and colleagues' reliability estimate for productivity ($r_{xx} = .63$) was used. For organizational citizenship behaviors, an average of effort, interpersonal competence, and compliance with authority was used ($r_{xx} = .53$).

Table 2
Regressions Combining Self + Acquaintance Ratings to Predict Performance Criteria

	Performance			In-Role			OCB		
	<i>R</i>	ΔR	β	<i>R</i>	ΔR	β	<i>R</i>	ΔR	β
Emotional Stability	.23 (.00, .27)			.26* (.00, .42)			.18 (.00, .35)		
Self		.05	-.16* (-.37, -.05)		.10*	-.22* (-.43, -.02)		.01	-.07 (-.28, .15)
Personal acquaintance		.17*	.24* (.04, .45)		.14*	.25* (.05, .46)		.17	.19 (-.02, .37)
Extraversion	.11 (.00, .27)			.14 (.00, .31)			.06 (.00, .20)		
Self		.05	-.12 (-.37, .14)		.07	-.15 (-.41, .10)		.03	-.06 (-.32, .20)
Personal acquaintance		.07	.13 (-.13, .29)		.10	.17 (-.08, .43)		.03	.07 (-.19, .33)
Openness	.06 (.00, .20)			.15 (.00, .32)			.07 (.00, .22)		
Self		.00	.02 (-.22, .26)		.00	-.04 (-.27, .20)		.03	.07 (-.17, .31)
Personal acquaintance		.02	.05 (-.19, .29)		.10	.17 (-.06, .41)		.04	-.08 (-.31, .16)
Agreeableness	.22 (.00, .38)			.25* (.00, .41)			.15 (.00, .32)		
Self		.07	-.17 (-.38, .04)		.11*	-.23* (-.43, -.02)		.02	-.09 (-.31, .12)
Personal acquaintance		.13*	.22* (.01, .43)		.12*	.24* (.03, .44)		.13	.17 (-.05, .38)
Conscientiousness	.33* (.11, .48)			.42* (.22, .56)			.19 (.00, .36)		
Self		.00	.05 (-.16, .27)		.00	.06 (-.15, .27)		.00	.04 (-.18, .27)
Personal acquaintance		.12*	.30* (.09, .51)		.16*	.38* (.19, .58)		.06	.16 (-.06, .39)

Note. $N = 97$. OCB = organizational citizenship behavior. ΔR indicates improvement in *R* when term is added in the combined model beyond the other rating source (e.g., ΔR for self indicates the change in *R* for regressing performance on self + personal acquaintance compared to personal acquaintance alone). Confidence intervals for effect sizes are provided parenthetically.
 * $p < .05$.

prediction (ΔR) for self-ratings and personal acquaintance ratings indicates the unique predictive power of each rating source, and the regression weights (β) indicate the relative importance of each rating source when combined. The results of these analyses are shown in Table 2.

Consistent with the clearer lens hypothesis, adding personal acquaintance ratings of Conscientiousness to self-ratings yielded noteworthy increments in multiple- R values for predicting overall job performance ($\Delta R = .12$) and for predicting in-role behaviors ($\Delta R = .16$). The converse, however, did not hold: adding a self-rating yielded no increases in multiple- R values above a single personal acquaintance. In addition, the β weights indicate that personal acquaintance ratings dominated the prediction. Thus, for Conscientiousness, self-ratings provided no information relevant for predicting performance criteria that was not contained in a personal acquaintance's rating.

Consistent with zero-order results, multiple- R values were modest for Extraversion and Openness even when self- and personal acquaintance ratings were combined. For predicting overall performance and in-role behaviors, combining self- and personal acquaintance ratings of Emotional Stability and Agreeableness produced noteworthy increments in multiple- R values both beyond a self-rating alone. The β weights indicate that both self-ratings and personal acquaintance ratings contain unique information for predicting performance criteria. However, note that these β s are similar in magnitude but opposite in direction for self- and personal acquaintance ratings. Such a pattern indicates that overall performance and in-role behaviors are (positively) predicted by the difference between personal acquaintance and self-ratings (Edwards, 1994).⁴ That is, overestimates from self-ratings of Emotional Stability and Agreeableness actually yield lower job performance. These effects indicate that misperceiving one's own Emotional Stability or Agreeableness has negative implications for overall performance and

4. Following procedures outlined by Edwards (1994), constraining the regression weights for self- and personal acquaintance ratings to be equal and opposite produced minimal (nonsignificant) reductions in variance explained. For Emotional Stability, $\Delta R = -.01$ and $\Delta R = .00$ for overall performance and in-role behaviors, respectively. For Agreeableness, $\Delta R = .00$ and $\Delta R = -.01$ for overall performance and in-role behaviors, respectively.

in-role behaviors, and personal acquaintances' ratings of Emotional Stability and Agreeableness may be most useful when viewed in contrast to self-ratings.

RESULTS PART 2: META-ANALYTIC INTEGRATION ACROSS SELF, PERSONAL ACQUAINTANCE, AND COLLEAGUE SOURCES

In Part 2 of our results, we integrate findings for personal acquaintances' ratings in our primary study with those of existing meta-analytic estimates for self-ratings and colleague ratings. Specifically, we compare (a) the operational validities of ratings from personal acquaintances from the primary study reported in Part 1 to meta-analytic estimates for self-reports and colleague ratings of personality and (b) the extent to which predictive validities for rating sources are driven by common versus unique perspectives of the target.

Research Question 3: Comparing Performance Prediction Across Rating Sources

Our third research question compared personal acquaintance performance prediction correlations from our primary study to the best available estimates for self- and colleague raters (i.e., meta-analytic correlations). We converted the correlations in our primary study between personal acquaintances' ratings of traits predicting overall job performance to operational validities by correcting for criterion unreliability. Table 3 compares these personal acquaintance operational validities to meta-analytic estimates of operational validities for self-ratings and colleague ratings of traits. The magnitude of these differences in operational validities across rating sources must be weighed against variability in correlations due to both (a) sampling error (as indicated by Table 3's confidence intervals) and (b) potential differences in underlying true correlations (as indicated by Table 3's credibility intervals). Because these comparisons have implied directions (i.e., whether personal acquaintance ratings are more predictive than self-reports, as predicted by the clearer lens hypothesis, and whether personal acquaintance ratings are less predictive than colleague ratings, as predicted by the narrow scope hypothesis), we used 90% confidence intervals (80% intervals are typical for credibility intervals; Hunter & Schmidt, 1994).

Table 3
Comparison of Personal Acquaintance Operational Validities in the Present Study to Meta-Analytic
Estimates for Self-Reports and Colleague Reports

	Operational Validities for Job Performance			Differences in Validities		
	Personal Acq. ρ_{ov}	Self-Reports ^a ρ_{ov} (80% Cred.)	Colleague ^b ρ_{ov} (80% Cred.)	P. Acq. – Self $\rho_{ov,pacq} - \rho_{ov,self}$ (90% Conf.)	P. Acq. – Coll. $\rho_{ov,pacq} - \rho_{ov,call}$ (90% Conf.)	Coll. – Self $\rho_{ov,call} - \rho_{ov,self}$ (90% Conf.)
Emotional Stability	.25	.11 (.00, .21)	.17 (.17, .17)	.14 (-02, .31)	.08 (-09, .25)	.06 (.01, .11)
Extraversion	.08	.11 (-05, .26)	.11 (-02, .24)	-.02 (-19, .14)	-.03 (-20, .15)	.00 (-05, .05)
Openness to Experience	.08	.04 (-10, .18)	.24 (.24, .24)	.04 (-13, .21)	-.15 (-33, .02)	.20 (.14, .25)
Agreeableness	.21	.11 (00, .23)	.16 (.16, .16)	.10 (-07, .26)	.04 (-13, .21)	.05 (00, .10)
Conscientiousness	.46	.20 (.07, .33)	.28 (.28, .28)	.26 (.11, .40)	.18 (.03, .34)	.07 (.03, .12)

Note. The 80% credibility intervals are calculated based on $SD_{\rho_{ov}}$'s presented in original meta-analyses. Italicized credibility intervals indicate credibility intervals that do not include the personal acquaintance point estimate of ρ_{ov} . The 90% confidence intervals represent the confidence interval around the differences in correlations. Boldfaced differences in validities indicate differences in validities whose 90% confidence intervals do not include zero.

^aBased on values from Barrick et al. (2001), independent samples only.

^bBased on values from Connelly and Ones (2010), coworkers only.

For Conscientiousness, operational validities for personal acquaintances are larger than self-reports by a substantial margin ($\rho_{ovpacq} - \rho_{ovself} = .26$). Contrary to what would be predicted by the narrow scope hypothesis, operational validities for personal acquaintances' ratings also exceeded those of colleagues for Conscientiousness, though the margin was less ($\rho_{ovpacq} - \rho_{ovcoll} = .18$). Emotional Stability and Agreeableness also showed advantages beyond self-ratings, though these advantages were less pronounced than those for Conscientiousness and their confidence intervals included zero. Openness was the only trait for which colleagues' ratings trended toward producing stronger operational validities than personal acquaintances' ratings ($\rho_{ovpacq} - \rho_{ovcoll} = -.15$) as the narrow scope hypothesis would predict, though the confidence interval of this difference included zero. These findings suggest that the validity advantage for personal acquaintance ratings above self-ratings for Conscientiousness (and, to a lesser extent, for Emotional Stability and Agreeableness) is not likely purely the result of sampling error or differences in study design.

Research Question 4: Validity From Unique Rater Perspectives

To separate the prediction of job performance that is due to common rater perspectives from that due to unique rater perspectives, we conducted a series of structural equation model (SEM) analyses. We solicited meta-analytic input intercorrelation matrices between job performance and self-ratings (Barrick et al., 2001), personal acquaintance ratings (the present primary study), and colleague ratings (Connelly & Ones, 2010) of personality separately for each of the Big Five. Thus, these sets of correlations formed observed (uncorrected) intercorrelations between job performance and self-, personal acquaintance, and colleague ratings. We fit five base models (one for each of the Big Five; see Figure 1) predicting job performance from the common factor across self-, personal acquaintance, and colleague perspectives of a target's standing on the trait. To correct for unreliability in the criterion (and to ensure that models were not underidentified), we created two indicators of overall job performance that represented two different raters of job performance. In the base models, the job performance factor is regressed on only the higher-order latent trait factor.

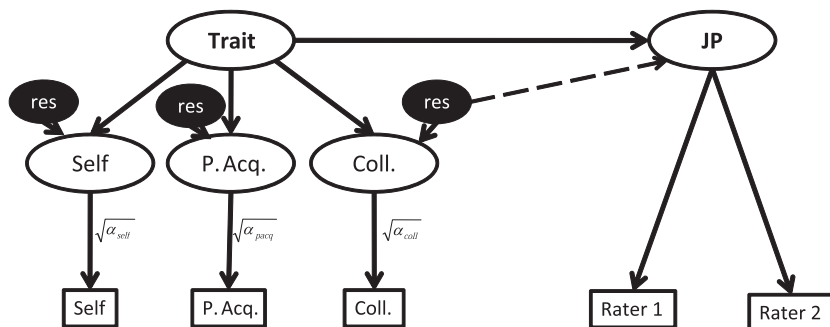


Figure 1

Base SEM (solid lines only) and colleague SEM (solid and dashed lines). JP = job performance; res = residual variance.

To test whether the unique information held by self-, personal acquaintance, and coworker ratings predicted job performance above and beyond the common variance across rating sources, paths from the residual variance in the self-, personal acquaintance, and coworker perspectives to the latent job performance factor were individually added and then removed from the base model. If raters have unique perspectives relevant for predicting job performance, adding these paths should (a) improve overall model fit, (b) increase the variance explained in the latent job performance factor, and (c) produce strong β weights for the unique rater perspective factor relative to the β weights for the common trait factor.

The results for the SEM models are shown in Table 4. These results show several general trends across traits. Note first that generally all base models fit their data well, with CFI (Comparative Fit Index) and TLI (Tucker-Lewis) values generally at or in excess of 1.0 and SRMR (Standardized Root Mean Square Residual) values generally below .05 (Openness represents the only exception). Thus, models in which the common variance alone predicts job performance are not inadequate for explaining the pattern of observed correlations. Because models specifying paths predicting job performance from unique rater variance have relatively little room within which to improve model fit, increases in R_{est} values and the relative strength of β weights serve as indicators of the importance of unique rater perspectives.

Table 4 shows that the common variance across raters was strongly predictive of job performance for Emotional Stability,

Table 4
SEM Results Examining Common and Unique Rater Prediction of Overall Job Performance

	Model Fit				Trait Loadings (SE _λ)				β weights (SE _β)			
	χ ²	df	CFI	TLI	SRMR	λ _{self}	λ _{peer}	λ _{col}	R _{cor} (SE _R)	Common	Rater	
Emotional Stability (N = 264)												
Common only	2.40	5	1.00	1.03	.03	.47 (.07)	.88 (.10)	.40 (.07)	.30 (.07)	.30 (.08)	—	
Colleague	.35	4	1.00	1.05	.01	.45 (.08)	.94 (.13)	.37 (.08)	.30 (.07)	.23 (.09)	.12 (.08)	
Self	1.50	4	1.00	1.04	.02	.50 (.08)	.83 (.10)	.42 (.07)	.34 (.08)	.38 (.12)	-.10 (.10)	
P. Acquaintance	2.31	4	1.00	1.02	.03	.46 (.08)	.90 (.13)	.39 (.08)	.34 (.16)	.50 (.90)	-.19 (.86)	
Extraversion (N = 269)												
Common only	1.79	5	1.00	1.03	.02	.55 (.06)	.93 (.07)	.57 (.06)	.11 (.06)	.11 (.08)	—	
Colleague	.44	4	1.00	1.04	.01	.54 (.06)	.94 (.07)	.56 (.06)	.13 (.06)	.04 (.10)	.11 (.09)	
Self	1.46	4	1.00	1.03	.02	.54 (.06)	.94 (.07)	.56 (.06)	.11 (.06)	.07 (.10)	.05 (.09)	
P. Acquaintance	.14	4	1.00	1.04	.01	.54 (.06)	.95 (.07)	.56 (.06)	.37 (.25)	1.10 (1.58)	-.96 (1.56)	
Openness (N = 266)												
Common only	12.64	5	.96	.93	.06	.63 (.07)	.79 (.08)	.45 (.06)	.15 (.07)	.16 (.09)	—	
Colleague	.11	4	1.00	1.05	.01	.62 (.07)	.80 (.08)	.44 (.07)	.29 (.06)	-.03 (.10)	.30 (.08)	
Self	11.26	4	.97	.91	.05	.64 (.08)	.75 (.08)	.47 (.07)	.24 (.09)	.32 (.16)	-.16 (.13)	
P. Acquaintance	10.55	4	.97	.92	.05	.59 (.07)	.82 (.08)	.46 (.07)	.35 (.14)	.60 (.38)	-.40 (.35)	
Agreeableness (N = 269)												
Common only	3.68	5	1.00	1.02	.03	.61 (.07)	.63 (.07)	.49 (.07)	.28 (.08)	.29 (.09)	—	
Colleague	2.66	4	1.00	1.02	.02	.62 (.08)	.63 (.08)	.47 (.07)	.27 (.08)	.21 (.12)	.10 (.09)	
Self	.06	4	1.00	1.06	.00	.66 (.08)	.59 (.07)	.48 (.07)	.43 (.11)	.56 (.19)	-.27 (.17)	
P. Acquaintance	2.66	4	1.00	1.02	.03	.64 (.08)	.60 (.08)	.49 (.07)	.25 (.07)	.15 (.15)	.13 (.12)	
Conscientiousness (N = 269)												
Common only	5.36	5	1.00	1.00	.04	.53 (.06)	.95 (.06)	.45 (.06)	.50 (.07)	.50 (.07)	—	
Colleague	1.44	5	1.00	1.03	.02	.51 (.05)	1.00 (.00)	.42 (.05)	.49 (.06)	.40 (.07)	.16 (.07)	
Self	1.80	4	1.00	1.02	.02	.58 (.06)	.87 (.06)	.48 (.06)	.58 (.08)	.67 (.12)	-.21 (.11)	
P. Acquaintance	5.36	4	1.00	.99	.04	.54 (.07)	.94 (.09)	.45 (.07)	.49 (.09)	.41 (.10)	.08 (1.01)	

Note. Ns are based on the harmonic mean of total sample sizes for correlations in the meta-analytic intercorrelation matrix (Viswesvaran & Ones, 1995).

Agreeableness, and Conscientiousness ($\beta = .30, .29$, and $.50$, respectively). These paths are notably twice the magnitude of past true score estimates for these traits' validities based on self-reports (Barrick et al., 2001). These findings suggest that the importance of Emotional Stability, Agreeableness, and Conscientiousness has been severely undervalued by overreliance on single self-report measures.

Across traits, personal acquaintances' perspectives had the strongest loadings on the common trait (ranging from $\lambda_{pacq} = .63$ for Agreeableness to $\lambda_{pacq} = .95$ for Conscientiousness). These loadings were generally so strong that models adding the path from the uniqueness of the personal acquaintance perspective to job performance produced large standard errors for R_{est} values and β weights. This pattern is akin to multicollinearity problems between personal acquaintance perspectives and the common trait. Thus, though personal acquaintances are strong indicators of how targets are generally perceived, they do not hold any unique predictive power.

Interestingly, Table 4 generally suggests minimal increases in predictive power for models predicting job performance from the common trait *plus* the unique perspectives of colleagues. For most traits, R_{est} values in these models show little improvement over base models, and models adding the path for colleagues did not substantially redistribute β weights assigned to the common trait. Findings for Openness, however, represent an important exception to this pattern observed for other traits ($R_{est} = .29$ for the colleague model vs. $R_{est} = .15$ for the base model; $\beta_{colleague} = .30$ vs. $\beta_{common} = -.03$). Thus, for Openness, colleagues' unique perceptions seem to carry the most weight for predicting job performance, though this pattern did not emerge for any other trait.

Finally, Table 4 shows a fascinating pattern for models in which the unique perspectives of self-raters were added to the common trait to predict job performance. For Agreeableness and Conscientiousness, R_{est} values in these models increased when the unique perspectives of self-raters were added. However, the β weights for self-perspective uniquenesses were negative, and the β weights for the common trait increased when the self-report uniqueness was added (for Agreeableness, $\beta_{com.rater} - \beta_{com} = .27$, 95% CI = $[.01, .53]$; for Conscientiousness, $\beta_{com.rater} - \beta_{com} = .17$, 95% CI = $[.02, .33]$). This indicates that self-reports' unique variance acts as a suppressor of the relationship between trait factors and performance. Thus, holding constant the way a target is generally perceived, targets who perceive

themselves as more agreeable and conscientious tend to perform worse on the job (i.e., overestimators are worse employees and underestimators are better employees).

GENERAL DISCUSSION

The purpose of these analyses was to examine the relative merit of the clearer lens and narrower scope explanations for the advantages previously found for observers' ratings of personality for predicting job performance. Findings across both the primary study and the integration with meta-analytic results were generally consistent with the clearer lens hypothesis. For Conscientiousness, the trait most predictive of performance, ratings from targets' personal acquaintances (1) correlated meaningfully with performance criteria, (2) yielded incremental prediction beyond self-ratings, and (3) were more strongly predictive of job performance than were self-ratings. In addition, (4) the unique variance associated with a colleague's perspective of the target's trait did not generally predict job performance beyond the common perception across raters. Furthermore, although the effects were more marginal, the point estimates generally suggested that observers may generally have clearer lenses for Agreeableness and Emotional Stability when it comes to the prediction of job performance.

Finding that self-ratings have "cloudy lenses" raises as many questions as it answers and indicates directions for future research. Traditionally, I/O psychologists have focused on response distortion in self-ratings as intentional and purposeful misrepresentation among job applicants hoping to secure a job offer. However, finding validity advantages for observers over self-raters—even self-raters with low incentive to distort—suggests that prediction from "honest" self-perceptions has depreciated because of self-misperceptions. Motivated response distortion may thus not fully explain the validity disadvantages of self-ratings, and future research would be well directed to contrast specific impression formation and item response thought processes of self-raters versus observers. Such contrasts may point toward methods for improving self-report measures if self-raters can be coaxed to mimic observers.

For the broader community of personality psychologists, the present findings raise important questions about whether self-reports' cloudy lenses affect predictions of other criteria. Although

observers have also shown advantages beyond self-reports for predicting grades (Connelly & Ones, 2010), early discharge from the military (Fiedler, Oltmanns, & Turkheimer, 2004), and the development of coronary heart disease (Smith et al., 2008), predicting emotions has typically favored self-reports over observers (e.g., Spain, Eaton, & Funder, 2000). This pattern of findings is consistent with Hogan's (1996) notion of inner versus outer aspects of personality differentially predicting inner versus outer criteria and with Vazire's (2010) Self-Other Knowledge Asymmetry model. However, most inner criteria like emotions tend to be measured via self-reports. Thus, self-reports of personality may benefit from a common-method advantage in predicting inner criteria because there can only be one self-rater (in contrast, it is possible to have separate observers rate personality and "outer" criteria). Thus, future research should examine whether self-reports have "clearer lenses" for predicting inner criteria not measured by self-reports (perhaps by studying psychophysiological outcomes of personality).

The narrower scope hypothesis held for only one trait, Openness. In the results of the primary study, personal acquaintances' ratings of Openness were not strongly predictive of job performance (whereas coworkers' ratings of Openness showed strong validity; Connelly & Ones, 2010). In addition, Openness was the one trait for which the unique perspective of coworkers incrementally predicted performance. This is perhaps surprising, given that self-report findings for Openness have generally shown it to be an especially weak predictor of job performance. However, several researchers (Chernyshenko, Stark, Woo, & Conz, 2008; DeYoung, Quilty, & Peterson, 2007) have noted that Openness has two distinguishable components: an intellectance meso-facet that describes enjoying complex thinking and an experiencing meso-facet that describes a propinquity toward external stimuli (e.g., aesthetics, imagination). This intellectance component aligns more closely with performance criteria and is likely more salient in the workplace (Chernyshenko et al., 2008). Work colleagues' ratings of Openness may more closely reflect this component of Openness and, as a result, produce stronger validities for Openness. On the other hand, the experiencing meso-facet may be more salient to personal acquaintances but less important in the workplace, making personal acquaintances' ratings of Openness "contaminated" when it comes to predicting performance criteria.

Particularly interestingly, discrepancies between self- and observer ratings of Agreeableness and Emotional Stability predicted performance, and self-ratings of Agreeableness and Conscientiousness produced suppression effects when combined with observer ratings (i.e., negative β weights for self-ratings). The residual variance associated with self-ratings is often regarded simply as random error. However, recent behavioral genetics research has found that unique self-views are genetically influenced and stable over time (for at least 13 years; Kandler et al., 2010). Indeed, personality researchers are increasingly viewing self-rater specific variance not as error but as indicating self-rater biases such as self-enhancement or self-deception tendencies. Our results complement these findings. These unique self-views have not only stability and genetic bases but also behavioral consequences: Individuals with inflated self-views of Agreeableness, Conscientiousness, and Emotional Stability may perform worse on the job. These unique self-views are clearly not purely random error but merit further scrutinizing of their etiology and outcomes, necessitating multirater approaches to measuring personality.

These suppression findings corroborate other lines of research suggesting that overly positive self-views may be maladaptive for performance outcomes: Self-deception is negatively related to actual learning performance (Hirschfeld, Thomas, & McNatt, 2008), and overraters of job performance perform worse on the job and in academic group work (Atwater, Ostroff, Yammarino, & Fleenor, 1998; Kwan, John, Kenny, Bond, & Robins, 2004). In addition, underestimators may be more skilled at adapting to compensate for their standing on the trait. For example, those with accurate or conservative estimates of a low standing on Agreeableness might avoid tasks particularly calling for interpersonal sensitivity, whereas those who overestimate Agreeableness might surge ahead toward likely failure. As an anonymous reviewer noted, it is unclear whether these suppression effects are driven more by the performance costs of self-enhancement or by the performance benefits of self-effacement. Future research using large primary samples could use polynomial regression models to disentangle the effects of self-enhancement versus self-effacement (e.g., Edwards, 1994; cf. Atwater et al., 1998). However, a broad literature has shown many negative effects of self-enhancement that may elicit resentment from others: Self-enhancers underestimate the requirements to learn and prepare for

important tasks (Hirschfeld et al., 2008) and are described as narcissistic, hostile, and defensive (Colvin, Block, & Funder, 1995; Paulhus, 1998). Thus, this suppression effect is most likely driven at least in part by costs of self-enhancement, but it would be interesting and useful to distinguish self-enhancement and self-effacement in future studies.

These findings have shown observer ratings to be a valuable method for measuring personality. This has practical implications for organizational measurement of personality, which has recently come under fire for producing modest observed correlations with performance (Morgeson et al., 2007). However, the operational validities from observers compete with validities reported for some of the most valid selection methods, such as tests of general mental ability and work sample tests (Schmidt & Hunter, 1998), underscoring the importance of differentiating the validity of constructs from measures (Hough, 2001). Although traditionally self-ratings have been used to collect information about current or future employees' personality traits, personnel psychologists may consider using personal acquaintances' ratings to complement employees' self-views. However, collecting observer reports as a selection tool also raises pragmatic questions about their fakability, susceptibility to stereotypes, and necessity for soliciting raters with particular relationships with targets. We believe that observer ratings have produced correlations with performance that are tantalizing enough to call for answering these pragmatic questions.

CONCLUSIONS

The findings presented here reify previous findings of the predictive validity merit in collecting personality ratings from nonself sources. Especially for the personality trait of Conscientiousness, our findings indicate that the advantages of observers extend even to observers outside the work context. As such, personality research across domains of psychology may benefit from using raters with clearer lenses and from studying how self-raters' lenses become sullied. The depth of useful personality information available from observers presents a major frontier to the breadth of personality research and practice in psychology that have otherwise relied heavily on self-report measures.

REFERENCES

- Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1998). Self-other agreement: Does it really matter? *Personnel Psychology*, *51*, 577–598.
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*, 9–30.
- Baumeister, R. F., & Newman, L. S. (1994). Self-regulation of cognitive inference and decision processes. *Personality and Social Psychology Bulletin*, *20*, 3–19.
- Binnewies, C., Sonnentag, S., & Mojza, E. J. (2009). Daily performance at work: Feeling recovered in the morning as a predictor of day-level job performance. *Journal of Organizational Behavior*, *30*, 67–93.
- Carlson, E. N., & Furr, R. M. (2009). Evidence of differential meta-accuracy: People understand the different impressions they make. *Psychological Science*, *20*, 1033–1039.
- Chang, L., Connelly, B. S., & Geeza, A. A. (2010, April). *Trait and method effects in personality ratings: A meta-analytic approach*. Poster presented at the annual meeting for the Society for Industrial and Organizational Psychologists, Atlanta, GA.
- Chernyshenko, O. S., Stark, S., Woo, S. E., & Conz, G. (2008, April). *Openness to experience: Its facet structure, measurement, and validity*. Paper presented at the annual conference for the Society for Industrial and Organizational Psychology, San Francisco.
- 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.
- Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*, 1092–1122.
- Crowne, D., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349–354.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*, 880–896.
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*, 69–106.
- Edwards, J. R. (1994). The study of congruence in organizational behavior research: A critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, *58*, 51–100.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, *59*, 395–430.
- Fiedler, E. R., Oltmanns, T. F., & Turkheimer, E. (2004). Traits associated with personality disorders and adjustment to military life: Predictive validity of self and peer reports. *Military Medicine*, *169*, 207–211.

- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, **80**, 1011–1027.
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, **102**, 652–670.
- 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.
- Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology*, **64**, 479–490.
- Funder, D. C., & West, S. G. (1993). Consensus, self-other agreement, and accuracy in personality judgment: An introduction. *Journal of Personality*, **61**, 457–476.
- Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences*, **7**, 385–400.
- Furr, R. M. (2009). Profile analysis in person–situation integration. *Journal of Research in Personality*, **43**, 196–207.
- Geeza, A. A., Connelly, B. S., & Chang, L. (2010, April). A meta-analytic examination of consistency in observers' perspectives across contexts. In B. S. Connelly & L. Chang (Chairs), *Stop being so self-centered! Researching and applying personality via observer-reports*. Paper presented at the annual meeting for the Society for Industrial and Organizational Psychologists, Atlanta, GA.
- Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel-selection. *Personnel Psychology*, **18**, 135–164.
- Heller, D., Ferris, D. L., Brown, D., & Watson, D. (2009). The influence of work personality on job satisfaction: Incremental validity and mediation effects. *Journal of Personality*, **77**, 1051–1084.
- Heller, D., Watson, D., Komar, J., Min, J. A., & Perunovic, W. Q. E. (2007). Contextualized personality: Traditional and new assessment procedures. *Journal of Personality*, **75**, 1229–1254. doi: 10.1111/j.1467-6494.2007.00474.x
- Hirschfeld, R. R., Thomas, C. H., & McNatt, D. B. (2008). Implications of self-deception for self-reported intrinsic and extrinsic motivational dispositions and actual learning performance: A higher order structural model. *Educational and Psychological Measurement*, **68**, 154–173.
- Hogan, R. (1996). A socioanalytic interpretation of the five-factor model. In J. S. Wiggins (Ed.), *The five-factor model of personality* (pp. 163–179). New York: Guilford Press.
- Hogan, R., & Shelton, D. (1998). A socioanalytic perspective on job performance. *Human Performance*, **11**, 129–144.
- Hough, L. M. (2001). I/O owes its advances to personality. In B. W. Roberts & R. Hogan (Eds.), *Personality psychology in the workplace* (pp. 19–44). Washington, DC: American Psychological Association.
- Hunter, J. E., & Schmidt, F. L. (1994). Estimation of sampling error variance in the meta-analysis of correlations: Use of average correlation in the homogeneous case. *Journal of Applied Psychology*, **79**, 171–177.

- Kandler, C., Bleidorn, W., Riemann, R., Spinath, F. M., Thiel, W., & Angleitner, A. (2010). Sources of cumulative continuity in personality: A longitudinal multiple-rater twin study. *Journal of Personality and Social Psychology*, **98**, 995–1008.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, **8**, 265–280.
- Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, **43**, 23–34.
- 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.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, **108**, 480–498.
- Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review*, **111**, 94–110.
- Lievens, F., De Corte, W., & Schollaert, E. (2008). A closer look at the frame-of-reference effect in personality scale scores and validity. *Journal of Applied Psychology*, **93**, 268–279.
- Mabe, P., & West, S. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, **67**, 280–296.
- McCrae, R. R., & Costa, P. T. (1995). Trait explanations in personality psychology. *European Journal of Personality*, **9**, 231–252.
- Mischel, W. (1968). *Personality and Assessment*. New York: Wiley.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, **102**, 246–268.
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology*, **60**, 683–729.
- Mount, M. K., Barrick, M. R., Scullen, S. M., & Rounds, J. (2005). Higher-order dimensions of the Big Five personality traits and the big six vocational interest types. *Personnel Psychology*, **58**, 447–478.
- Mount, M. K., Barrick, M. R., & Strauss, J. (1994). Validity of observer ratings of the Big Five personality factors. *Journal of Applied Psychology*, **79**, 272–280.
- Oh, I. S., Wang, G., & Mount, M. K. (2011). Validity of observer ratings of the five factor model of personality traits: A meta-analysis. *Journal of Applied Psychology*, **96**, 762–773. doi: 10.1037/a0021832
- 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. 71–109). New York: Guilford Press.
- Ostendorf, F., & Angleitner, A. (1994, July). *Psychometric properties of the German translation of the NEO Personality Inventory (NEO-PI-R)*. Paper presented at the seventh conference for the European Association for Personality Psychology, Madrid, Spain.

- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, **46**, 598–609.
- Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology*, **74**, 1197–1208.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, **124**, 262–274.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame of reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, **80**, 607–620.
- Smith, T. W., Uchino, B. N., Berg, C. A., Florsheim, P., Pearce, G., Hawkins, M., et al. (2008). Associations of self-reports versus spouse ratings of negative affectivity, dominance, and affiliation with coronary artery disease: Where should we look and who should we ask when studying personality and health? *Health Psychology*, **27**, 676–684.
- 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.
- Staufenbiel, T., & Hartz, C. (2000). Organizational Citizenship Behavior: Entwicklung und erste Validierung eines Meßinstruments [Organizational Citizenship Behavior: Development and validation of a measurement instrument]. *Diagnostica*, **46**, 73–83.
- Swann, W. B., Johnson, R. E., & Bosson, J. K. (2009). Identity negotiation at work. In B. Staw & A. Brief (Eds.), *Research in organizational behavior* (Vol. 29, pp. 81–109). Amsterdam, The Netherlands: Elsevier.
- Vazire, S. (2010). Who knows what about a person? The Self-Other Knowledge Asymmetry (SOKA) Model. *Journal of Personality and Social Psychology*, **98**, 281–300.
- Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, **95**, 1202–1216.
- Viswesvaran, C., & Ones, D. S. (1995). Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology*, **48**, 865–885.
- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, **81**, 557–574.
- Wright, J. C., & Mischel, W. (1987). A conditional approach to dispositional constructs: The local predictability of social behavior. *Journal of Personality and Social Psychology*, **53**, 1159–1177.