A Narrower Scope or a Clearer Lens for Personality?

Examining Sources of Observers’ Advantages over Self-Reports

For Predicting Performance

Brian S. Connelly

University of Toronto

Ute R. Hülsheger

Maastricht 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) using 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 predicted 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.
A Narrower Scope or a Clearer Lens?

Examining the Validity of Personality Ratings from Observers Outside the Workplace

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., Mischel, 1968; Guion & Gottier, 1965), finding consensus across raters’ perceptions of personality traits validated self-report measures and bolstered the notion that traits are cross-situationally stable (Kenrick & 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 and 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 of 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 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 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

\(^1\) Based on operational validities described in Connelly & Ones (2010). Big Five meta-analytic intercorrelations were drawn for observer ratings from Chang, Connelly, & Geeza (2010) and for self-ratings from Mount, Barrick, Scullen, & Rounds (2005).
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 paper 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 Michel’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 intra-individual 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, Connelly, & Chang, 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 a respondent 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 suggests 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 self-reports being more accurate. However, Hogan and colleagues (R. Hogan, 1996; R. Hogan & Shelton, 1998) have 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) discusses 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) argue 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) argue 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) presents 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 $\rho_{ov}$), 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 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. 111 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. 33% 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.” Personal acquaintances knew targets on average for 14.8 years
($SD = 13.08$), reported to know targets very well (mean = 7.94, $SD = .99$ on a scale from 1 to 9),
reported having a good relationship with targets (mean = 8.42, $SD = .78$ on a scale from 1 to 9),
and met targets frequently in their spare time (mean = 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 years), knew targets well (mean = 6.55, SD = 1.32), and reported having a good relationship with targets (mean = 7.34, SD = 1.24). However, colleagues reported meeting targets only occasionally in their spare time (mean = 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, 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, Bleidorn, Riemann, Spinath, Thiel, & Angleitner, 2010). Cronbach’s α’s for the NEO-PI-R domain scales ranged between .84 and .95.

---

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.
Job performance criteria. Job performance criteria were assessed by colleagues’ ratings. Both, in-role behavior and organizational citizenship behavior (OCB) were assessed with scales provided by Staufenbiel and Hartz (2000). This questionnaire is a well-established set of scales that is widely used in German speaking 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 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, $\rho_{ov_{pacq}} = .46$ vs. $\rho_{ov_{self}} = .29$; for in-role behaviors, $\rho_{ov_{pacq}} = .52$ vs. $\rho_{ov_{self}} = .31$; for organizational citizenship behaviors, $\rho_{ov_{pacq}} = .26$ vs. $\rho_{ov_{self}} = .18$. 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

---

3 Interrater reliabilities were drawn from Viswesvaran, Ones, and Schmidt (1996) to make performance unreliability corrections. For correlations with overall job performance, Viswesvaran et al.’s reliability estimate for overall job performance ($r_{xx} = .52$) was used. For correlations with in-role behaviors, Viswesvaran et al.’s 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 ($r_{xx} = .53$).
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- and personal acquaintance-ratings in multiple regression compared to using either rating source individually. The incremental prediction ($\Delta R$) for self-ratings and personal acquaintance-ratings indicates the unique predictive power of each rating source, and the regression weights ($\beta$) 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 $\beta$ 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 $\beta$-weights indicate that both self-ratings and personal acquaintance-ratings contain unique information for predicting performance criteria. However, note that these $\beta$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). That is, over-estimates from self-ratings of Emotional Stability and of Agreeableness actually yield lower job performance. These effects indicate that mis-perceiving one’s own Emotional Stability or Agreeableness has negative implications for overall performance and 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.

---

4 Following procedures outlined by Edwards (1994), constraining the regression weights for self- and personal acquaintance-ratings to be equal and opposite produced minimal (non-significant) 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.
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).

For Conscientiousness, operational validities for personal acquaintances are larger than self-reports by a substantial margin ($\rho_{ov_{paca}} - \rho_{ov_{off}} = .26$). 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 ($\rho_{ov_{paca}} - \rho_{ov_{col}} = .18$). 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
towards producing stronger operational validities than personal acquaintances’ ratings  

\( \rho_{\text{ov-pacq}} - \rho_{\text{ov-ral}} = -.15 \) 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 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-acquaintances-, 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 under-identified), 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.

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 $\beta$-weights for the unique rater perspective factor relative to the $\beta$-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 and TLI values generally at or in excess of 1.0 and SRMR 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 $R_{\text{est}}$ values and the relative strength of $\beta$-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, 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 $\lambda_{\text{pacq}} = .63$ for Agreeableness to $\lambda_{\text{pacq}} = .95$ 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 $R_{\text{est}}$’s
and $\beta$s. 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 $\beta$ 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 traits.

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 $\beta$-weights for self-perspective uniquenesses were negative, and the $\beta$-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’ rating 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. In addition, although the effects were more marginal, the point estimates generally suggested that observers may generally have Clearer Lens of 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 beta-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 multi-rater 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 over-raters 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 towards 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 vs. self-effacement (e.g., Edwards, 1994; c.f., 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 non-self 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 has otherwise relied heavily on self-report measures.
References


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. doi:10.1037/0033-295X.102.2.246


validation of a measurement instrument]. *Diagnostica, 46*, 73-83. doi:10.1026/0012-1924.46.2.73.


## Table 1

**Means, Standard Deviations, Reliabilities and Intercorrelations Among Primary Study Variables**

|          | M    | SD   | 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 |     |     |     |     |     |     |     |
| 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; 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 are presented in gray and should be interpreted cautiously. Correlations whose absolute value is greater than r = .20 have 95% confidence intervals that do not include zero/p <.05.
Table 2: Regressions Combining Self + Acquaintance Ratings to Predict Performance Criteria

<table>
<thead>
<tr>
<th></th>
<th>Performance</th>
<th>In-Role</th>
<th>OCB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R )</td>
<td>( \Delta R )</td>
<td>( \beta )</td>
</tr>
<tr>
<td><strong>Emotional Stability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self</td>
<td>.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal Acquaintance</td>
<td>(.00, .27)</td>
<td>(-.37, -.05)</td>
<td>.05</td>
</tr>
<tr>
<td><strong>Extraversion</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self</td>
<td>.05</td>
<td>-.12</td>
<td></td>
</tr>
<tr>
<td>Personal Acquaintance</td>
<td>(.00, .27)</td>
<td>(-.37, .14)</td>
<td>.07</td>
</tr>
<tr>
<td><strong>Openness</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self</td>
<td>.00</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Personal Acquaintance</td>
<td>(.00, .20)</td>
<td>(-.22, .26)</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Agreeableness</strong></td>
<td>.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self</td>
<td>(.00, .38)</td>
<td>-.17</td>
<td></td>
</tr>
<tr>
<td>Personal Acquaintance</td>
<td>(-.38, .04)</td>
<td>(-.38, .04)</td>
<td>.13*</td>
</tr>
<tr>
<td><strong>Conscientiousness</strong></td>
<td>.33*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self</td>
<td>(.11, .48)</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>Personal Acquaintance</td>
<td>(.22, .56)</td>
<td>(-.16, .27)</td>
<td>.12*</td>
</tr>
<tr>
<td>Acquaintance</td>
<td>(.09, .51)</td>
<td>(.19, .58)</td>
<td>.16*</td>
</tr>
</tbody>
</table>
*p < .05.

Note. N = 97; 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.
Table 3:
Comparison of Personal Acquaintance Operational Validities in the Present Study to Meta-Analytic Estimates for Self- and Colleague-Reports

<table>
<thead>
<tr>
<th></th>
<th>Operational Validities for Job Performance</th>
<th>Differences in Validities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Personal Acq.</td>
<td>Self-Reports(^a)</td>
</tr>
<tr>
<td></td>
<td>(\rho_{ov}) (80% Cred.)</td>
<td>(\rho_{ov}) (80% Cred.)</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>.25</td>
<td>.11</td>
</tr>
<tr>
<td>Extraversion</td>
<td>.08</td>
<td>.11</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>.08</td>
<td>.04</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>.21</td>
<td>.11</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>.46</td>
<td>.20</td>
</tr>
</tbody>
</table>

\(\rho_{ov}\)’s presented in original meta-analyses. Italicized credibility intervals indicate credibility intervals that do not include the personal acquaintance point estimate of \(\rho_{ov}\). 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.

\(^a\)Based on values from Barrick et al. (2001), independent samples only.

\(^b\)Based on values from Connelly and Ones (2010), coworkers only.

Note: 80% credibility intervals calculated based on \(SD_{\rho_{ov}}\)’s presented in original meta-analyses.
<table>
<thead>
<tr>
<th>Trait</th>
<th>Model Fit</th>
<th>Trait Loadings (SE)</th>
<th>$R_{ext}$ (SE)</th>
<th>$\beta$ weights (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>df</td>
<td>CFI</td>
<td>TLI</td>
</tr>
<tr>
<td><strong>Emotional Stability (N = 264)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common only</td>
<td>2.40</td>
<td>5</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td>Colleague</td>
<td>.35</td>
<td>4</td>
<td>1.00</td>
<td>1.05</td>
</tr>
<tr>
<td>Self</td>
<td>1.50</td>
<td>4</td>
<td>1.00</td>
<td>1.04</td>
</tr>
<tr>
<td>P. Acquaintance</td>
<td>2.31</td>
<td>4</td>
<td>1.00</td>
<td>1.02</td>
</tr>
<tr>
<td><strong>Extraversion (N = 269)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common only</td>
<td>1.79</td>
<td>5</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td>Colleague</td>
<td>.44</td>
<td>4</td>
<td>1.00</td>
<td>1.04</td>
</tr>
<tr>
<td>Self</td>
<td>1.46</td>
<td>4</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td>P. Acquaintance</td>
<td>.14</td>
<td>4</td>
<td>1.00</td>
<td>1.04</td>
</tr>
<tr>
<td><strong>Openness (N = 266)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common only</td>
<td>12.64</td>
<td>5</td>
<td>.96</td>
<td>.93</td>
</tr>
<tr>
<td>Colleague</td>
<td>.11</td>
<td>4</td>
<td>1.00</td>
<td>1.05</td>
</tr>
<tr>
<td>Self</td>
<td>11.26</td>
<td>4</td>
<td>.97</td>
<td>.91</td>
</tr>
<tr>
<td>P. Acquaintance</td>
<td>10.55</td>
<td>4</td>
<td>.97</td>
<td>.92</td>
</tr>
<tr>
<td><strong>Agreeableness (N = 269)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common only</td>
<td>3.68</td>
<td>5</td>
<td>1.00</td>
<td>1.02</td>
</tr>
<tr>
<td>Colleague</td>
<td>2.66</td>
<td>4</td>
<td>1.00</td>
<td>1.02</td>
</tr>
<tr>
<td>Self</td>
<td>.66</td>
<td>4</td>
<td>1.00</td>
<td>1.06</td>
</tr>
<tr>
<td>P. Acquaintance</td>
<td>2.66</td>
<td>4</td>
<td>1.00</td>
<td>1.02</td>
</tr>
<tr>
<td><strong>Conscientiousness (N = 269)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common only</td>
<td>5.36</td>
<td>5</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Colleague</td>
<td>1.44</td>
<td>5</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td>Self</td>
<td>1.80</td>
<td>4</td>
<td>1.00</td>
<td>1.02</td>
</tr>
<tr>
<td>P. Acquaintance</td>
<td>5.36</td>
<td>4</td>
<td>1.00</td>
<td>.99</td>
</tr>
</tbody>
</table>

**Note.** N’s are based on the harmonic mean of total sample sizes for correlations in the meta-analytic intercorrelation matrix.
Figure 1. Base SEM Model (solid lines only) and Colleague SEM Model (solid and dashed lines).