CONVERGENCE OF MULTIPLE RATING SOURCES WITH OBJECTIVE MEASURES OF WORK PERFORMANCE: A META-ANALYSIS

BY

BERTHA RANGEL

THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Arts in Psychology in the Graduate College of the University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Adviser:

Professor Nichelle C. Carpenter
ABSTRACT

Given the common use of subjective-ratings (e.g., supervisor, peer or self ratings) for performance appraisal, the purpose of this study was to evaluate the extent to which subjective ratings converge with, and account for unique variance in, objective measures of work performance. It is important to determine the extent to which subjective measures (prone to rater biases) converge with measures that do not have the same vulnerabilities. Results demonstrated that peer-ratings had the highest convergence with objective measures ($\rho = .31$), self-ratings were the next highest ($\rho = .20$), while supervisor-ratings had the lowest convergence ($\rho = .16$), these correlations were statistically significantly different from each other. Regression analysis demonstrated that self-ratings accounted for significant variance in objective measures, beyond peer and supervisor-ratings. We also investigated possible moderators of the subjective-objective performance relationship. For strict measures of task performance, peer-ratings were found to have the greatest convergence with objective measures of task performance ($\rho = .34$). Moreover, the subjective-objective relationship was stronger when both the subjective-rating and the objective measure represented precisely the same construct and when both were at a similar level of specificity.
# TABLE OF CONTENTS

Introduction ......................................................................................................................... 1
Method ................................................................................................................................. 11
Results ................................................................................................................................. 17
Discussion .......................................................................................................................... 22
References ......................................................................................................................... 27
Tables ................................................................................................................................. 39
Introduction

Subjective ratings continue to be the most popular method of measuring workplace performance (Bernardin & Beatty, 1984; Murphy & Cleveland, 1995). Subjective ratings can be defined as judgments/evaluations obtained from supervisors, peers, subordinates, customers or the self (Cascio, 1991). The use of multisource feedback continues to be widespread (Morgeson, Mumford & Campion, 2005), wherein subjective ratings are collected from multiple sources, with supervisor ratings being the most common, followed by peer ratings, and also including the focal employee’s self-ratings (Cleveland, Murphy & Williams, 1989; Dalessio, 1998). Most of our understanding of employee performance and competence has been obtained via subjective ratings. Indeed, Atwater and Waldman (1998) reported that multisource assessments are used by 90% of Fortune 1000 companies.

The widespread use of subjective ratings has garnered interest in their validity, which has become the focus of numerous validity and psychometric studies (e.g. Conway & Huffcutt, 1997; Viswesvaran, Schmidt & Ones, 2002). Recent evidence has demonstrated the merit of subjective-ratings via moderate relationships with other subjective ratings of the construct. Specifically, meta-analyses have shown that self- and observer ratings are moderately correlated in measures of organizational citizenship behavior ($\rho = .26$; OCB; Carpenter, Berry, & Houston, 2014), counterproductive work behavior ($\rho = .38$; CWB; Berry, Carpenter & Barratt, 2012), and job performance ($\rho = .22$ with supervisors and $\rho = .19$ with peers; Conway & Huffcutt, 1997). Additionally, both Conway and Huffcutt ($\rho = .34$; 1997) and Viswesvaran et al. ($r = .46$; 2002) found moderate correlations between peer and supervisor ratings of work performance.

A number of assumptions about subjective ratings have resulted from these comparisons. Specifically, supervisor and peer ratings tend to be viewed as “good” rating sources, given that
the supervisor-peer interrater relationships tend to be large. In contrast, the use of self-ratings tends to be discouraged, because self-supervisor and self-peer interrater relationships are typically lower than supervisor-peer relationships (Conway & Huffcutt, 1997). Thus, the implicit assumption about the validity of subjective ratings is that supervisors and peers give more valid ratings than self-ratings.

Despite advances in understanding the validity of subjective ratings, one important limitation remains: the validity of subjective ratings is most typically demonstrated by comparing ratings from one source with other subjective rating sources. That is, our understanding of the merits of subjective ratings is based on how subjective ratings from one source are correlated with subjective ratings from another source. In the past, researchers have assumed that low convergence between raters suggests the presence of error, such that at least one of the raters is providing invalid information (Viswesvaran, Ones & Schmidt, 1996). However, this assumption neglects the possibility that each rating source may be tapping into different, yet equally valid, information (Mount, Judge, Scullen, Sytsma & Hezlett, 1998). Hence, low convergence with another rating source does not necessarily discount the validity of a given source (Borman, 1997; LeBreton, Burgess, Kaiser, Atchley, & James, 2003). In fact, advocates of multisource feedback argue that low convergence is desired, as it suggests that each rating source offers unique and valid information about the target employee (Borman, 1997; Craig and Hannum, 2006). If comparisons among multiple sources of subjective ratings do not necessarily allow for conclusions regarding these ratings’ validity, an alternate approach is needed to assess the validity of subjective ratings.

The current study demonstrates that an important means of evaluating the extent to which subjective ratings convey valid information is to understand the relationship between subjective
ratings and objective measures of work phenomena. Objective measures reflect non-judgmental (i.e., non-subjective) reports of work performance and typically comprise a quantitative count of the results of work, such as number of disciplinary actions, production output, or time to complete a job. Objective measures have the assumed advantage of being less contaminated by bias and errors resulting from human judgment (Landy & Farr, 1983). If each rating source converges with an objective measure, then this indicates that each source contains at least some valid information about the target employee. Further, the extent to which each rating source converges with an objective performance measure can be compared, across subjective rating sources, to evaluate whether each rating source accounts for unique variance in objective measures, over and above the other sources. This will inform us of the level of convergence between independent rating sources and objective measures and allow us to directly test the assumption made by multisource feedback systems—that multiple sources provide valid and unique performance-relevant information about the target employee (Lance, Baxter, & Mahan, 2006; Morgeson et al., 2005).

The purpose of the current study is to investigate the convergence between subjective ratings and objective measures of performance and to examine the incremental validity of different subjective rating sources of performance for predicting objective performance measures. To this end, we first meta-analytically examine the correlations between subjective ratings and objective measures of performance to determine the extent to which these two measures represent overlapping information. We then use the results from these subjective-objective meta-analyses to compare the extent to which each subjective rating source accounts for unique variability in objective measures, relative to other rating sources. Next, we test whether the type of criterion measure (e.g., CWB vs. task performance) moderates the
relationship between subjective ratings and objective measures of performance. Finally, we address a critical limitation of prior research by examining whether the subjective rating and objective measure tap matching constructs. Overall, this study contributes to understanding the construct validity of subjective ratings of performance by examining the extent to which subjective ratings actually overlap with and account for unique variance in objective measures of performance.

**Convergence between Subjective Ratings and Objective Measures**

Researchers have argued that the discrepancy between subjective ratings results from measurement source/rater biases (Kenny & Berman, 1980; Viswesvaran et al., 1996). This perspective discourages the use of self-ratings for performance measures as it assumes that employees cannot rate themselves objectively (Fox, Spector, Goh, & Bruursema, 2007; Levine, Flory & Ash, 1977). Self-reports are considered to be affected by biases particularly due to self-enhancement motives, whereas observer ratings are not considered to be contaminated by these biases (DeNisi & Shaw, 1977). Therefore, this argument proposes that supervisor and peer ratings contain less rater bias than do self-ratings. The greater observed correlation between peer- and supervisor ratings than between self-observer ratings can be interpreted as resulting from these self-enhancement biases. Objective measures are also not affected by these rater source biases (Landy & Farr, 1983). For this reason we expect to see lower convergence of self-ratings with objective measures of work performance than supervisor and peer-ratings convergence with objective measures.

*Hypothesis 1:* The relationship of objective measures with (a) peer ratings and (b) supervisor ratings will be greater than with self-ratings.
Incremental Validity of Rating Sources

The ecological perspective can be used to understand whether each rating source is uniquely and incrementally predictive of performance outcomes. This perspective argues that each rating source provides different but valid information (Gibson, 1979; Kavanagh, Borman, Hedge & Gould, 1986). If this is the case then we would expect different rating sources to contribute incremental validity. Multisource feedback systems support this perspective as they are anchored in the belief that ratings from different sources reflect diverse and valid information about the target employee’s performance, in which each rating source captures different parts of the total criterion space (Lance et al., 2006).

Derived by social psychologists in the area of social judgment accuracy (Gibson, 1979), the ecological perspective states that information presented to raters represents affordances that the rater uses to act upon the environment (Beauvois & Dubois, 2000), implying that raters pay the most attention to information that is relevant to their interactional goals. Therefore, raters who share similar affordances (e.g., individuals in the same level of the organization) will focus their attention on the same or similar events in their environment. Conway and Huffcutt’s (1997) findings support this perspective, demonstrating that within-source (e.g., supervisor-supervisor) agreement is higher than between source (e.g., supervisor-peer) agreement. These findings suggest that individuals from the same level (or same role) in the organization tend to have greater amounts of converging information than individuals from different levels in the organization.

Moreover, this perspective argues that each rater focuses on different aspects of employee behavior because raters’ perceptions of the target employee are guided by different goals, and these goals influence the type of information the raters seek out (Gibson, 1979). For
instance, a supervisor with the goal of being seen as an effective leader will attend more closely to which employees complete tasks and follow rules and instructions (e.g. Oh & Berry, 2009).

As another example, a coworker with the goal of getting along with others will seek information concerning whether the target employee is a “team player.” In this framework, we expect to see differences between raters because the target employee represents a different set of affordances to each rating source. As raters in varying organizational roles may be attuned to different affordances, a variety of interaction goals may be set, depending on the rater. Additionally, this perspective states that the target employee may interact differently with each of the raters, as they too have different interaction goals with each of the raters (Baron & Boudreau, 1987).

In sum, the ecological perspective suggests that self-ratings tap into different information than peer and supervisor ratings (Gibson, 1979). These distinctive perspectives result in each subjective rating source likely accounting for unique variance in objective measures, beyond the other sources. For this study we specifically focus on the incremental variance of self-ratings beyond supervisor and peer ratings, because supervisor and peer ratings are already popular ways of measuring performance (Cleveland, Murphy & Williams, 1989; Murphy & Cleveland, 1995).

Based on the ecological perspective we hypothesize the following:

**Hypothesis 2**: Self-ratings account for incremental variance in objective measures of (a) overall performance, (b) task performance, and (c) CWB and withdrawal, above and beyond supervisor- and peer-ratings.

**Type of Objective Measure as Moderator**

Potential moderators of the relationship between self-ratings and objective measures of performance also merit consideration. One variable that may moderate the subjective-objective relationship is *task acquaintance*, which refers to the degree to which the rater has the
opportunity to observe the target employee engaging in workplace behaviors (Kingstrom & Mainstone, 1985). For instance, peers often have the opportunity to observe different aspects of task performance than supervisors, as peers tend to spend more time in close proximity with the target employee (Borman, 1974). Furthermore, Murphy and Cleveland (1995 p.134; Murphy 1989a) theorize that access to information relating to work performance varies among rating sources. Specifically, self-raters tend to have the most access to information for both task and interpersonal behaviors relative to peers and supervisors, whereas peers tend to have more access to information for interpersonal behaviors, but approximately the same access to information for task performance relative to supervisors. This indicates that the convergence between subjective ratings and objective measures may depend on the performance dimension being measured. In the current study, we focus on OCB, CWB (withdrawal) and task performance.

First, we examine the subjective-objective relationships for OCB and CWB. OCB is defined as extra-role behaviors that go beyond the formal job requirements (Organ, 1988) and CWB is defined as voluntary behaviors that violate organizational norms (Robinson & Bennett, 1995). Compared to supervisors, peers have two advantages for rating these types of behaviors. First, peers work more closely with the target employee and may have more interpersonal interaction with the employee than do supervisors (Borman, 1997; Murphy & Cleveland, 1995). Second, employees are more likely to alter their behavior when they know their supervisor is watching (Baron & Boudreau, 1987). This may be especially true for CWB and OCB because the employee is more likely to hide the negative workplace behavior from the supervisor and more likely to purposely engage in OCB when the supervisor is present.

Similarly, self-ratings should have greater convergence with objective measures of CWB than do supervisor-ratings, because self-raters have more knowledge concerning their enactment
of these often-hidden behaviors. Recently, Berry et al. (2012) demonstrated that self- and observer reports of CWB were moderately correlated at $\rho = .38$. Moreover, this meta-analysis found that employees admit to engaging in more CWB than observers report, indicating that self-reports of CWB may be less downwardly biased than previously assumed. For example, an employee will tend to refrain from stealing office supplies in front of a supervisor but may engage in this behavior when no one else is around. As a result, the only person with access to this information is the target employee. This leads us to our next hypotheses:

**Hypothesis 3:** The convergence between subjective ratings and objective measures of (a) CWB and (b) OCB are greater for peer ratings than supervisor ratings.

**Hypothesis 4:** The convergence between subjective ratings and objective measures of CWB is greater for self-ratings than for (a) supervisor ratings and (b) peer-ratings.

Second, we focus on task performance, defined as the effectiveness with which an employee performs the activities that directly contribute to the organization’s technical core (Borman & Motowidlo, 1993). Murphy and Cleveland (1995) would expect that peers and supervisors have approximately the same access to information regarding task performance outcomes. While peer raters may have more opportunity to observe the employee engaging in task performance behaviors, supervisors may have access to different types of information not available to peers, such as goal attainment or organizational records of productivity, disciplinary actions, past performance reviews, etc. The fact the supervisors have access to objective metrics of the target employee’s performance should result in greater convergence with objective measures of task performance compared with peer ratings.

**Hypothesis 5:** The convergence between subjective ratings and objective measures of task performance is greater for supervisor ratings than for peer ratings.
Construct Precision and Compatibility Principle

We will examine construct match as a moderator, which is whether the subjective ratings and objective measures of interest represent precisely the same performance dimension. There is an implicit assumption that correlations calculated in Bommer et al. (1995) and Conway et al. (2001) represent the relationship between subjective ratings and objective measures of the same performance dimension (e.g., the correlation between subjective ratings of task performance and an objective measure of task performance), but this may not be the case. When evaluating the convergence between subjective ratings and objective measures, it is potentially important to note whether the construct is held constant. The present study addresses this critical issue by comparing the subjective-objective relationship when the construct is held constant versus when it is not.

Research should take into consideration the (lack of) construct match between subjective ratings and objective measures, because if the two represent distinct constructs, contamination can result due to construct-irrelevant variance. In other words, when the subjective ratings and objective measures represent different constructs, the lack of construct match will likely attenuate the subjective-objective relationship. The low convergence between subjective ratings and objective measures may thus be a result of the construct-irrelevant contamination that comes from correlating two distinct constructs, rather than a consequence of the rating sources.

Hypothesis 6: The construct match between the subjective rating and the objective measure moderates the subjective-objective performance relationship, such that the relationship is stronger for matched constructs.

Another aspect of measurement that may affect the subjective-objective correlation is whether the constructs are similar in breadth, defined as the specificity or generality of the
constructs being measured. When subjective ratings and objective measures are dissimilar in breadth, an effect on covariation could result from construct under-representation. Construct-underrepresentation occurs when a measure does not sample from the entire domain of behaviors that encompass that construct (Messick, 1989). For subjective and objective job performance measures, the subjective rating tends to be broader than the objective measure. For example, Motowildo (1982) collected self-ratings of overall performance and correlated them with the total dollar value of sales made over 11 months. Although total dollar sales represents an aspect of overall job performance, it does not encompass the entire broad construct domain. If the objective measure is more precise (narrow) than the subjective rating, the objective measure will not take into account all of the dimensions that are represented in the subjective rating, which will result in attenuation of the subjective-objective correlation.

In related research on attitudes, Ajzen and Fishbein (1977) demonstrated that the relationship between attitudes and behaviors was strongest when the specificity or generality of the attitude matched that of the behavior, a phenomenon referred to as the compatibility principle. Harrison, Newman and Roth (2006) argued that specific work attitudes (e.g. attitude towards lateness) were better at predicting specific work behaviors (e.g. coming late to work) than were general work attitudes (e.g. overall job satisfaction) and vice versa. This supports the notion that we will find greater overlap between constructs measured at the same level of abstraction—i.e. those similar in breadth.

**Hypothesis 7**: The *breadth* match between subjective evaluations and the objective measures moderates the subjective-objective performance relationship, such that the relationship is stronger for measures that are matched in construct *breadth*. 
Method

Literature Search

To locate primary studies, we first consulted the reference sections of previous meta-analyses that included subjective-objective relationships (i.e. Bommer et al., 1995; Conway et al., 2001; and Mabe & West, 1982). Next, we used PsycINFO, Dissertation Abstracts International, and Google Scholar databases to search for published and unpublished articles through 2014. Example search terms included self-ratings, self-rated, self-evaluation, self-assessment, supervisor-ratings, supervisor-rater, supervisor-evaluation, supervisor-assessment, multisource ratings, job performance, work performance, objective performance, objective measures, and externally measured performance. Lastly, we carried out a manual search of the 2009-2014 conference programs for Society of Industrial and Organizational Psychology annual meetings.

Inclusion Criteria and Procedure

Studies were deemed eligible for inclusion if they contained either (a) a correlation between a subjective rating (i.e. self, peer or supervisor) and an objective measure of CWB, OCB, withdrawal or task performance, or (b) adequate information to extract an effect size in the absence of a reported correlation. These inclusion criteria resulted in 92 studies, comprising 71 supervisor-objective independent samples, 49 self-objective independent samples, and 23 peer-objective independent samples. Of these studies, 10 were unpublished and 82 were published.

For each sample, we coded the correlation between a subjective rating and an objective measure, along with the source of the subjective rating (i.e. peer, self or supervisor). Specifically, each study had to provide a self, peer, or supervisor-rating of CWB, OCB, withdrawal or task performance (this constituted the subjective rating). This subjective rating had to be correlated
with a *quantitative count* of OCB (e.g., amount of overtime), CWB (e.g., number of demerits), withdrawal (e.g. days absent) or task performance (e.g. sales volume)—this constituted the objective measure. All three authors coded the same 11 studies (10.3%) in order to ensure reliability and accuracy of the coding. The authors met to discuss minor discrepancies in the coding and once total agreement was reached, the remainder of the studies were divided between two of the coders.

Next, we coded possible moderator variables. First, the type of objective measure was coded. Each objective measure was coded as one of four types: CWB, OCB, withdrawal or task performance. Measures were coded as *CWB* if they contained counts of negative work behaviors. Examples of measures coded as CWB are number of employee complaints, number of days late, or number of disciplinary actions. Importantly, we decided to collapse withdrawal behaviors (i.e. lateness and absences) into the CWB category, as commonly done in the CWB literature (Berry & Carpenter, 2014). Affirming this decision, the subjective-objective correlations for CWB and withdrawal were not statistically different from each other (self-objective $z = 1.24$; supervisor-objective $z = -1.01$; n.s.). Measures of positive work behaviors that are distinct from work tasks were coded as *OCB*. Only two studies in our meta-analysis fit this category, and both measured amount of overtime worked. Lastly, we categorized a measure as *task performance* if it measured activities that are formally recognized as part of the job. Examples of measures categorized as performance are total sales in a month, score on a work sample test, or number of publications.

Second, each correlation was coded for construct match between the subjective and objective measures, such that studies with “no construct match” were coded as 1, “some construct match” coded as 2, and “complete construct match” coded as 3. For example, if a study provided supervisor ratings of an employee’s mechanical ability (subjective rating) along
with the employee’s number of days late in the last month (objective measure), we would code this as a 1, “no construct match.” This is because the subjective rating and objective measure represent two distinct constructs (ability vs. CWB). In the same example, if the subjective measure was supervisor ratings of CWB, it would be coded as 2, “some construct match.” While number of days late represents a type of CWB, it does not encompass the entire construct domain. In the same example, if the subjective rating were a supervisor’s estimate of the number of days they believed the employee was late in the previous month, it would be coded as 3, “complete construct match”. Here, both the subjective rating and objective measure represent precisely the same construct.

Third, we coded breadth match, indicating whether the subjective and objective measures were similar in breadth, or alternatively whether the objective measure was more precise. For example, if a study collected peer-ratings of OCB along with the target employee’s number of absences in the past week, this would represent an example of the objective measure’s being more precise. Here, the OCB measure is more general; it asks about a variety of positive workplace behaviors, whereas the objective measure only gathers data concerning one specific negative workplace behavior. On the other hand, if the peer were specifically asked to rate how much overtime the target employee worked during the past week, this would be coded as “same breadth” because both the subjective rating and the objective measure are asking about specific workplace behaviors.

In order for a relationship to be meta-analyzed there needed to be at least three independent samples. We were unable to attain three independent samples containing subjective-objective correlations of OCB, and we did not have 3 independent samples with peer-objective correlations for CWB. For this reason we were unable to test Hypothesis 3ab.
Linear composites

To avoid violating the assumption of independence, when a sample provided several correlations with multiple dimensions of performance we calculated a linear composite to estimate the relationship with overall performance. For example, if a study used two objective measures of performance for a bank teller, such as productivity (i.e. the number of transactions a teller processed in a month divided by the number of hours worked) and dollar shortages (i.e. the amount of money unaccounted for by financial transactions per month), the correlations were combined to create a composite of overall performance. Ghiselli, Campbell, and Zedeck’s (1981) theory of composites was used to aggregate the correlations. Using these composites ensured that each sample was only included once in each given meta-analysis. When not enough information was given to calculate a composite, the mean of the correlations was used.

Meta-Analysis

Hunter and Schmidt’s (2004; Schmidt & Hunter, 2014) artifact distribution meta-analysis procedures were used, and unreliability of the subjective-rating was corrected using internal consistency reliability for subjective ratings of performance. Artifact distributions were created using the data collected for this study. The alpha reliabilities obtained were .78 ($k = 16, N = 3,473$) for self-ratings, .80 ($k = 5, N = 801$) for peer ratings and .80 ($k = 27, N = 4,254$) for supervisor ratings. We assumed that objective measures were free of measurement error. Raju and Brand’s (2003) formulas were used to assess the statistical significance of the differences between the moderator categories.

Incremental Validity Analyses

In order to test the incremental validity of each rating source against objective measures of performance, we created a meta-analytic correlation matrix. This matrix can be found in
Table 5 and includes objective measures as well as supervisor, peer and self-ratings. There are a total of six correlations amongst these variables, but only the subjective-objective (i.e. self-objective, peer-objective and supervisor-objective) meta-analytic correlations were original to our current study. Similar to the method used in Conway et al. (2001), we used the meta-analytic correlations for subjective-subjective ratings (i.e. self-supervisor, self-peer and supervisor-peer) that had been reported from previous meta-analyses.

Specifically, the correlation between self-supervisor ratings of combined performance measures (we refer to these as overall in our results) was taken from Heidemeier and Moser (2009; \( r = .24, k = 115, N = 37,752 \)), the self-peer correlation came from Huffcutt (2009; \( r = .19, k = 17, N = 6,359 \)) and the peer-supervisor correlation came from Viswesvaran, Schmidt and Ones (2002; \( r = .41, k = 31, N = 6,252 \)). One issue with this method is that each value in our correlation matrix has a different corresponding \( N \). We followed recommendations from Viswesvaran and Ones (1995) and used the harmonic mean of the sample sizes found in the meta-analytic correlation matrix (Table 5, harmonic mean \( N = 7,825 \)) as the basis for our multiple regression analyses.

Next, we created the same matrix but this time only for the studies that had objective measures of task performance (Table 7). For this correlation matrix the self-supervisor correlation came from Heidemeier and Moser (2009; \( r = .20, k = 67 \)), the self-peer correlation came from Conway and Huffcutt (1997; \( r = .19, k = 17, N = 6,359 \)), and the peer-supervisor correlation came from Viswesvaran et al. (2002; \( r = .39, k = 13, N = 2,481 \)). For the self-peer task performance meta-analytic correlation we combined the job performance dimensions of productivity and quality from Viswesvaran et al.’s (2002) meta-analysis. The Heidemeier and Moser (2009) meta-analysis did not provide the \( N \) for the task performance self-supervisor
correlation. Therefore, we extrapolated an $N$ by dividing the overall meta-analytic $N = 37,752$ by the overall meta-analytic $k = 115$ and multiplying that value by the task performance $k = 67$. This resulted in an $N$ of 21,889. The overall harmonic mean $N$ for this analysis was 5,474.

Similarly, Table 9 contains the meta-analytic correlation matrix for incremental validity analysis of CWB measures. This matrix contains three meta-analytic correlations. Two of the meta-analytic correlations (self-objective and supervisor-objective) were computed in this study, and we used Berry et al.’s (2012) meta-analytic estimate of the relationship between self-ratings and supervisor ratings. Berry et al. found the mean uncorrected correlation between self- and supervisor ratings of CWB to be .31, based on 11 independent samples and a total $N$ of 2,044.

Lastly, Table 11 has the meta-analytic correlation matrix that was used as input for incremental validity analysis for those studies with complete construct match. The matrix contains six meta-analytic correlations, in which the subjective-objective correlations were original and were calculated in the current study, and the subjective-subjective correlations came from previous meta-analyses. The “borrowed” correlations for this matrix are the same as the correlations used for the overall performance analysis (see Table 11). The harmonic mean for this analysis was $N = 2,885$. 
Results

Overall Relationship between Subjective Ratings and Objective Measures

Our meta-analytic results are presented in Table 1. We found a moderate correlation between self-ratings and objective measures. Specifically, the uncorrected correlation was .18, and after correcting for unreliability of the subjective measures, the correlation was .20. The 95% confidence interval for this correlation did not include zero. We also examined supervisor-objective and peer-objective meta-analytic relationships, and we compared these relationships with the self-objective meta-analytic correlation. We found a moderate correlation between supervisor ratings and objective measures ($\rho = .16$) and a stronger correlation between peer ratings and objective measures ($\rho = .31$); these correlations were statistically different from each other ($z = 8.97$, $p < .05$). Similarly, the self-objective relationship was statistically different from the overall supervisor-objective relationship ($z = 2.77$, $p < .05$), and the overall peer-objective relationship ($z = -5.58$, $p < .05$), demonstrating partial support for Hypothesis 1.

These results demonstrate that self-ratings have greater overlap with objective measures than do supervisor ratings, but peer ratings have a greater overlap with objective measures than both self and supervisor ratings.

Incremental Validity Results

In our next step, we tested whether self-ratings accounted for incremental variance in objective measures of workplace performance above other rating sources. Multiple regression analyses were used to answer this question. The meta-analytic correlation matrix used for these overall performance analyses can be found in Table 5, and the results for the regression analyses are located in Table 6. For model A we first entered supervisor-ratings into the equation and the $R^2$ was .020. When peer-ratings were entered to the equation the $R^2$ increased by .065 to .084.
Adding self-ratings as a third rating source resulted in a .016 increase to $R^2 = .100$ for all three rating sources together. Next, for Model B, we entered peer-ratings first and supervisors second, which resulted in $\Delta R^2$ of .001. Adding self-ratings along with supervisor and peer resulted in a $\Delta R^2$ of .016 to a total $R^2$ of .100 for all three sources. In sum, these findings support Hypothesis 2a that self-ratings account for unique variance in objective measures of performance, above supervisor and peer-ratings.

Next, we tested whether self-ratings accounted for a unique source of variance in objective measures of task performance above supervisor and peer-ratings. For model A (Table 8) we first entered supervisor ratings followed by peer ratings and self-ratings. When supervisor-ratings were entered alone, the $R^2$ was .044. Once peer-ratings were added the $R^2$ increased by .056 to .100. Adding self-ratings as a third source resulted in a .020 increase to .120 for all three sources together. Next, when we entered peer-ratings first (Model B) and then added supervisor-ratings, the $\Delta R^2$ was .010. Adding self-ratings along with supervisor and peer resulted in a $\Delta R^2$ of .020 to a total $R^2$ of .120 for all three sources. These findings show some support for our Hypothesis 2b. Specifically, they demonstrate that self-ratings account for unique variance in objective measures of task performance above the other rating sources. The meta-analytic correlation matrix for these results is located in Table 7, and the regression results can be found in Table 8.

We then tested whether each rating source contributed incremental validity for explaining objective measures of CWB. When supervisor-ratings were entered into the regression equation in the first step and self-ratings were entered in the second step, the $\Delta R^2$ was .007, partially supporting Hypothesis 2c. These findings can be found in Table 10.
Results for Moderator Analyses

Importantly, for both the self-objective and supervisor-objective correlations, the credibility intervals included zero; and for all three rating sources the credibility intervals were very wide (see Tables 2, 3, & 4), indicating there may be substantive moderators of the subjective-objective relationships. Next, we tested the extent to which several variables moderated the subjective-objective relationship. Results for the moderator analyses can be found in Table 2 for self-objective relationships, Table 3 for supervisor-objective relationships, and Table 4 for peer-objective relationships.

We expected the convergence between subjective ratings and objective measures of CWB would be larger for self-ratings than for supervisor-ratings. Results demonstrate that the meta-analytic correlation between self-ratings and objective measures of CWB was .04, whereas the supervisor-objective meta-analytic correlation was -.15, and these correlations were statistically different from each other ($z = 4.95, p < .05$). This finding demonstrates that the magnitude of the relationship between supervisor-objective measures is greater than that of self-objective measures, but—interestingly—the negative correlation signifies that as objective measures report greater enactment of CWB, supervisors report decreasing levels of CWB.

We posited that the supervisor-objective relationship would be greater than the peer-objective relationship for measures of task performance. Contrary to our Hypothesis 5, we found the peer-objective correlation ($\rho = .34$) was significantly larger than the supervisor-objective correlation ($\rho = .23$) for measures of task performance ($z = 5.66, p < .05$). This result signifies peer-ratings have a greater overlap with objective measures of task performance than do supervisor-ratings.
Next, we posited that the construct match between subjective ratings and objective measures (i.e. subjective rating of CWB and objective measure of CWB) moderates the subjective-objective relationship, such that the relationship gets stronger as the construct match increases. Consistent with expectations, the findings showed the relationship for supervisor-ratings and objective measures was -.04 for no match, .18 for some match and .35 for complete construct match. For the self-objective relationships the correlations were .07 for some construct match and .43 for complete construct match. Finally, the peer-objective relationships were .02 for no construct match, .27 for some construct match and .50 for complete construct match. These correlations were statistically different from each other within each rating source, supporting Hypothesis 6. Regardless of rating source, the objective-subjective relationship got stronger as the match between the constructs increased. These findings suggest that there is higher convergence between subjective and objective measures of performance when the two are measuring the same construct, and the convergence decreases when the match decreases.

Finally, we tested whether breadth match (general vs. specific) moderated the subjective-objective relationship. We found that the subjective-objective relationship was stronger when the breadth of the measures was similar than when the objective measure was more narrow than the subjective rating. Specifically, the self-objective correlation increased from .10 when the objective measure was more narrow to .30 when they both were similar in breadth ($z = 7.15, p < .05$). Similarly, the correlation increased from .21 to .38 for peer ratings ($z = 5.97, p < .05$) and .12 to .19 for supervisor ratings ($z = 3.49, p < .05$), supporting Hypothesis 7. In sum, we found that matching the specificity or generality of the subjective measure to the specificity or generality of the objective measure makes the relationship stronger.
Given the results of the construct match analyses, we decided to test the incremental validity of self-ratings for predicting objective measures, but this time using exclusively those primary studies with complete construct match. When supervisor ratings were entered into the equation first, the $R^2$ was .096. Next when we entered the peer-ratings the $R^2$ increased by .125 to .221. In the final step, when the self-ratings were entered the $\Delta R^2$ was .078 for a total $R^2$ of .299 for the three rating sources together. These results can be found in Table 12. This signifies that each of the rating sources accounts for unique variance in objective measures, when the subjective ratings and objective measures represent precisely the same construct.
Discussion

The current study examined the relationship between subjective ratings and objective measures of performance. First, we found the overall self-objective correlation to be .20, the supervisor-objective correlation to be .16, and the peer-objective correlation to be .31. The performance dimension being measured moderated this relationship, such that self-ratings had the most convergence with objective measures of ability, and peer ratings had the most convergence with objective measures of task performance. Next, both construct match and breadth match were found to moderate the convergence between subjective ratings and objective measures, such that the convergence was greater when the match was high. Importantly, we found that self-ratings accounted for unique variance in objective measures, above and beyond supervisor and peer ratings. This incremental variance of self-ratings was even greater when we only included studies in which the subjective ratings and objective measures represented precisely the same construct.

Some of the most important findings in the current study were those from the regression analysis. These analyses demonstrated that each rating source accounts for incremental variance in objective measures, indicating that each source offers unique and valid information on ratings of work performance. The findings are important because they support the idea that using multiple operations is generally better than only using one method of measurement. The sole use of supervisor ratings is pervasive in HR research, but practitioners should be encouraged to consider using more than one rating source. These findings support the continued use of multiple rating sources. In the future, practitioners and researchers should focus on developing ways to assess potential differences in the content of the valid portion that comes from each rating source.
Additionally, the current study updated the peer-objective and supervisor-objective meta-analytic summaries. The peer-objective correlation found in this study ($\rho = 0.31, k = 23$) was consistent with the relationship found in Conway et al’s ($\rho = 0.29, k = 9; 2001$) previous meta-analysis. However, the current study found that the observed correlation between supervisor ratings and objective measures is smaller than previously found. Bommer et al. (1995) found a .39 correlation using 50 independent samples, whereas the current study found a .16 correlation using 71 independent samples.

Moreover, the current work addressed some significant potential limitations found in Bommer et al. (1995) and Conway et al. (2001) meta-analyses. Bommer et al’s meta-analysis has two important limitations: 1) only one source of subjective ratings was used 2) all dimensions of work performance were collapsed to overall performance. Conway, Lombardo and Sanders (2001) did include multiple sources in their subjective-objective meta-analyses (subordinate and peer), but once again examined only overall performance. Notably, neither of these previous meta-analyses included self-ratings. Past research has noted that the target employee is likely to hold unique information about his/her work behavior that is not available to other ratings sources (Berry et al., 2012; DeNisi, Cafferty & Meglino, 1984) and for this reason they can provide a unique source a variance beyond supervisor and peer ratings. The current meta-analysis addressed these issues by including multiple ratings sources (containing self-ratings), and by analyzing the relationships of different performance dimensions in addition to overall performance.

Furthermore, the current study addressed an important issue that was not adequately addressed in previous meta-analyses—the issue of construct match. Past research did not take into consideration whether the subjective rating and the objective measure were tapping into the
same performance dimension. This relationship is closer to the *true* subjective-objective relationship because the measurement error that results from correlating distinct constructs was removed. This study found that when we only included the studies in which the subjective-rating and objective measure represented precisely the same performance dimension, the convergence was significantly greater than when construct match was inconsistent. Importantly, when these correlations were used for the regression analysis we found that each rating source accounted for greater unique variance in objective measures, compared to the analysis conducted when construct match was not taken into account. Further supporting the use of multiple rating sources for performance appraisal, for each rating source provides unique and valid information about the target employee.

The results in this study support Bommer et al.’s (1995) conclusion that subjective ratings and objective measures are not interchangeable. We did not find very high convergence between the subjective ratings and objective measures. These results should be interpreted cautiously, though, as these moderate relationships do not necessarily indicate that subjective ratings are low in validity. Both the subjective ratings and objective measures are imperfect ways of measuring performance, each including unique problems with contamination and deficiency. Consequently, these imperfections in the measurement of performance will further reduce the convergence between subjective ratings and objective measures. The moderate convergence between the objective criteria and subjective ratings provides evidence for the potential unique validity of each rating source.

Furthermore, our study provides additional evidence in support of the continued use of self-ratings in the workplace. Specifically, by demonstrating that *self-ratings have greater convergence with objective criteria than do supervisor ratings*, this study demonstrates that self-
ratings should not be regarded as inherently flawed. In summary, the findings of this study demonstrate that self-ratings of withdrawal, CWB, and task performance overlap with, and account for unique variance in, objective measures; representing important evidence of the construct validity of self-ratings.

Our study has several important managerial implications. First, managers should not refrain from using self-ratings out of concern that they do not reflect the same amount of information as supervisor ratings. Second, we demonstrate that the type of outcome measure used can affect the convergence of self-ratings with objective measures, meaning that managers should use care when deciding which source to use when rating withdrawal, CWB, or task performance. Third, our findings support the use of multiple raters for evaluating work performance, as each rating source provides unique information about the target employee.

Despite this study’s contributions, some limitations should be noted. First, we did not take into account some possible situational variables, such as job complexity. When jobs are more complex it may be even more difficult to grasp one’s level of performance through an objective measure. As such, jobs with high job complexity may have lower subjective-objective relationships. Another important moderator variable that should be taken into account is whether the target employee chose which peer would complete the peer-ratings or if the peer-rater was appointed by a third party. If the target employee chose the peer rater it is likely that they chose someone whom they interact with the most, and this in turn could be why we see a greater convergence between peer-ratings and objective measures.

Furthermore, we were unable to test all the subjective-objective relationships that we set out to test. OCB is an important part of the criterion space but we were unable to obtain enough studies to analyze these subjective-objective relationships for it. We also did not locate enough
studies to calculate the peer-objective correlation for CWB. It is important for future research to focus on these relationships, as doing so would allow for a richer comparison of the convergence of multiple rating sources with objective measures across dimensions of performance. In conclusion, previous studies have shown that there is disagreement between sources when rating work performance, and the current study provides evidence indicating that at least some of the unique variance from each source represents valid information about the target employee.
References

* indicates the study was included in the meta-analysis


Table 1

Correlations between Subjective and Objective Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Supervisor</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>$k$</td>
<td>$r_m$</td>
<td>$\rho$</td>
<td>$N$</td>
<td>$k$</td>
<td>$r_m$</td>
<td>$\rho$</td>
<td>$N$</td>
<td>$k$</td>
</tr>
<tr>
<td>Overall</td>
<td>13,979</td>
<td>71</td>
<td>0.14</td>
<td>0.16</td>
<td>6,756</td>
<td>49</td>
<td>0.18</td>
<td>0.2</td>
<td>4,917</td>
<td>23</td>
</tr>
<tr>
<td>Task Performance</td>
<td>11,199</td>
<td>55</td>
<td>0.21</td>
<td>0.23</td>
<td>5,433</td>
<td>40</td>
<td>0.21</td>
<td>0.24</td>
<td>4,651</td>
<td>22</td>
</tr>
<tr>
<td>CWB/Withdrawal</td>
<td>2,781</td>
<td>15</td>
<td>-0.13</td>
<td>-0.15</td>
<td>1,210</td>
<td>8</td>
<td>0.04</td>
<td>0.04</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Note. $r_m$ = mean sample size-weighted correlation; $\rho$ = mean sample size-weighted correlation corrected for unreliability.
<table>
<thead>
<tr>
<th>Variable</th>
<th>(N)</th>
<th>(k)</th>
<th>(r_m)</th>
<th>(SD_{r})</th>
<th>(\rho)</th>
<th>(SD_{\rho})</th>
<th>% Var</th>
<th>(CV_{10})</th>
<th>(CV_{90})</th>
<th>(CL)</th>
<th>(CU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Performance</td>
<td>6,756</td>
<td>49</td>
<td>0.18</td>
<td>0.23</td>
<td>0.20</td>
<td>0.23</td>
<td>13.44</td>
<td>-0.10</td>
<td>0.50</td>
<td>0.13</td>
<td>0.28</td>
</tr>
<tr>
<td>Outcome Variable Performance</td>
<td>5,433</td>
<td>40</td>
<td>0.21</td>
<td>0.23</td>
<td>0.24</td>
<td>0.25</td>
<td>12.33</td>
<td>-0.08</td>
<td>0.55</td>
<td>0.16</td>
<td>0.32</td>
</tr>
<tr>
<td>CWB/Withdrawal</td>
<td>1,210</td>
<td>8</td>
<td>0.04</td>
<td>0.12</td>
<td>0.04</td>
<td>0.09</td>
<td>48.89</td>
<td>-0.07</td>
<td>0.16</td>
<td>-0.05</td>
<td>0.14</td>
</tr>
<tr>
<td>Construct Match</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Match</td>
<td>4,199</td>
<td>29</td>
<td>0.06</td>
<td>0.15</td>
<td>0.07</td>
<td>0.14</td>
<td>29.69</td>
<td>-0.11</td>
<td>0.25</td>
<td>0.01</td>
<td>0.13</td>
</tr>
<tr>
<td>Some Match</td>
<td>2,488</td>
<td>19</td>
<td>0.38</td>
<td>0.19</td>
<td>0.43</td>
<td>0.19</td>
<td>15.67</td>
<td>0.18</td>
<td>0.68</td>
<td>0.33</td>
<td>0.53</td>
</tr>
<tr>
<td>Complete Match</td>
<td>3,470</td>
<td>17</td>
<td>0.26</td>
<td>0.23</td>
<td>0.30</td>
<td>0.25</td>
<td>15.67</td>
<td>-0.03</td>
<td>0.62</td>
<td>0.17</td>
<td>0.42</td>
</tr>
<tr>
<td>Breadth Match</td>
<td>3,083</td>
<td>25</td>
<td>0.09</td>
<td>0.18</td>
<td>0.10</td>
<td>0.18</td>
<td>23.70</td>
<td>-0.13</td>
<td>0.33</td>
<td>0.02</td>
<td>0.19</td>
</tr>
<tr>
<td>Same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obj- more precise</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publication Status</td>
<td>7,522</td>
<td>53</td>
<td>0.19</td>
<td>0.21</td>
<td>0.22</td>
<td>0.22</td>
<td>14.76</td>
<td>-0.06</td>
<td>0.50</td>
<td>0.16</td>
<td>0.28</td>
</tr>
<tr>
<td>Not Published</td>
<td>1,556</td>
<td>5</td>
<td>0.00</td>
<td>0.07</td>
<td>-0.01</td>
<td>0.05</td>
<td>61.51</td>
<td>-0.07</td>
<td>0.06</td>
<td>-0.08</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Note. \(r_m\) = mean sample size-weighted correlation; \(SD_{r}\) = sample size-weighted observed standard deviation of correlations; \(\rho\) = mean sample size-weighted correlation corrected for unreliability; \(SD_{\rho}\) = corrected standard deviation of corrected correlations; % Var. = percentage of variance attributable to statistical artifacts; \(CV_{10}\) and \(CV_{90}\) = 10% and 90% credibility values, respectively; \(CL\) and \(CU\) = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean correlation.
<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>k</th>
<th>$r_m$</th>
<th>$SD_r$</th>
<th>$\rho$</th>
<th>$SD_\rho$</th>
<th>% Var</th>
<th>CV10</th>
<th>CV90</th>
<th>CI_L</th>
<th>CI_U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>13,979</td>
<td>71</td>
<td>0.14</td>
<td>0.26</td>
<td>0.16</td>
<td>0.27</td>
<td>7.53</td>
<td>-0.19</td>
<td>0.51</td>
<td>0.09</td>
<td>0.22</td>
</tr>
<tr>
<td>Objective Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td>11,199</td>
<td>55</td>
<td>0.21</td>
<td>0.24</td>
<td>0.23</td>
<td>0.25</td>
<td>8.09</td>
<td>-0.09</td>
<td>0.55</td>
<td>0.16</td>
<td>0.30</td>
</tr>
<tr>
<td>CWB/Withdrawal</td>
<td>2,781</td>
<td>15</td>
<td>-0.13</td>
<td>0.29</td>
<td>-0.15</td>
<td>0.31</td>
<td>6.34</td>
<td>-0.54</td>
<td>0.25</td>
<td>-0.31</td>
<td>0.02</td>
</tr>
<tr>
<td>Construct Match</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Match</td>
<td>2,383</td>
<td>12</td>
<td>-0.04</td>
<td>0.27</td>
<td>-0.04</td>
<td>0.29</td>
<td>6.69</td>
<td>-0.42</td>
<td>0.33</td>
<td>-0.22</td>
<td>0.13</td>
</tr>
<tr>
<td>Some Match</td>
<td>12,619</td>
<td>57</td>
<td>0.16</td>
<td>0.27</td>
<td>0.18</td>
<td>0.28</td>
<td>6.10</td>
<td>-0.19</td>
<td>0.54</td>
<td>0.10</td>
<td>0.26</td>
</tr>
<tr>
<td>Complete Match</td>
<td>1,519</td>
<td>10</td>
<td>0.31</td>
<td>0.15</td>
<td>0.35</td>
<td>0.15</td>
<td>23.73</td>
<td>0.16</td>
<td>0.54</td>
<td>0.25</td>
<td>0.45</td>
</tr>
<tr>
<td>Breadth Match</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same</td>
<td>4,666</td>
<td>20</td>
<td>0.17</td>
<td>0.22</td>
<td>0.19</td>
<td>0.24</td>
<td>8.22</td>
<td>-0.11</td>
<td>0.49</td>
<td>0.08</td>
<td>0.30</td>
</tr>
<tr>
<td>Obj- more precise</td>
<td>11,276</td>
<td>66</td>
<td>0.11</td>
<td>0.27</td>
<td>0.12</td>
<td>0.28</td>
<td>8.17</td>
<td>-0.24</td>
<td>0.48</td>
<td>0.05</td>
<td>0.19</td>
</tr>
<tr>
<td>Publication Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Published</td>
<td>12,185</td>
<td>65</td>
<td>0.15</td>
<td>0.27</td>
<td>0.17</td>
<td>0.29</td>
<td>7.12</td>
<td>-0.19</td>
<td>0.54</td>
<td>0.10</td>
<td>0.24</td>
</tr>
<tr>
<td>Not Published</td>
<td>1,794</td>
<td>6</td>
<td>0.06</td>
<td>0.11</td>
<td>0.07</td>
<td>0.10</td>
<td>28.23</td>
<td>-0.06</td>
<td>0.20</td>
<td>-0.03</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Note. $r_m$ = mean sample size-weighted correlation; $SD_r$ = sample size-weighted observed standard deviation of correlations; $\rho$ = mean sample size-weighted correlation corrected for unreliability; $SD_\rho$ = corrected standard deviation of corrected correlations; % Var. = percentage of variance attributable to statistical artifacts; CV10 and CV90 = 10% and 90% credibility values, respectively; CI_L and CI_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean correlation.
Table 4
Meta-analytic results: Relationships between peer-ratings and objective measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>k</th>
<th>rm</th>
<th>Dr</th>
<th>p</th>
<th>Dp</th>
<th>% Var</th>
<th>CV10</th>
<th>CV90</th>
<th>Cl_L</th>
<th>Cl_U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>4,917</td>
<td>23</td>
<td>0.29</td>
<td>0.22</td>
<td>0.31</td>
<td>0.08</td>
<td>8.16</td>
<td>0.03</td>
<td>0.23</td>
<td>0.09</td>
<td>0.17</td>
</tr>
<tr>
<td>Objective Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task Performance</td>
<td>4,651</td>
<td>22</td>
<td>0.30</td>
<td>0.22</td>
<td>0.34</td>
<td>0.23</td>
<td>8.36</td>
<td>0.04</td>
<td>0.63</td>
<td>0.24</td>
<td>0.44</td>
</tr>
<tr>
<td>CWB/Withdrawal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construct Match</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Match</td>
<td>664</td>
<td>3</td>
<td>0.02</td>
<td>0.12</td>
<td>0.02</td>
<td>0.11</td>
<td>31.25</td>
<td>-0.12</td>
<td>0.16</td>
<td>-0.13</td>
<td>0.18</td>
</tr>
<tr>
<td>Some Match</td>
<td>3,395</td>
<td>19</td>
<td>0.24</td>
<td>0.13</td>
<td>0.27</td>
<td>0.13</td>
<td>27.87</td>
<td>0.11</td>
<td>0.43</td>
<td>0.20</td>
<td>0.33</td>
</tr>
<tr>
<td>Complete Match</td>
<td>1,481</td>
<td>6</td>
<td>0.45</td>
<td>0.26</td>
<td>0.50</td>
<td>0.29</td>
<td>3.71</td>
<td>0.14</td>
<td>0.87</td>
<td>0.27</td>
<td>0.74</td>
</tr>
<tr>
<td>Breadth Match</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Same</td>
<td>3,360</td>
<td>14</td>
<td>0.34</td>
<td>0.23</td>
<td>0.38</td>
<td>0.25</td>
<td>6.14</td>
<td>0.07</td>
<td>0.70</td>
<td>0.25</td>
<td>0.52</td>
</tr>
<tr>
<td>Obj- more precise</td>
<td>2,009</td>
<td>14</td>
<td>0.18</td>
<td>0.12</td>
<td>0.21</td>
<td>0.10</td>
<td>44.85</td>
<td>0.08</td>
<td>0.33</td>
<td>0.14</td>
<td>0.28</td>
</tr>
<tr>
<td>Publication Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Published</td>
<td>4,014</td>
<td>20</td>
<td>0.29</td>
<td>0.23</td>
<td>0.32</td>
<td>0.24</td>
<td>7.97</td>
<td>0.01</td>
<td>0.63</td>
<td>0.21</td>
<td>0.43</td>
</tr>
<tr>
<td>Not Published</td>
<td>903</td>
<td>3</td>
<td>0.29</td>
<td>0.17</td>
<td>0.33</td>
<td>0.18</td>
<td>9.75</td>
<td>0.10</td>
<td>0.55</td>
<td>0.11</td>
<td>0.54</td>
</tr>
</tbody>
</table>

Note. rm = mean sample size-weighted correlation; Dr = sample size-weighted observed standard deviation of correlations; p = mean sample size-weighted correlation corrected for unreliability; Dp = corrected standard deviation of corrected correlations; % Var. = percentage of variance attributable to statistical artifacts; CV10 and CV90 = 10% and 90% credibility values, respectively; Cl_L and Cl_U = lower and upper bounds, respectively, of the 95% confidence interval around the corrected mean correlation.
Table 5
*Meta-Analytic Correlation Matrix for Incremental Validity for Overall Performance*

<table>
<thead>
<tr>
<th>Objective</th>
<th>Self</th>
<th>Peer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self</td>
<td>.18 (this study)</td>
<td></td>
</tr>
<tr>
<td>Peer</td>
<td>.29 (this study)</td>
<td>.19 (Conway)</td>
</tr>
<tr>
<td>Supervisor</td>
<td>.14 (this study)</td>
<td>.24 (Heidemeier)</td>
</tr>
</tbody>
</table>


Table 6
*Incremental Validity Analysis for Overall Performance*

<table>
<thead>
<tr>
<th>Model/Step</th>
<th>Sources Included</th>
<th>β</th>
<th>R²</th>
<th>ΔR²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A</td>
<td>Supervisor</td>
<td>.140*</td>
<td>.020</td>
<td>.020*</td>
</tr>
<tr>
<td></td>
<td>+ Peer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Supervisor</td>
<td>.025</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Peer</td>
<td>.280*</td>
<td>.084</td>
<td>.065*</td>
</tr>
<tr>
<td>3</td>
<td>Supervisor</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Peer</td>
<td>.265*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Self</td>
<td>.130*</td>
<td>.100</td>
<td>.016*</td>
</tr>
<tr>
<td>Model B</td>
<td>Peer</td>
<td>.290*</td>
<td>.084</td>
<td>.073*</td>
</tr>
<tr>
<td>2</td>
<td>Peer</td>
<td>.280*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Supervisor</td>
<td>.025*</td>
<td>.085</td>
<td>.001*</td>
</tr>
<tr>
<td>3</td>
<td>Peer</td>
<td>.25*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Supervisor</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Self</td>
<td>.130*</td>
<td>.100</td>
<td>.016*</td>
</tr>
</tbody>
</table>

*Note.* Harmonic mean N = 7,825

*p<.05.
Table 7
Meta-Analytic Correlation Matrix for Incremental Validity for Task Performance

<table>
<thead>
<tr>
<th>Objective</th>
<th>Self</th>
<th>Peer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self</td>
<td>.21 (this study)</td>
<td></td>
</tr>
<tr>
<td>Peer</td>
<td>.30 (this study)</td>
<td>.19 (Conway)</td>
</tr>
<tr>
<td>Supervisor</td>
<td>.21 (this study)</td>
<td>.20 (Heidemeier)</td>
</tr>
</tbody>
</table>


Table 8
Incremental Validity Analysis for Task Performance

<table>
<thead>
<tr>
<th>Model/Step</th>
<th>Sources Included</th>
<th>β</th>
<th>$R^2$</th>
<th>$\Delta R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Supervisor</td>
<td>.210*</td>
<td>.044</td>
<td>.044*</td>
</tr>
<tr>
<td>2</td>
<td>Supervisor + Peer</td>
<td>.110*</td>
<td>.100</td>
<td>.056*</td>
</tr>
<tr>
<td>3</td>
<td>Supervisor + Peer+ Self</td>
<td>.088*</td>
<td>.120</td>
<td>.020*</td>
</tr>
<tr>
<td><strong>Model B</strong></td>
<td>Peer + Supervisor</td>
<td>.300*</td>
<td>.090</td>
<td>.090*</td>
</tr>
<tr>
<td>2</td>
<td>Peer + Supervisor</td>
<td>.257*</td>
<td>.110*</td>
<td>.010*</td>
</tr>
<tr>
<td>3</td>
<td>Peer + Supervisor + Self</td>
<td>.238*</td>
<td>.088*</td>
<td>.147*</td>
</tr>
</tbody>
</table>

Note. Harmonic mean N= 5,474
*p<.05.
Table 9
Meta-Analytic Correlation Matrix for Incremental Validity for CWB

<table>
<thead>
<tr>
<th>Source</th>
<th>Objective</th>
<th>Self</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self</td>
<td>.04 (this study)</td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>-.13 (this study)</td>
<td>.31 (Berry)</td>
</tr>
</tbody>
</table>

Note. Berry = Berry, Carpenter & Barratt, 2012.

Table 10
Incremental Validity Analysis for CWB

<table>
<thead>
<tr>
<th>Model/Step</th>
<th>Sources Included</th>
<th>β</th>
<th>R^2</th>
<th>ΔR^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Supervisor</td>
<td>-.130*</td>
<td>.017</td>
<td>.017*</td>
</tr>
<tr>
<td>2</td>
<td>Supervisor</td>
<td>-.158*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Self</td>
<td>.089*</td>
<td>.023</td>
<td>.007*</td>
</tr>
</tbody>
</table>

Note. Harmonic mean N= 1,790
*p<.05.
### Table 11

*Meta-Analytic Correlation Matrix for Incremental Validity for Complete Construct Match*

<table>
<thead>
<tr>
<th></th>
<th>Objective</th>
<th>Self</th>
<th>Peer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self</td>
<td>.38 (this study)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer</td>
<td>.45 (this study)</td>
<td>.19 (Conway)</td>
<td></td>
</tr>
<tr>
<td>Supervisor</td>
<td>.31 (this study)</td>
<td>.24 (Heidemeier)</td>
<td>.41 (Viswesvaran)</td>
</tr>
</tbody>
</table>

*Note. N= 5939; Conway= Conway & Huffcut, 1997; Heidemeier = Heidemeier & Moser, 2009; Viswesvaran= Viswesvaran, Schmidt & Ones, 2002.*

### Table 12

*Incremental Validity Analysis for Task Performance*

<table>
<thead>
<tr>
<th>Model/Step</th>
<th>Sources Included</th>
<th>( \beta )</th>
<th>( R^2 )</th>
<th>( \Delta R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Supervisor</td>
<td>.310*</td>
<td>.096</td>
<td>.096*</td>
</tr>
<tr>
<td></td>
<td>+ Peer</td>
<td>.388*</td>
<td>.221</td>
<td>.125*</td>
</tr>
<tr>
<td>2</td>
<td>Supervisor</td>
<td>.151*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Peer</td>
<td>.356*</td>
<td>.299</td>
<td>.078*</td>
</tr>
<tr>
<td>3</td>
<td>Supervisor</td>
<td>.094*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Peer</td>
<td>.290*</td>
<td>.299</td>
<td>.078*</td>
</tr>
<tr>
<td></td>
<td>+ Self</td>
<td>.290*</td>
<td>.299</td>
<td>.078*</td>
</tr>
<tr>
<td><strong>Model B</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Peer</td>
<td>.450*</td>
<td>.202</td>
<td>.203*</td>
</tr>
<tr>
<td>2</td>
<td>Peer</td>
<td>.388*</td>
<td>.151*</td>
<td>.019*</td>
</tr>
<tr>
<td></td>
<td>+ Supervisor</td>
<td>.221</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Peer</td>
<td>.356*</td>
<td>.290*</td>
<td>.078*</td>
</tr>
<tr>
<td></td>
<td>+ Supervisor</td>
<td>.290*</td>
<td>.299</td>
<td>.078*</td>
</tr>
<tr>
<td></td>
<td>+ Self</td>
<td>.290*</td>
<td>.299</td>
<td>.078*</td>
</tr>
</tbody>
</table>

*Note. Harmonic mean \( N= 2,885 \)

*p<.05*