

# The Social Perception of Emotional Abilities: Expanding What We Know About Observer Ratings of Emotional Intelligence

Hillary Anger Elfenbein  
Washington University in St. Louis

Sigal G. Barsade  
University of Pennsylvania

Noah Eisenkraft  
University of North Carolina, Chapel Hill

We examine the social perception of emotional intelligence (EI) through the use of observer ratings. Individuals frequently judge others' emotional abilities in real-world settings, yet we know little about the properties of such ratings. This article examines the social perception of EI and expands the evidence to evaluate its reliability and cross-judge agreement, as well as its convergent, divergent, and predictive validity. Three studies use real-world colleagues as observers and data from 2,521 participants. Results indicate significant consensus across observers about targets' EI, moderate but significant self-observer agreement, and modest but relatively consistent discriminant validity across the components of EI. Observer ratings significantly predicted interdependent task performance, even after controlling for numerous factors. Notably, predictive validity was greater for observer-rated than for self-rated or ability-tested EI. We discuss the minimal associations of observer ratings with ability-tested EI, study limitations, future directions, and practical implications.

*Keywords:* emotional intelligence, social perception, observer ratings, self-ratings, ability tests, social relations model

*Supplemental materials:* <http://dx.doi.org/10.1037/a0038436.supp>

The public interest and academic research in emotional intelligence (EI) is at least partly inspired by the commonsense notion that those around us vary in their emotional skills. It is relatively easy to point out examples and cautionary tales of EI in literary works and in real life (e.g., Oatley, 2004). Indeed, Averill (2004) argued that among the reasons for the prominence of the EI construct “is that it has an easily recognizable grain of truth. We all know people who are emotionally adept” (p. 228). In the present article, we focus on this phenomenon—that is, people's beliefs about other people's emotional abilities. We take a social perception lens to the construct of emotional intelligence and, in doing so, attempt systematically to develop a greater understanding of observer ratings.

We are interested in the social perception of emotional abilities as an epiphenomenal construct. That is, judgments of EI are theoretically important in their own right because they exist in our minds and guide our behavior. People continually make these judgments, and these judgments have meaningful interpersonal

implications. As Funder and West (1993) argued, interpersonal perceptions are the basis of common and consequential decisions, regardless of whether they correspond to an objective reality. In this article, we argue that individuals in organizational settings continually make judgments of other people's emotional abilities. The modern workplace attempts explicitly to recruit, select, and promote individuals who have interpersonal skills and, as such, social perceptions of EI should be central to individuals' career advancement. Given that judgments of other people's emotional abilities are involved in consequential decisions, it is worthwhile to understand them better. Indeed, Roberts and colleagues (2001) argued that in the field of EI “validation studies of this type [that is, observer perceptions] . . . [are] in urgent need of detailed investigation” (p. 201). The present investigation attempts to fill this gap by expanding the empirical basis of evidence about observers' judgments about other people's emotional abilities.

## Understanding Observer Reports of EI

The goal of this article is to provide a comprehensive empirical investigation of the social perception of EI through a large-scale study of observer ratings. This investigation includes key psychometric properties involved with evaluating constructs and measures, which are detailed in the following paragraphs.

## Consensus Among Judges

In attempting to understand the meaning of observer perceptions, it is important to know whether observers agree. Given that

---

Hillary Anger Elfenbein, Olin School of Business, Washington University in St. Louis; Sigal G. Barsade, Wharton School, University of Pennsylvania; Noah Eisenkraft, Kenan-Flagler Business School, University of North Carolina, Chapel Hill.

The authors thank Jennifer Boldry, Stéphane Côté, Jeff Edwards, and David Kenny for helpful comments and collegueship.

Correspondence concerning this article should be addressed to Hillary Anger Elfenbein, Washington University in St. Louis, Olin School of Business, 1 Brookings Drive, Saint Louis, MO 63130. E-mail: [hillary@post.harvard.edu](mailto:hillary@post.harvard.edu)

observer judgments are often used in organizational contexts related to job selection and advancement, it is important to know to what extent different judges tend to agree with each other. If judges fail to agree meaningfully, then this suggests there may be a somewhat arbitrary component to the pervasive phenomenon of judging others' social skills.

Some evidence already exists for moderate consensus in judging emotional abilities. To date, a handful of EI studies have included observer ratings—often, but not always, in an attempt to validate the self-report instruments that were of primary interest to the investigations (e.g., Law, Wong, & Song, 2004). For this reason, most of these studies used a single judge or used multiple judges without reporting cross-judge agreement (e.g., Law et al., 2004; van der Zee, Thijs, & Schakel, 2002). Kellett, Humphrey, and Sleeth (2006) reported intraclass correlations for cross-judge agreement of .32 for emotional expression ability and .23 for emotion perception ability in their study. Likewise, researchers have reported convergence between the EI ratings made by participants' peers and supervisors ( $r = .34$ , Law et al., 2004;  $r_s = .23-.30$ ; Lopes, Grewal, Kadis, Gall, & Salovey, 2006). Charbonneau and Nicol (2002) reported that interrater agreement ( $r_{wg}$ ) on interpersonal EI was greater than .70 for most of their sample. Further, they reported that peer and supervisor ratings correlated at .22, which is moderate but suggests some underlying signal that observers detected in common. These values compare moderately with the relatively good convergence between peer and supervisor ratings ( $r = .48$ ) found in Harris and Schaubroeck's (1988) meta-analysis of workplace job performance. The current article expands this evidence by examining every factor separately from two different four-factor models of EI, using advanced statistical techniques with the social relations model for social perception (Kenny, 1994) and by greatly increasing the existing body of data with over 2,500 participants.

### Convergent Validity

Although observer reports of emotional abilities are important epiphenomenal constructs in their own right, it is worthwhile to examine to what extent they converge with existing ability tests of EI. Observer reporting of EI is an important phenomenon, regardless of its association or lack of association with ability tests. As an epiphenomenal construct, social perception is important simply because people engage in it and because it has important consequences. Even so, to understand the properties of this construct, it is worthwhile to examine its potential convergence with ability measures of EI. Theorists now recognize the ability model as the most authoritative approach to EI (Matthews, Roberts, & Zeidner, 2004; Mayer et al., 2008; Rivers, Brackett, Salovey, & Mayer, 2007). This model begins with the larger concept of intelligence, the various definitions of which typically involve a person's ability to "deal effectively with his [or her] environment" (Wechsler, 1944, p. 3, see also Roberts, Zeidner, & Matthews, 2001). Top theorists of EI argue that the criteria for what it means to deal effectively on an emotional level are socially shared. With this reasoning, Mayer, Salovey, Caruso, and Sitarenios (2003) argued that consensus scoring is an appropriate method of determining correct answers for their test because their "model of EI hypothesizes that emotional knowledge is embedded within a general, evolved, social context of communication" that allows even un-

trained samples to define correct responses (p. 98). Mayer et al. (2003) found empirical evidence that there is a high level of agreement regarding the correct answers to the items on their ability test and a high rate of agreement between the responses of researchers and those of the general public used for consensus coding. There is a possibility that if untrained observers can perceive skilled responses on a test of EI ability, then they might also be able to perceive to what degree other people typically make these skilled responses in actual situations. However, it is alternately possible that observers form their judgments on the basis of behavioral features that are unrelated to the content of ability tests. The present study provides data to consider these issues.

There exists some limited evidence for significant associations between observer ratings and ability tests of emotional skills in nonverbal communication ( $r = .20$ ; Rosenthal et al., 1979) and interpersonal sensitivity ( $r = .48$ ; Costanzo & Archer, 1989). Lopes et al. (2006) found convergence between an ability test of overall EI with peer and supervisor reports of trait EI. However, other studies have shown no association among ability tests and observer reports, such as Carney and Harrigan's (2003) investigation of both emotion perception accuracy and social situation understanding.

### Divergent Validity

A longstanding concern among theorists of EI is the criticism that EI might merely be a repackaging of old wine in new bottles (Matthews, Zeidner, & Roberts, 2002; Mayer, Caruso, & Salovey, 1999). As such, it is important to examine any concept related to EI with other constructs that have positive valence. We begin with liking and positive regard more generally, for their potential halo effect that could cast a positive glow across ratings of other desirable characteristics. We also include observer ratings of emotional stability—that is, low levels of neuroticism, for the potentially confounding nature of this other emotion-relevant personality trait. In the little research conducted so far on this topic, Song et al. (2010) found only a small association ( $r = .12$ ) between liking and observer-rated EI. The current research adds to existing evidence, with an expanded list of potential confounds and with confirmatory factor models based on large numbers of participants.

### Predictive Validity

The predictive validity of EI has been considered a 'holy grail' in the field, and quests to find it have been met with varying levels of success (Côté & Miners, 2006; Elfenbein 2007; Joseph & Newman, 2010; Matthews et al., 2002; Mayer, Roberts, & Barsade, 2008). As such, it is worthwhile to examine the predictive validity of observer-rated EI. There is initial evidence that observer-reported EI predicts important life outcomes. Law et al. (2004) examined observer reports as a supporting measure in validating their self-report scale of EI. They presented evidence that observer-rated EI had substantial predictive validity for job performance, even when controlling for the Big Five personality dimensions. Indeed, the predictive validity of observer-rated EI was higher than that of self-reported EI in their study. Sala (2005) reviewed evidence showing that observer ratings on the emotional competence inventory predict a range of organizationally relevant criteria such as job performance across several industries, percep-

tions of leadership, and lower levels of stress (see also Boyatzis & Sala, 2004). Evidence also shows that observer-rated EI can predict social outcomes outside of the workplace (Law et al., 2004; Sala, 2005; Song, Huang, Peng, Law, Wong, & Chen, 2010; van der Zee et al., 2002). Our study extends the body of evidence in the workplace domain by examining the predictive power of observer-rated EI beyond that of a wide range of control variables, including conventional intelligence and existing personality traits.

Drawing from Staw, Sutton, and Pelled's (1994) theoretical discussion of positive affect and job performance, there are three theoretical mechanisms that could produce a positive association between observer-rated EI and job performance. First, in a veridical association, higher EI could help individuals outperform their low EI counterparts, through greater awareness of their own and others emotions, using emotions to facilitate cognition, understanding emotions, and regulating their own and others emotions. Any or all of these could lead, for example, to better leadership, decision making, negotiation, and other aspects of work performance (e.g., Elfenbein, Foo, White, Tan, & Aik, 2007; Mayer et al., 2008). Second, higher EI can endear a person to their colleagues, who provide tangible support and resources that facilitate performance. Third, higher EI can lead to a halo effect of competence, above and beyond being liked or regarded more positively in general. Colleagues could perceive greater performance from individuals high in EI when their actual performance is no better than that of their low EI counterparts. We argue that all three of these mechanisms, singly and in combination, can lead to the predictive validity of observer-judged EI. Note that other than the first mechanism (i.e., a veridical association), the other two mechanisms are mediated by colleagues' behavior. Given that colleagues' behaviors are likely to be most proximally influenced by their own assessments, observer ratings are particularly important to understand.

### Self-Observer Agreement

It is also worthwhile to explore the extent to which observer reports converge with self-perceptions. In doing so, it is important to emphasize that self-reported EI cannot be taken at face value as a veridical construct. Abilities can be self-reported the most accurately in those domains for which there are 'gold standards' for testing and for which people receive explicit feedback about where they stand. This includes mathematical and verbal intelligence as well as athletic abilities (Kuncel, Credé, & Thomas, 2005; Mabe & West, 1982). This criterion is not met for EI in that we live in a 'white lie' society (Swann, Stein-Seroussi, & McNulty, 1992) that provides relatively little clear feedback about our interpersonal skills. As such, the accuracy of self-reported EI is likely to be imperfect. Expecting individuals to report about their own emotional skills may be asking them to tell more than they could know (Nisbett & Wilson, 1977). Further, somewhat paradoxically, those without EI may lack the metacognitive skills to be aware of their low EI (Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006), which adds to arguments against taking the self-report of EI at face value. Some theorists have even referred to trait models of EI as "emotional self-efficacy" (Petrides et al., 2007, p. 151; Tett, Fox, & Wang, 2005, p. 859), suggesting that self-reported emotional traits represent one's priorities, motivations, and hopes for emotional processes as much as one's skill, per se. Thus, like observer

reports, self-reports of EI are also epiphenomenal—that is, they are theoretically important in their own right because they exist in our minds. Without considering self-reported EI as a valid construct at face value, the studies below examine self-observer agreement to consider the role of social perception from the lens of both the observer and target.

There are two distinct theoretical mechanisms that could produce significant self-observer agreement. First, there can be some accurate signal that makes its way through the noise of both self-awareness and observer perception. People can observe their own behaviors to see whether they achieve their emotionally desired ends, and other people can give their feedback along the way. In terms of the possibility that there is an accurate signal, it is worth noting that the field of personality treats self-other agreement as a gold standard to validate self-reports (Funder, 1991; Kolar, Funder, & Colvin, 1996). The logic is that observers are in a better position than targets to assess how these targets behave in the social environment. As such, self-other agreement could indicate accuracy on the part of self-ratings. Without accepting the strong version of this argument in the case of EI, the argument adds to the value of exploring self-other agreement. A second potential mechanism is through self-verification processes. People can help to shape others' perceptions in order to persuade them, however subtly, to view them the way they view themselves (Swann, Chang-Schneider, & Larsen McClarty, 2007).

There is already some evidence in support of significant, even if moderate, associations between observer reports and self-reported EI. Law et al. (2004) demonstrated good self-observer agreement for total EI rated by parents, workplace observers, and supervisors, with coefficients ranging from .28–.41. These values are on par with the self-observer agreement that the same authors found for the Big Five personality traits. Other reported self-observer correlations in emotional domains have been  $r = .20$  for emotional self-awareness,  $r = .43$  for emotional self-control, and  $r = .29$  for empathy (Sala, 2005);  $r = .33$  for emotional self-awareness (Bar-On & Handley, 2003), and  $r = .13$  for emotion perception accuracy (Carney & Harrigan, 2003). In research using measures from the trait EI perspective, there have been reports of  $r = .09$  for interpersonal EI (Charbonneau & Nicol, 2002),  $r = .18$  for an overall EI inventory (Jordan & Ashkanasy, 2005), and  $r = .12$  for empathy rated by unacquainted observers (D'Augelli, 1973). One report found self-observer agreement correlations as high as .56–.65 (van der Zee et al., 2002), whereas others tend to be more moderate; for example,  $r = .26$  to  $.35$  in Libbrecht, Lievens, and Schollaert (2010). This general picture of self-observer agreement across emotional skills is in line with the level of self-peer agreement of  $r = .24$  for job performance based on Harris and Schaubroeck's (1988) meta-analysis. For comparison, observer ratings of traditional cognitive intelligence, self-other agreement is typically in the range of low .30s for community samples and rarely greater than .30 for college students (Paulhus, Lysy, & Yik, 1998). As with observer agreement, we expand existing evidence with a comprehensive examination of two different four-factor models.

### Measurement Equivalence Between Self- and Observer Ratings

It is worthwhile to expand the evidence for evaluating the measurement equivalence between self and observer reports of EI.

This indicates to what extent the measures are used similarly by these two sources, in terms of factor loadings, intercepts, and mean values among factors in confirmatory models. Existing data have shown that self- and other ratings of EI are largely equivalent in measuring the same underlying construct, albeit with some leniency bias for self-ratings (Joseph & Newman, 2010; Libbrecht et al., 2010).

### The Current Studies

Below we report the results of three studies conducted with real-world colleagues to examine the properties of social judgments of EI. Study 1 offers a first examination of observer reports within a large sample of teams of full-time MBA students who worked closely together and uses observer reports adapted from a well-validated self-report questionnaire. Study 2 extends these findings using observer reports of EI and performance reviews from individuals' workplace colleagues and supervisors, respectively, and expands the evidence by adapting observer reports from a different EI instrument. Study 3 attempts to address potential limitations of the first two studies by demonstrating that observer ratings of EI are distinct from liking, positive regard, and emotional stability; includes an ability test of EI to evaluate convergent validity; and tests the predictive validity of observer-rated EI, while controlling for liking and positive regard. Taken as a whole, these studies present the most comprehensive test to date for evaluating the social perception of EI.

#### Study 1

We take a social perception approach to the study of EI and, for this reason, use the authoritative conceptual model developed for the study of interpersonal perception, Kenny's (1994) social relations model (SRM; see also Kenny & LaVoie, 1984).

The SRM describes how all interpersonal ratings, such as those of EI, consist of four mutually exclusive and exhaustive components. First, *target effects* consist of consensus across raters, in that the group as a whole systematically rates some people as having higher versus lower EI. This is our most important coefficient, and it is the operational definition of observer-rated EI. Second, *rater effects* consist of response bias on the part of raters who vary in the average scores that they assign when judging others' EI. This is a matter of leniency bias. Third, *relationship effects* consist of dyadic perceptions whereby some people uniquely find particular others to be very high or low in EI—even after controlling for the perceiver's general rating of others and the target's typical rating by others. The fourth component of the SRM is measurement error, which results from imperfect inter-item correlations within the questionnaire. For the purpose of this study, we use the SRM first to examine the extent to which interpersonal ratings of EI show consensus across raters. We use it second to construct variables for further analysis that represent each individual's aggregate ratings from multiple teammates. To implement the SRM, participants took part in a round robin design, in which each member of a group rated each other member of the group (Kenny & LaVoie, 1984). In analyzing a round robin, it is important to use specialized algorithms versus conventional analysis, which would be vulnerable to artifacts that are due to systematically missing data because there can be no observer ratings of the self. As such,

statistics such as the within-group interrater index  $r_{wg}$  cannot be used because this agreement measure requires interchangeable raters of a single target (James, Demaree, & Wolf, 1993). The SRM also corrects for biases such as leniency, which can influence observer and self-reports. These positive features of the SRM make it the optimal approach for expanding evidence to evaluate observer ratings of EI.

### Method

**Participants.** This study sampled full-time students in a large MBA program on the East Coast of the United States. Students completed measures as part of their required Organizational Behavior course. The Institutional Review Board approved these data to be used afterward for research purposes without individual identifiers. In their first-year coursework, 998 participants were assigned into student teams of 5 to 6 members who completed assignments together across all courses. The amount of usable data was 93.2%, for 931 individuals and 186 teams ( $M = 5.31$  individuals per team,  $SD = 0.72$ ). Because SRM analyses cannot accommodate missing data (Kenny, 1998), individuals were excluded from analysis if they did not receive complete observer ratings, regardless of their own participation. This affected 6.7% of individuals. Before participants arrived at the university, team composition was determined by a process that ensured diverse representation in terms of gender, national origin, and the nature of previous work experience. This process reduced, but did not eliminate, the likelihood of prior acquaintance.

We chose this population for a number of reasons. First, the observers were well-acquainted after working closely together full-time for approximately 2 months, across a range of settings with real consequences; for example, quantitative and qualitative course assignments, including problem sets, business case analyses, and a major group project requiring extensive coordination. Observer ratings tend to have greater agreement with increasing acquaintance, particularly with opportunities to view the attributes in question (Funder & Colvin, 1988; Watson & Clark, 1991). Second, ability measurement theorists have argued that empirical data should be collected among individuals who are relatively comparable along attributes such as age, gender, education level, and other factors that might influence performance (Carroll, 1993). Third, interpersonal judgments tend to have the greatest consensus among raters who have been exposed to each other in group settings—giving all observers experiences in common versus separate one-on-one contact (Kenny, 1994). Fourth, for observer ratings of personality, Costa and McCrae (1992) found that four appeared to be the optimal number of raters, with diminishing returns for larger numbers. Last, the individuals collaborated on tasks that had real consequences, which allowed them to form opinions about ecologically valid outcome variables, notably work performance.

**Measures.** Participants completed measures in two waves. First, no later than the second day of the course, participants completed all the individual difference measures except as specified below. Second, at the end of the course—approximately 8 weeks later—participants completed observer ratings of EI and performance, as well as the remaining individual difference measures. In the case of missing data on EI or performance, analyses

below used list-wise deletion. In the case of missing data on control measures, analyses used mean substitution.

**Self-ratings of EI.** Among the self-rating scales that have been developed to measure EI, some of the most extensive published validity data has been documented for the Wong and Law Emotional Intelligence Scale (WLEIS; Wong & Law, 2002). The 16-item WLEIS is based on Davies, Stankov, and Roberts' (1998) model of EI derived from factor analysis of most EI-related scales available at the time. This model includes the following factors: (a) *Self-Emotions Appraisal* (SEA), which is intrapersonal awareness of one's own emotional states; (b) *Others'-Emotions Appraisal* (OEA), which is interpersonal awareness of other people's emotional states; (c) *Use of Emotion* (UOE), which is using one's own emotional states to bring out one's best motivation and achievement; and (d) *Regulation of Emotion* (ROE), which is the ability to control one's own emotions. The authors selected four questionnaire items for each factor on the basis of factor loadings and correlations with criterion variables across multiple field samples. Law et al. (2004) presented further validity data on the scale in the form of confirmatory factor analysis of their model, convergent validity with respect to reports from peers, parents, and supervisors, and incremental predictive validity of criteria such as life satisfaction and job performance, while controlling for self-reported personality traits. Even with this validity evidence, it is worth noting that the UOE scale has items that are double-barreled and that have construct overlap with self-efficacy, and the following results are interpreted in this light.

Self-ratings of EI appeared in the same session as observer ratings, which took place at the end of the course to maximize acquaintance. Instructions for these evaluations emphasized their confidential nature—that is, that they would be shared with targets only in aggregate—which is particularly important given findings that ratings of ability are more accurate with such assurances (Mabe & West, 1982).

**Observer ratings of EI.** In keeping with the convention of converting self-reported personality scales into observer-reported versions (e.g., McCrae & Costa, 1987), we adapted the Law et al. (2004) scale by replacing the word “I” with the teammate's first name, the word “me” with “him/her,” and the word “my” with “his or her.” For questions that were highly intrapersonal, we also added “seems to” before the verb, to legitimize taking one's best guess in rating another person's internal functioning.

**Performance ratings from teammates.** For the purpose of developmental feedback, team members rated each other's performance at the end of the semester. Feedback was provided after the course ended in an aggregated form to preserve the privacy of individual respondents. They rated each colleague's performance as a *teammate* (“How effective a member of your team was [name]?”), *individual* (“How effective an individual contributor was [name]?”), and *leader* (“How effective a leader was [name]?”) on scales ranging from 1 (*not at all effective*) to 5 (*extremely effective*).

**Demographic background.** Given that EI and/or performance ratings could be differentially related to gender (e.g., Brackett, Mayer & Warner, 2004; Brackett et al., 2006), age (Mayer, Caruso, & Salovey, 1999), and sharing the culture's dominant first language (Downs, Strand, & Cerna, 2007), we included these factors as control variables in predictive validity analyses.

**Personality scales.** To establish the divergent validity of EI measures with respect to conventional personality traits, participants completed several self-report scales. We assessed the traits of Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness using John and Srivastava's (1999) Big Five Inventory. Tellegen's (1982) multidimensional personality questionnaire (MPQ) measured Trait Positive Affectivity (Trait PA) and Trait Negative Affectivity (Trait NA). Trait PA and NA were particularly worthwhile to include in light of the association between workplace outcomes and trait affectivity (e.g., Côté, 1999; Staw & Barsade, 1993; Staw et al., 1994). Because of time constraints, the MPQ appeared in the second wave of measures. We note that the expansive use of personality characteristics is worthwhile in research that includes self-reported EI in that self-reports tend to be confounded with a wide range of other factors (Joseph, Jin, Newman, & O'Boyle, 2014).

**Cognitive intelligence.** As a measure of cognitive intelligence, we used participants' verbal, quantitative, and total scores for the Graduate Management Admissions Test (GMAT). Cognitive intelligence—particularly verbal intelligence—is a worthwhile control variable when establishing the predictive validity of EI (Mayer et al., 2008). Although values were self-reported, meta-analytic evidence shows very high associations between students' self-reports and actual standardized test scores (Newman & Lyon, 2009).

## Results

Supplementary Table S1 summarizes descriptive statistics, reliability levels, and bivariate correlations for all study measures.

**Consensus among judges.** Table 1 summarizes the output of a series of SRMs. As discussed above, the SRM provides estimates for the amount of variance in colleagues' ratings that can be attributed to the target, the perceiver, the idiosyncratic relationship between target and perceiver and measurement error. In the current model, target variance indicates the degree of consensus among judges in the EI of each focal individual. The magnitude of variance can be interpreted akin to an  $R^2$  value in conventional statistical models. In support of agreement among judges, the level of target variance for each of the four components of EI ranged from 9% to 25%. This variance was statistically significant in each case and exceeded the 5% threshold suggested for interpreting SRM effects as meaningful (Kenny, 1994).

Although this exploration is focused on target variance, which indicates cross-judge consensus, for completeness with respect to the SRM and for the sake of exploring other properties of observer-rated EI, we also report effect sizes for rater variance. In the present design, in which individuals recorded impressions of each other's EI using a rating scale, rater variance can be interpreted as a type of artifact or response bias. It refers to the extent to which individual raters use higher versus lower mean values in assigning all of their observer ratings. This response bias is controlled in the following analyses via the SRM.

**Self-observer agreement.** There was significant agreement between observer and self-ratings, and Table 2 presents correlations among these variables. Self-observer agreement is positive and significant for all four branches of EI: SEA ( $r = .08$ ; 95% CI .02–.15), OEA ( $r = .14$ ; 95% CI .07–.20), UOE ( $r = .24$ ; 95% CI

Table 1  
*Variance Partitioning of Observer-Rated Emotional Intelligence (EI) (Studies 1 and 3)*

| Factor                   | Source of variance |       |                    |                    |
|--------------------------|--------------------|-------|--------------------|--------------------|
|                          | Target             | Rater | Relationship/Error | Dyadic reciprocity |
| a. Study 1               |                    |       |                    |                    |
| Emotional Intelligence   |                    |       |                    |                    |
| Self-emotions appraisal  | <b>9%</b> **       | 40%** | 51%                | .02                |
| Other-emotions appraisal | <b>21%</b> **      | 28%** | 51%                | .14**              |
| Use of emotion           | <b>25%</b> **      | 26%** | 49%                | .09**              |
| Regulation of emotion    | <b>21%</b> **      | 28%** | 50%                | .05                |
| Total EI                 | <b>18%</b> **      | 41%** | 41%                | .13**              |
| b. Study 3               |                    |       |                    |                    |
| Emotional Intelligence   |                    |       |                    |                    |
| Perceiving Emotions      | <b>16%</b> **      | 26%** | 58%                | .16                |
| Use of Emotions          | <b>21%</b> **      | 12%** | 67%                | -.11*              |
| Understanding Emotions   | <b>11%</b> **      | 34%** | 55%                | .15**              |
| Managing Emotions        | <b>23%</b> **      | 24%** | 54%                | .19                |
| Total EI                 | <b>21%</b> **      | 26%** | 53%                | .20**              |
| Liking/Regard            | <b>24%</b> **      | 26%** | 50%                | .44**              |
| Emotional stability      | <b>26%</b> **      | 17%** | 57%                | -.01               |

*Note.* Study 1 included 3,990 ratings from  $N = 931$  individuals in 186 teams. Study 3 included 1,153 ratings from 299 individuals in 62 teams. Boldface indicates coefficients representing consensus among judges.  
 \*  $p < .05$ . \*\*  $p < .01$ ; all values two-tailed.

.18–.31), ROE ( $r = .25$ ; 95% CI .19–.31), and overall EI ( $r = .15$ ; 95% CI .08–.21).

**Convergent and divergent validity.** The data presented in Table 2 provide the information necessary to use Campbell and Fiske's (1959) classic multitrait multimethod (MTMM) approach to explore convergent and divergent validity. Using MTMM, researchers assess distinct constructs using two or more methods, and they examine several criteria. First, scores for a particular construct should be similar across different measurement strategies—that is, the self–observer agreement hypothesized in the preceding paragraphs. Second, scores for the same construct should be more similar than scores for distinct constructs. In their large-scale study of trait EI, Law et al. (2004) conducted an MTMM analysis to contrast overall EI with conventional personality traits, and they found moderate evidence for discriminant validity. In the present case, the variables are four subfactors within the single construct of EI. Our analyses explore MTMM among these four factors, instead of contrasting EI with distinct constructs such as personality traits. That is, the analysis examines whether individual components of EI show convergence when measured using different methods as well as divergence with respect to each other. Note that Study 1 used only two methods to assess EI. Therefore, there were insufficient degrees of freedom to make a formal test of these criteria using structural equation modeling versus visual inspection.

Along these lines, it is worth noting that this MTMM is a highly conservative and stringent test. The analysis is typically applied to variables that are theoretically more distinct than four factors within a single construct, as in the present case. The veridical relationships among the four factors of EI would make it more difficult to establish construct validity. The possibility of halo effects or stereotypes about who might have higher versus lower EI should also influence all four factors equally and should make it harder to establish construct validity. Theories posit that the

multiple factors within higher order intelligences should have a positive manifold—that is, nonnegative correlations with each other (Gutman & Levy, 1991; Roberts et al., 2001). As such, we still expect positive correlations among the factors of EI, which is theoretical substance rather than mere methodological artifact. Thus, the MTMM analyses presented below are a highly conservative test of the construct validity of observer-rated EI.

These values provide evidence, albeit modest, for the criteria outlined by Campbell and Fiske (1959). First, the reliability coefficients—listed in the first column of Table 2—are the largest values in any associated row or column. Second, as reported earlier, the coefficients for self–observer agreement, although not large, are significantly different from zero. Third, these data indicate some discriminant validity. The criterion to apply is that the convergent correlation (i.e., self–observer agreement) should be higher than any of the other values in its row or column within the  $4 \times 4$  heteromethod block—that is, the values in bold font should be higher than the values appearing in underline. This criterion is met with the exception of self-emotions appraisal. Fourth, there appear to be common method variance challenges, in that the italicized numbers are generally larger than the degree of observer–self agreement. This suggests that raters are imperfect in their ability to distinguish among the four factors within the umbrella of EI. However, also as discussed above, this could indicate a veridical property of the EI construct. In particular, there should be nonnegative correlations among multiple factors of intelligence (Gutman & Levy, 1991), and this positive manifold should apply to both self- and observer ratings. Given the findings described above, this MTMM analysis of EI ratings shows fair construct validity.

**Predictive validity of observer ratings.** Research in EI has been called upon to demonstrate the power of measures to predict effectiveness for real-world outcomes of interest. Table 3 summarizes the results of partial correlations that examine observer

Table 2  
Multi-Trait Multi-Method Analysis of Emotional Intelligence (EI) Ratings (Studies 1, 2, and 3)

| Variable                              | $\alpha$ | 1            | 2            | 3            | 4            | 5            | 6            | 7             | 8            | 9            | 10          | 11           | 12           | 13           | 14           |
|---------------------------------------|----------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|--------------|--------------|-------------|--------------|--------------|--------------|--------------|
| a. Study 1 (N = 931)                  |          |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| Self-rated Emotional Intelligence     |          |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 1. Self-emotions appraisal            | .84      |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 2. Other-emotions appraisal           | .86      | <i>.42**</i> |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 3. Use of emotion                     | .76      | <i>.33**</i> | <i>.20**</i> |              |              |              |              |               |              |              |             |              |              |              |              |
| 4. Regulation of emotion              | .90      | <i>.36**</i> | <i>.32**</i> | <i>.34**</i> |              |              |              |               |              |              |             |              |              |              |              |
| 5. Total Self-rated EI                | .87      | <i>.71**</i> | <i>.69**</i> | <i>.61**</i> | <i>.78**</i> |              |              |               |              |              |             |              |              |              |              |
| Observer-rated Emotional Intelligence |          |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 6. Self-emotions appraisal            | .92      | <b>.08*</b>  | <b>.08*</b>  | <b>.05</b>   | <b>.00**</b> | .11**        |              |               |              |              |             |              |              |              |              |
| 7. Other-emotions appraisal           | .93      | <i>.06</i>   | <b>.14**</b> | <b>.00</b>   | <i>.12**</i> | <i>.12**</i> | <i>.65**</i> |               |              |              |             |              |              |              |              |
| 8. Use of emotion                     | .84      | <i>.02</i>   | <i>-.06</i>  | <b>.24**</b> | <b>.03</b>   | <i>.06*</i>  | <i>.37**</i> | <i>.29**</i>  |              |              |             |              |              |              |              |
| 9. Regulation of emotion              | .93      | <b>.05</b>   | <b>.06</b>   | <i>-.02</i>  | <b>.25**</b> | <i>.14**</i> | <i>.48**</i> | <i>.59**</i>  | <i>.19**</i> |              |             |              |              |              |              |
| 10. Total Observer-rated EI           | .93      | <i>.06*</i>  | <i>.08*</i>  | <i>.08*</i>  | <i>.17**</i> | <b>.15**</b> | <i>.79**</i> | <i>.86**</i>  | <i>.59**</i> | <i>.78**</i> |             |              |              |              |              |
| b. Study 2 (N = 135)                  |          |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| Self-rated Emotional Intelligence     |          |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 1. Perceiving emotion                 | .74      |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 2. Use of emotion                     | .87      | <i>.22*</i>  |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 3. Understanding emotion              | .87      | <i>.37**</i> | <i>.30**</i> |              |              |              |              |               |              |              |             |              |              |              |              |
| 4. Managing emotion                   | .67      | <i>.45**</i> | <i>.01</i>   | <i>.25**</i> |              |              |              |               |              |              |             |              |              |              |              |
| 5. Total Self-rated EI                | .85      | <i>.73**</i> | <i>.38**</i> | <i>.75**</i> | <i>.74**</i> |              |              |               |              |              |             |              |              |              |              |
| Observer-rated Emotional Intelligence |          |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 6. Perceiving emotion                 | .78      | <b>.17*</b>  | <b>.14</b>   | <b>.05</b>   | <b>.09</b>   | .15          |              |               |              |              |             |              |              |              |              |
| 7. Use of emotion                     | .57      | <b>.07</b>   | <b>.19*</b>  | <b>.14</b>   | <i>-.14</i>  | <i>.06</i>   | <i>-.01</i>  |               |              |              |             |              |              |              |              |
| 8. Understanding emotion              | .82      | <b>.09</b>   | <b>.20*</b>  | <b>.27**</b> | <b>.06</b>   | <i>.21*</i>  | <i>.50**</i> | <i>.21*</i>   |              |              |             |              |              |              |              |
| 9. Managing emotion                   | .74      | <b>.04</b>   | <b>.07</b>   | <i>-.03</i>  | <b>.28**</b> | .15          | <i>.64**</i> | <i>-.27**</i> | <i>.33**</i> |              |             |              |              |              |              |
| 10. Total Observer-rated EI           | .86      | .13          | .19*         | .11          | .19*         | <b>.22*</b>  | <i>.83**</i> | <i>.08</i>    | <i>.70**</i> | <i>.85**</i> |             |              |              |              |              |
| Ability-tested Emotional Intelligence |          |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 11. Perceiving emotion                | —        | <b>.06</b>   | <b>.00</b>   | <b>.14</b>   | <b>.03</b>   | .09          | <i>-.07</i>  | <b>.08</b>    | <b>.07</b>   | <i>-.11</i>  | <i>-.04</i> |              |              |              |              |
| 12. Use of emotion                    | —        | <b>.11</b>   | <b>-.01</b>  | <b>.00</b>   | <b>.12</b>   | .08          | <i>-.02</i>  | <b>.00</b>    | <b>.04</b>   | <i>-.03</i>  | <i>.00</i>  | <i>.61**</i> |              |              |              |
| 13. Understanding emotion             | —        | <b>.02</b>   | <b>-.10</b>  | <b>.15</b>   | <i>-.02</i>  | .04          | <b>.02</b>   | <i>-.01</i>   | <b>.16</b>   | <i>-.01</i>  | <i>.05</i>  | <i>.28**</i> | <i>.26**</i> |              |              |
| 14. Managing emotion                  | —        | <b>.27**</b> | <b>.04</b>   | <b>.13</b>   | <b>.30**</b> | <i>.28**</i> | <b>.10</b>   | <b>.17</b>    | <b>.16</b>   | <b>.04</b>   | <i>.14</i>  | <i>.25**</i> | <i>.34**</i> | <i>.27**</i> |              |
| 15. Total Peer-rated EI               | —        | .15          | .00          | .15          | .13          | <b>.16</b>   | <i>-.01</i>  | <i>.08</i>    | <i>.14</i>   | <i>-.05</i>  | <b>.04</b>  | <i>.84**</i> | <i>.80**</i> | <i>.55**</i> | <i>.57**</i> |
| c. Study 3 (N = 299)                  |          |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| Self-rated Emotional Intelligence     |          |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 1. Perceiving emotion                 | .72      |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 2. Use of emotion                     | .75      | <i>.04</i>   |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 3. Understanding emotion              | .94      | <i>.37**</i> | <i>.23**</i> |              |              |              |              |               |              |              |             |              |              |              |              |
| 4. Managing emotion                   | .75      | <i>.52**</i> | <i>-.08</i>  | <i>.27**</i> |              |              |              |               |              |              |             |              |              |              |              |
| 5. Total Self-rated EI                | .82      | <i>.71**</i> | <i>.36**</i> | <i>.74**</i> | <i>.75**</i> |              |              |               |              |              |             |              |              |              |              |
| Observer-rated Emotional Intelligence |          |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 6. Perceiving emotion                 | .83      | <b>.14*</b>  | <b>.13*</b>  | <b>.11*</b>  | <b>.12*</b>  | .19**        |              |               |              |              |             |              |              |              |              |
| 7. Use of emotion                     | .67      | <i>-.03</i>  | <b>.27**</b> | <b>.06</b>   | <i>-.05</i>  | <i>.07</i>   | <i>.12*</i>  |               |              |              |             |              |              |              |              |
| 8. Understanding emotion              | .89      | <b>.02</b>   | <b>.17**</b> | <b>.11</b>   | <b>.06</b>   | <i>.13*</i>  | <i>.68**</i> | <i>.21**</i>  |              |              |             |              |              |              |              |
| 9. Managing emotion                   | .87      | <b>.07</b>   | <i>-.04</i>  | <i>-.03</i>  | <b>.15**</b> | <i>.07</i>   | <i>.74**</i> | <i>-.09</i>   | <i>.52**</i> |              |             |              |              |              |              |
| 10. Total Observer-rated EI           | .90      | .07          | .11          | .06          | .12*         | <b>.14*</b>  | <i>.89**</i> | <i>.24**</i>  | <i>.79**</i> | <i>.89**</i> |             |              |              |              |              |
| Ability-tested Emotional Intelligence |          |              |              |              |              |              |              |               |              |              |             |              |              |              |              |
| 11. Understanding emotion             | .69      | .03          | .06          | <b>.08</b>   | <b>.07</b>   | .10          | <i>-.04</i>  | <i>-.03</i>   | <b>.12*</b>  | <i>-.06</i>  | <i>-.02</i> |              |              |              |              |
| 12. Managing emotion                  | .85      | <i>-.03</i>  | .03          | <b>.10</b>   | <b>.13*</b>  | .11          | <i>-.04</i>  | <i>.08</i>    | <b>.11</b>   | <i>-.04</i>  | <i>.01</i>  | <i>.59**</i> |              |              |              |

Note. Bold typeface refers to “monotrait-heteromethod” coefficients (i.e., observer-self agreement, observer-ability agreement, self-ability agreement). Italicized typeface, appearing in a triangle, refers to “monomethod-heterotrait” coefficients (i.e., convergence among distinct branches of EI). Underlined typeface refers to “heterotrait-heteromethod” coefficients (i.e., challenges to divergent validity across branches of EI). \*  $p < .05$ . \*\*  $p < .01$ ; all values two-tailed.

ratings of effectiveness as predicted by observer-rated and self-rated EI. For analyses of observer-rated EI, we used bootstrap simulations (Efron, 1979; Efron & Tibshirani, 1993), which eliminate same-source bias in that both EI and performance were rated by the same observers (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). In each of 10,000 bootstrap iterations, we randomly sampled a focal individual, one teammate to provide observer ratings

of performance, and a separate teammate to provide observer ratings of EI. All of the reported partial correlations control for demographic background, cognitive intelligence, personality, and trait affectivity. This allows us to examine the incremental predictive validity of EI above and beyond those factors. As seen in Table 3, higher observer-rated EI significantly predicts higher effectiveness ratings, which held across ratings of effectiveness as

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3  
*Partial Correlations Between Emotional Intelligence (EI) and Performance Ratings, Controlling for Demographic Variables, Personality Traits, and Cognitive Intelligence (Study 1)*

| Measure                  | Performance ratings from teammates |             |          |             |          |             |
|--------------------------|------------------------------------|-------------|----------|-------------|----------|-------------|
|                          | Individual                         |             | Teammate |             | Leader   |             |
|                          | <i>M</i>                           | 95% CI      | <i>M</i> | 95% CI      | <i>M</i> | 95% CI      |
| a. Observer-rated EI     |                                    |             |          |             |          |             |
| Self-emotions appraisal  | .09**                              | (.03, .16)  | .11**    | (.04, .17)  | .11**    | (.05, .17)  |
| Other-emotions appraisal | .12**                              | (.06, .19)  | .16**    | (.10, .22)  | .14**    | (.08, .21)  |
| Use of emotion           | .25**                              | (.19, .31)  | .23**    | (.17, .29)  | .17**    | (.11, .23)  |
| Regulation of emotion    | .08**                              | (.01, .15)  | .11**    | (.04, .17)  | .10**    | (.04, .17)  |
| Total observer-rated EI  | .16**                              | (.10, .23)  | .19**    | (.12, .25)  | .16**    | (.10, .23)  |
| b. Self-rated EI         |                                    |             |          |             |          |             |
| Self-emotions appraisal  | .04                                | (-.04, .11) | .04      | (-.03, .11) | .00      | (-.06, .06) |
| Other-emotions appraisal | .02                                | (-.07, .10) | .05      | (-.03, .13) | .04      | (-.03, .11) |
| Use of emotion           | .16**                              | (.09, .23)  | .13**    | (.05, .20)  | .08*     | (.01, .15)  |
| Regulation of emotion    | -.02                               | (-.09, .05) | .02      | (-.06, .09) | -.04     | (-.11, .04) |
| Total self-rated EI      | .06                                | (-.02, .14) | .08*     | (.00, .16)  | .03      | (-.04, .10) |

*Note.*  $N = 931$  individuals in 186 teams. Estimates for observer-rated EI are based on bootstrap simulations that sample one observer for performance ratings and a distinct observer for EI ratings, in order to reduce same-source bias. The estimates and significance levels of the bootstrap simulations are based on 10,000 samples. These partial correlations control for age, gender, native English speaker, positive affect, negative affect, extraversion, conscientiousness, agreeableness, neuroticism, openness to experience, and total Graduate Management Admissions Test scores.

\* $p < .05$ . \*\* $p < .01$ ; all values two-tailed.

a teammate, individual, and leader. It is worth noting that there were significant zero-order correlations between self-rated EI and performance (see Table S1 in supplemental materials). However, unlike observer-rated EI, for self-rated EI these associations became nonsignificant after including control variables, as seen in the present Table 3. Baseline models of control variables appear in Table S2 in supplemental materials).

#### Measurement equivalence of self- and observer ratings.

Table 4 presents the results of analyses of measurement equivalence between self- and observer ratings of EI. Establishing mea-

surement equivalence involves comparing a series of nested multigroup confirmatory factor analysis models, which provide increasing constraints on the comparison of CFA solutions for self-reported data versus observer-reported data. The baseline model (configural invariance) examines to what extent there is the same pattern of factor loading across both observer ratings and self-ratings, while allowing the values of these factor loadings to differ. The nested weak invariance model adds a constraint of equal factor loadings. The strong invariance model adds a further constraint of equal intercepts. The final model adds the constraint

Table 4

*Analysis of Measurement Equivalence Between Observer-Reported and Self-Reported Emotional Intelligence (Studies 1, 2, and 3)*

| Confirmatory factor analysis                               | $\chi^2$ | <i>df</i> | CFI  | Change in CFI | RMSEA | BIC      |
|--|----------|-----------|------|---------------|-------|----------|
| a. Study 1   |          |           |      |               |       |          |
| Model 1: Configural Invariance                             | 2428.0   | 196       | .972 | —             | .060  | 227472.3 |
| Model 2: Weak Invariance (equal loadings)                  | 2754.1   | 208       | .969 | .004          | .063  | 227693.4 |
| Model 3: Strong Invariance (equal loadings and intercepts) | 3029.7   | 220       | .965 | .003          | .064  | 227864.1 |
| Model 4: Equal loadings, intercepts, and means             | 3248.0   | 224       | .963 | .003          | .066  | 228047.4 |
| b. Study 2   |          |           |      |               |       |          |
| Model 1: Configural Invariance                             | 538.6    | 196       | .954 | —             | .057  | 60531.3  |
| Model 2: Weak Invariance (equal loadings)                  | 561.9    | 208       | .952 | .002          | .056  | 60470.7  |
| Model 3: Strong Invariance (equal loadings and intercepts) | 632.8    | 220       | .944 | .008          | .059  | 60457.7  |
| Model 4: Equal loadings, intercepts, and means             | 717.4    | 224       | .933 | .011          | .064  | 60514.4  |
| c. Study 3   |          |           |      |               |       |          |
| Model 1: Configural Invariance                             | 1213.7   | 196       | .918 | —             | .086  | 78108.7  |
| Model 2: Weak Invariance (equal loadings)                  | 1267.9   | 208       | .914 | .003          | .085  | 78075.8  |
| Model 3: Strong Invariance (equal loadings and intercepts) | 1395.6   | 220       | .905 | .009          | .087  | 78116.6  |
| Model 4: Equal loadings, intercepts, and means             | 1606.4   | 224       | .888 | .017          | .094  | 78298.3  |

*Note.* CFI = Comparative Fit Index; RMSEA = Root mean square error of approximation; BIC = Bayesian Information Criterion.

of equal mean values for the two types of measurement. The accepted criterion for comparing the relative fit is the change in fit using the CFI index, for which a value smaller or equal to  $-0.01$  indicates there is no significant difference between the two measurements (Cheung & Rensvold, 2002; Joseph & Newman, 2010). This criterion was met for each nested model, indicating support for measurement equivalence.

## Discussion

This study evaluated systematically the social perception of EI via observer ratings. It is the first to examine the topic using the SRM (Kenny, 1994), which is based in the psychology of interpersonal perception, and is the authoritative method for examining this topic.

The results provided evidence for some positive psychometric properties, albeit of modest effect sizes. There was consensus across raters in their judgments, which suggests that there is likely to be some kind of underlying signal about EI on which observers agreed. Second, there were significant levels of self-observer agreement. Third, the criteria outlined by Campbell and Fiske (1959) for MTMM models of construct validity were modestly met. In a test of predictive validity, higher observer-rated EI predicted better performance ratings, even when controlling for cognitive intelligence, personality, and trait affectivity. By contrast, self-rated EI did not have predictive validity above control variables. The bootstrap methodology corrected for the potential artifacts of common method bias due to response tendencies or idiosyncratic liking between individuals.

It is noteworthy that the effect sizes appeared to vary across the four EI components represented in the study. The self-emotions appraisal branch had lower consensus among raters and lower self-observer agreement than did the other branches. We speculate that this component—referring to the intrapersonal awareness of one's own emotional states—may be associated with less visible cues that observers may use to form their judgments. Last, self- and observer ratings appear to satisfy the criteria for measurement equivalence.

## Study 2

Study 2 expands the body of evidence in several ways. First, it increases the generality of the findings beyond the specific questionnaire items used by adapting a different EI self-report instrument for observer reports. Study 1 started the investigation with the widely cited Wong and Law (2002) scale. No single scale appears to include all of the components of EI that have been covered in theoretical models—and in this case, it was worth additionally including the nonverbal expression of emotion and managing the emotions of others (e.g., Mayer, Salovey, & Caruso, 2000; Tett et al., 2005). Expanding generality is also worthwhile, as described earlier, because the use of emotion scale has questions that are double-barreled or overlapping with the self-efficacy construct. Second, student teams may not generalize directly to other types of observers. The teams in Study 1 interacted intensively and had tangible course grades at stake, and the individuals were MBA students who had significant job experience and concern about maintaining a positive professional reputation while engaging in team tasks

where performance relatively visible to their peers. Even so, Study 2 increases external validity by sampling ratings from members of participants' real-world full-time workplaces and by soliciting criterion variables from their actual supervisors. Further, we examine the influence of observer-rated EI not only on traditional job performance, but also on other workplace factors that have been theorized as relevant to ability EI, such as transformational leadership and organizational citizenship behaviors (Côté & Miners, 2006; cf., Day & Carroll, 2004). Past work on self-reported EI has found positive associations with ratings of leadership and organizational citizenship, albeit substantially more so for self-reported versus ability measures of EI (Carmeli & Josman, 2006; Harms & Credé, 2010; Modassar & Singh, 2008). As such, it is worthwhile to examine the associations between these factors and observer-reported EI.

## Method

**Participants.** Students enrolled in a part-time MBA program in the Midwest United States completed study measures as part of a leadership course and consented to allow their data to be used for research purposes. Students in this program typically maintained full-time employment in professional-level positions at corporations that helped to sponsor their part-time studies and most had leadership roles with at least one supervisee. Participants solicited peer and supervisor ratings from colleagues at their workplace. A total of 135 focal individuals completed all measures. A total of 1,291 people contributed data to the study, including 328 supervisors (average of 2.4 each) and 828 peers (average of 6.1 each).

Participants were asked to invite only knowledgeable others to rate them. This criterion is important to increase the validity of ratings (John & Robins, 1993; Kolar et al., 1996), and so data were only included from observers who answered "How closely have you worked with this person?" with a three or higher on a five-point scale (1 = *a little*; 5 = *very closely*). This criterion eliminated 74 raters.

**Measures.** Before the course began, participants completed a 1.5-hr battery of self-rated and ability-tested measures, and sent invitations to peers and supervisors to complete 5–10 minute questionnaires each.

**Self-ratings of EI.** Brackett et al.'s (2006) 19-item Self-Report Emotional Intelligence Scale (SREIS) was written and validated to correspond to the Mayer and Salovey (1997) EI model. This model includes the following four branches: (a) *Perceiving Emotion* (the ability to identify emotions in oneself and others), (b) *Using Emotion* (the ability to harness feelings toward cognition), (c) *Understanding Emotion* (the ability to use language and thinking to analyze emotion), and (d) *Management of Emotion* (the ability to regulate one's own emotions and to influence the emotional states of other people). Brackett et al. (2006) validated the SREIS's construct validity, factor structure, and found moderate but significant convergence with a frequently used EI ability test, the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT, Mayer, Salovey, & Caruso, 2002). Note that the Use of Emotion branch contains two double-barreled questions that are worded with socially desirable responses in the direction that indicates low EI ("I am a rational person and I rarely, if ever, consult my feelings to make a decision," and "I am a rational person and don't like to rely on my feelings to make decisions.").

For this reason, we provide the following analyses that include and exclude these two items.

**Observer ratings of EI.** As in Study 1, we adapted the self-report scale of EI into an observer-rated format (McCrae & Costa, 1987). This was rated by the peers.

**Ability measure of EI.** Following the self-rated measures, participants completed the MSCEIT ability measure of EI (Mayer, Salovey & Caruso, 2002). Mayer, Salovey, Caruso, and Sitarenios (2003) provide reliability and validity data for the test and its resulting factor structure for both consensus and expert based coding. Given that Roberts et al. (2001) documented great similarity between the consensus scores generated by experts and lay people, we present scores based on the latter group in analyses. However, we also conducted all analyses below using expert scoring and found no differences in the results.

**Supervisor-rated job performance ( $\alpha = .77$ ).** Supervisors completed a five-item measure of general job performance from Côté and Miners (2006). Sample items include “How would you characterize your subordinate’s effectiveness in displaying job knowledge and skill?” and “How would you characterize your subordinate’s degree to which they set high standards and strive to meet them?”

**Supervisor-rated transformational leadership ( $\alpha = .88$ ).** Supervisors completed the 22-item questionnaire on transformational leadership behaviors from Herold, Fedor, Caldwell, and Liu (2008). These authors adapted the items from the work of Rubin, Munz, and Bommer (2005) and Podsakoff, MacKenzie, and Bommer (1996). Sample items include “My subordinate provides individuals with new ways of looking at things which are puzzling to them” and “My subordinate seeks new opportunities for our organization.” Although transformational leadership is frequently rated by subordinates rather than supervisors, subordinates did not take part in the students’ professional development exercise.

**Supervisor-rated organizational citizenship behaviors (OCBs;  $\alpha = .94$ ).** Supervisors completed a 16-item scale of OCBs from Lee and Allen (2002). This instrument asked about the frequency of OCBs directed at helping individuals within the organization and at helping the organization as a whole. Sample items include “Willingly give his or her time to help others who have work-related problems” and “Show genuine concern and courtesy toward coworkers, even under the most trying business or personal situations.”

**Demographic background.** As in Study 1, participants’ self-reported gender, age, and whether English was their first language were used as control variables.

**Personality scales.** Participants completed self-report scales of Trait PA and Trait NA using the Positive and Negative Affect Scale (PANAS; Watson, Clark, & Tellegen, 1988). The Big Five traits of Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness were tested using the Ten Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003). The TIPI has relatively low internal consistency because the two-item scales for each trait were designed to maximize content validity with nonredundant content. It shows high convergent validity with other widely used Big Five measures and very good test–retest reliability ( $r = .72$  across traits; Gosling et al., 2003).

**Cognitive intelligence.** As a measure of cognitive intelligence, we used participants’ self-reported GMAT scores.

## Results

Table S3 (see supplemental materials) contains bivariate correlations, descriptive statistics, and reliability values for study variables. The structure of the data in Study 2—as opposed to the data in Study 1—did not require the use of the SRM (Kenny & LaVoie, 1984) because each rater provided responses about only one focal individual. Because these data were statistically independent, conventional analyses were used to construct variables based on the average values across all raters.

**Consensus among judges.** There was significant yet modest agreement across raters in making assessments of individuals’ EI. Intraclass correlation (ICC–1) values, representing the average agreement between any two raters, were perceiving emotions,  $ICC(1) = .071$ ; using emotions,  $ICC(1) = .028$ ; understanding emotions,  $ICC(1) = .098$ ; and managing emotions,  $ICC(1) = .142$ .

**Self–observer agreement.** As illustrated in Table 2, self–observer agreement was significant for all four branches, Perceiving ( $r = .17$ ; 95% CI .00–.34), Use ( $r = .19$ ; 95% CI .02–.36), Understanding ( $r = .27$ ; 95% CI .10–.44), and Managing ( $r = .28$ ; 95% CI .11–.45).

**Convergent and divergent validity.** All branches of EI except Use of Emotion met at least modestly the criteria of the MTMM (Campbell & Fiske, 1959). That is, the diagonal values in bold should be statistically significant and larger than any other value in the same row or column. This is the case for all factors except the Use of Emotion branch in the case of one out of four values in the same column.

By contrast with self-reported EI, the ability test of EI showed no appreciable agreement with observer-reported EI. The results of a confirmatory factor analysis of these three sources of data were consistent with this observation made by visual inspection. A model converged but fit with only marginally acceptable statistics,  $\chi^2(33) = 37.95$ , (CFI = .98; unrestricted log likelihood =  $-2035.01$ ; RMSEA = .03; 95% CI .00–.08). There were significant loadings onto the EI trait from self-rated and observer-rated EI (values for Perceiving, Use, Understanding, and Managing of .42/.51, .40/.49, .52/.61, and .85/.33, respectively, for self-rated–observer-rated EI.). The only significant loading from the ability-tested factors of the MSCEIT was Managing (.25), with the remainder nonsignificant (–.16 to .08). There was substantial method variance in all three sources: self-rated EI (.16–.69), observer-rated EI (–.14–.75), and ability-tested EI (.35–.73). (See supplementary materials, Table S4 for more details of these analyses).

**Predictive validity of observer ratings.** Because these data were statistically independent, unlike those in Study 1 using the SRM, these results are based on conventional analyses rather than bootstrapping. The first 3 columns of Table 5 summarize the results of ordinary least squares multiple regression analyses in which observer-rated EI significantly predicts supervisor ratings, while controlling for personality, trait affect, cognitive ability and the other control variables. Each entry in the table is a regression coefficient from a separate model that includes age, gender, native English, Trait PA, Trait NA, Extraversion, Conscientiousness, Agreeableness, Neuroticism, and Openness. Baseline models including control variables appear in Table S2 (see supplemental materials). Because the observer closeness variable (even with its restricted range of three or higher) correlated highly with observer

Table 5

Regression Coefficients Illustrating Predictive Validity for Observer-Rated Emotional Intelligence, Controlling for Demographic Variables, Personality Traits, and Cognitive Intelligence (Studies 2 and 3)

|                        | Study 2   |                             |                 | Study 3                            |                        |   |                        |
|------------------------|---|-----------------------------|-----------------|------------------------------------|------------------------|---|------------------------|
|                        | Performance ratings from supervisors              |                             |                 | Performance ratings from teammates |                        | Performance ratings from teammates controlling for liking |                        |
|                        | Organizational citizenship                        | Transformational leadership | Job performance | Individual effectiveness           | Teammate effectiveness | Individual effectiveness                                  | Teammate effectiveness |
|                        | a. Without controlling for Cognitive Intelligence |                             |                 |                                    |                        |   |                        |
|                        | N = 135   |                             |                 | N = 283                            |                        |   |                        |
| Perceiving emotion     | .16   | .27** (.06)                 | .09             | .49** (.21)                        | .65** (.40)            | .00   | .17** (.01)            |
| Use of emotion         | .22* (.04)  | .21* (.04)                  | .12             | .22** (.04)                        | .30** (.08)            | -.07  | -.03                   |
| Understanding emotion  | .11   | .11                         | .07             | .50** (.21)                        | .57** (.27)            | .17** (.02)   | .18** (.02)            |
| Managing emotion total | .26** (.05)                                       | .34** (.09)                 | .17             | .45** (.20)                        | .66** (.42)            | -.05  | .19** (.02)            |
| Total EI               | .26**   | .34** (.09)                 | .16             | .52** (.25)                        | .70** (.47)            | .02   | .24** (.02)            |
| CI for total EI        | (.07, .45)  | (.16, .53)                  | (-.03, .36)     | (.42, .62)                         | (.61, .79)             | (-.11, .16)   | (.13, .34)             |
| df                     | (1, 125)  | (1, 125)                    | (1, 125)        | (1, 272)                           | (1, 272)               | (1, 271)  | (1, 271)               |
|                        | b. Controlling for Cognitive Intelligence         |                             |                 |                                    |                        |   |                        |
|                        | N = 97  |                             |                 | N = 266                            |                        |   |                        |
| Perceiving emotion     | .16   | .27** (.09)                 | .08             | .52** (.25)                        | .68** (.43)            | .06   | .20** (.02)            |
| Use of emotion         | .17   | .18                         | .15             | .24** (.05)                        | .29** (.08)            | -.05  | -.04                   |
| Understanding emotion  | .08   | .12                         | .08             | .50** (.21)                        | .57** (.29)            | .17** (.02)   | .18** (.02)            |
| Managing emotion total | .19   | .26** (.06)                 | .09             | .48** (.22)                        | .66** (.44)            | .00   | .21** (.02)            |
| Total EI               | .18   | .26** (.07)                 | .09             | .54** (.28)                        | .71** (.49)            | .08   | .26** (.03)            |
| CI for total EI        | (.03, .39)  | (.06, .46)                  | (-.09, .28)     | (.44, .64)                         | (.62, .79)             | (-.06, .21)   | (.15, .36)             |
| df                     | (1, 86)   | (1, 86)                     | (1, 86)         | (1, 254)                           | (1, 254)               | (1, 253)  | (1, 253)               |

Note. For significant coefficients, the incremental  $R^2$  appears in parentheses. Double-barreled questions in the Use of Emotion branch are removed from analysis.

\*  $p < .05$ . \*\*  $p < .01$ ; all values two-tailed.

ratings of EI, we used this closeness rating as an additional control variable in these regression models. This helped to correct for the possible extraneous influence of interpersonal regard. Results show evidence for predictive validity. Higher observer-rated EI significantly predicts better supervisor ratings of transformational leadership and organizational citizenship. Note that inclusion of cognitive intelligence as a control variable reduced the sample size substantially and also reduced the effect sizes for predictive validity, so that only transformational leadership remained significant. The first 3 columns of Table S5a and Table S5b in the supplemental materials report coefficients for self-reported and ability-tested EI. Self-reported EI did not predict performance ratings, and ability-tested EI had unexpected negative associations.

**Measurement equivalence of self- and observer ratings.** Table 4 presents the results of analyses of measurement equivalence between self and observer ratings of EI. The CFA models are described above in Study 1. The criterion was met for each nested model, in that the change in fit using the CFI index was smaller or equal to  $-0.01$ . This indicated support for measurement equivalence.

## Discussion

This study extends the body of evidence for understanding the social perception of EI. First, the same basic pattern of results of the first study generalized to observer reports based on a different self-report instrument, which was developed to follow Mayer and Salovey's (1997) four-factor model of EI. Second, this study had greater external validity in that observer and supervisor ratings were collected from actual colleagues at participants' full-time

jobs. The data included not only job performance, but also leadership and organizational citizenship behaviors, for which a link to EI has been theorized (Côté & Miners, 2006). Indeed, observer ratings of EI predicted organizational citizenship and transformational leadership, above and beyond Big Five personality factors, trait affectivity, and observer closeness to the targets—yet there were no such relationships apparent for self-rated EI. When controlling for cognitive intelligence in the subset of the sample for which data were available, predictive validity for transformational leadership remained significant.

## Study 3

The two studies above, while providing worthwhile data regarding the social perception of EI, benefit from additional evidence to evaluate observer perceptions of EI.

First, it is important to examine whether observer-reported EI is distinct from other observer-reported concepts, notably liking and positive regard, and other positive-valence judgments such as emotional stability that are potentially related to perceptions of EI. These socially desirable interpersonal constructs could be confounded with observer judgments of EI. With the goal of a comprehensive examination in this study, we administer observer reports of liking, positive regard, and the Big Five personality traits. In addition to conducting structural equation models to test directly the divergence of observer-rated EI from these factors, we also include liking and positive regard as control variables in tests of predictive validity.

Second, it is helpful to explore further the possible association between observer ratings and ability-tested EI. Study 2 used the MSCEIT test of EI (Mayer et al., 2002), and Study 3 expands evidence by using the ability tests of EI developed by MacCann and Roberts (2008).

The Situational Test of Emotion Management (STEM; MacCann & Roberts, 2008) examines hypothetical scenarios and scores responses with respect to expert judgments about behavioral responses that are more versus less appropriate. It is interesting to note that the STEM has also been used in an observer-report format, with parents responding about their children (MacCann, Wang, Matthews, & Roberts, 2010). Instead of judging their children's likely scores, as in other studies of EI observer-self agreement, participants provided predictions about the behaviors they believe their children would exhibit. The scores that resulted from these two sets of behavioral responses correlated at .19 (MacCann et al., 2010). This relatively modest overlap between self-rated hypothetical behaviors and the predictions made by closely knowledgeable observers suggests distinct perspectives between the two sources.

The Situational Judgment Test of Emotional Understanding (STEU; MacCann & Roberts, 2008) examines "understanding the relations between, and transitions among, emotions and between emotions and circumstances" (p. 540). Rather than relying on consensus or expert judgments, the test scores participant responses with respect to theoretical predictions from the appraisal model of emotion (Ellsworth & Scherer, 2003; Lazarus, 1991; Roseman, 2001).

## Method

**Participants.** Undergraduate students at a university on the East Coast of the United States completed measures as part of their required introductory course in Management and signed consent forms for research purposes. In this course, students were assigned into student teams averaging 5 members who worked closely throughout the semester. We collected observer ratings in the context of these teams. A total of 299 individuals took part in this study, in 62 teams. Because these measures were used to generate individualized assessments for developmental coaching as a required component of the course, there was missing data only for control variables, which led to 266 for analyses involving controls.

### Measures.

**Self-reported and observer-reported EI.** Respondents completed the same self-ratings and observer ratings of EI as in Study 2, adapted from Brackett et al. (2006). Each person completed ratings of each other team member in a round robin format as in Study 1.

**Ability tests of EI.** The STEU (MacCann & Roberts, 2008) contains 42 items describing situations, and participants select the likely emotions elicited by these situations using multiple-choice responses. The STEM (MacCann & Roberts, 2008) contains 44 items describing situations, with multiple-choice response options.

**Self-reported and teammate reported personality.** Participants completed self-report scales of the Big Five traits of Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness using the TIPI. They also completed peer-report version of the same scale for each teammate in a round robin format.

**Observer ratings of liking and positive regard.** Participants rated their liking and positive regard for each teammate in a round robin format on a scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). For Liking, the following statements were used: "I like my group member as a personal friend" and "I spend time (or would enjoy spending time) socially with my group member." Positive regard used the following statements: "I have strong positive regard for my group member" and "I have deep respect for my group member." These two sets of statements correlated at .93 and so they appear together in the following results.

**Performance ratings from teammates.** For the purpose of developmental feedback, team members rated each other's performance. These ratings were collected at the end of the semester, with feedback aggregated to preserve the privacy of individual respondents. They rated each colleague's performance as an individual and as a teammate on scales ranging from 1 (*not at all effective*) to 7 (*extremely effective*).

**Demographic background.** As control variables, participants self-reported their gender, age, and whether English was their first language.

**Trait affect.** Participants completed measures of Trait PA and Trait NA using the PANAS.

**Cognitive intelligence.** As a measure of cognitive intelligence, we used participants' self-reported Scholastic Achievement Test scores.

## Results

See supplemental Table S6 for bivariate correlations, descriptive statistics, and reliability values for study variables.

**Consensus among judges.** Table 1 presents the results of SRM analyses of EI, liking, and emotional stability. There was significant agreement about the EI of each particular target individual, both for individual branches as well as total EI (ranging from 16% to 23%). There was also significant target variance for liking (24%) and emotional stability (26%), which means that raters also tended to agree regarding who was more versus less likable and who was more versus less emotionally stable.

Note that there was moderate dyadic reciprocity in ratings of EI ( $r = .20, p < .01$ ), which means that a rater who gave particularly high ratings to a target was, in turn, moderately likely to receive high ratings from that target. For liking, however, dyadic reciprocity was substantially higher ( $r = .44, p < .01$ ). This is consistent with past research showing that feelings of liking are particularly mutual (Kenny, 1994). These results document a divergence between the underlying properties of observer-rated EI versus liking.

**Self-observer agreement.** As seen in Table 2, self-observer agreement was significant for the Perceiving ( $r = .14; 95\% \text{ CI } .02-.25$ ), Use ( $r = .27; 95\% \text{ CI } .15-.38$ ), and Managing ( $r = .15; 95\% \text{ CI } .03-.26$ ) branches of EI, but not Understanding Emotion ( $r = .11; 95\% \text{ CI } .01-.23$ ).

**Convergent and divergent validity.** With the exception of the Understanding Emotion branch, the other three branches of EI demonstrate modest convergent and discriminant validity in terms of the criteria of the MTMM (Campbell & Fiske, 1959). For these three factors, the diagonal values in bold are statistically significant and larger than underlined values in the same row or column.

As in Studies 1 and 2, there was generally a positive manifold among the four branches of EI, for both self- and observer ratings.

**Degree of convergence between observer-rated and ability-tested EI.** There was only limited support for convergence between observer-reports and ability tests of EI. Peer ratings of the Emotional Understanding branch correlated significantly with the STEU ( $r = .12, p < .05$ ), particularly when controlling for peer ratings of Liking (partial  $r = .20, p < .01$ ). However, this association appeared to lack divergent validity, in that the observer reports of emotional understanding have essentially the same association with ability-tested emotional understanding and emotion management ( $r = .12$  and  $r = .11$ , respectively). Self-reported Emotional Understanding did not correlate with the STEU ( $r = .08, ns$ ). Peer ratings of the Emotion Management branch did not correlate with the STEM ( $r = -.04, ns$ ), even when controlling for Liking ( $r = -.04, ns$ ). More details appear in Table 2.

**Divergence of observer-rated EI from other observer-rated constructs.** Table 6 summarizes confirmatory factor analyses that support the notion that observer reports of EI are distinct from the related constructs of liking, positive regard, and emotional stability, that is, low neuroticism. The latent factors were allowed to correlate, and these correlation matrices for all models are available upon request from the corresponding author. In each case, model fit was greatly improved by separating these constructs from EI, even though the overall model fit was fair. Also available upon request from the corresponding author are results for similar models that show the divergence of observer-reported EI from the traits of extraversion, agreeableness, openness, and conscientiousness.

**Predictive validity of observer ratings.** Columns 4–6 of Table 5 provided support for the predictive validity of observer-rated EI. The table presents the results of ordinary least squares multiple regression analyses in which observer-rated EI significantly predicts performance ratings from teammates. Each entry in the table is a regression coefficient from a separate model that

includes age, gender, native English, and self-reports for the factors Trait PA, Trait NA, Extraversion, Conscientiousness, Agreeableness, Neuroticism, and Openness. Baseline models including control variables appear in supplemental Table S2. Many coefficients for peer-rated EI are significant predictors of performance ratings whereas, by contrast as reported in columns 4–6 of supplemental Table S5, no coefficients were significant for self-reported EI. After controlling for Liking, the magnitude of the coefficients for observer-rated EI decreased in magnitude, but many remained significant—particularly for teammate versus individual effectiveness. It is interesting to note that the opposite was found for the STEM ability test of managing emotions. After controlling for liking, it predicted better individual but not teammate effectiveness. It is worth noting that the additional explanatory power of these variables was relatively modest. Note that, as mentioned above, these analyses exclude the two double-barreled questions from the Use of Emotion branch. For comparison, results are also presented with these two questions included, which tend to be negative predictors.

**Measurement equivalence of self- and observer ratings.** Table 4 presents the results of analyses of measurement equivalence between self- and observer ratings of EI, as described in more detail in Study 1. The criterion to conclude measurement equivalence was met for every stage of the model except the final constraint, that is the equality of mean values. This indicated evidence for leniency bias in this sample, in that the undergraduates' self-reports were significantly higher than observer reports. The underlying factor structure was otherwise equivalent.

**Discussion**

This study extended the body of evidence for evaluating the social perception of EI. First, it replicated the results of the first two studies demonstrating significant consensus among observers, significant self–observer agreement, and modest fit with criteria

Table 6  
Confirmatory Factor Analysis Models Distinguishing Observer-Rated Emotional Intelligence (EI) From Related Constructs (Study 3)

| Fit statistic               | Distinguishing EI and liking          |   |   | Distinguishing EI and observer-rated neuroticism |  |  |
|-----------------------------|---------------------------------------|---|---|--|--|--|
|                             | Global factor combining EI and liking | Two-factor model separating EI and liking | Five-factor model separating EI four factors and liking | Global factor combining EI and neuroticism       | Two-factor model separating EI and neuroticism | Five-factor model separating EI four factors and neuroticism |
| $\chi^2$                    | 2315.7**                              | 1814.6**                                  | 1300.4**  | 1866.4**   | 1714.2**                                       | 1132.5**   |
| df                          | 252                                   | 251                                       | 242   | 189  | 188  | 179  |
| $\chi^2/df$                 | 9.19                                  | 7.23                                      | 5.37  | 9.88   | 9.12   | 6.33   |
| Baseline $\chi^2$           | 5939.9**                              | 5939.9**                                  | 5939.9**  | 4405.7**   | 4405.7**                                       | 4405.7**   |
| Change in $\chi^2$          | —                                     | 501.2**                                   | 514.2**   | —  | 152.2**  | 581.7**  |
| CFI                         | .64                                   | .72                                       | .81   | .60  | .64  | .77  |
| TLI                         | .60                                   | .70                                       | .79   | .56  | .59  | .73  |
| Log-Likelihood              | -8358.2                               | -8107.6                                   | -7850.5   | -7629.4  | -7553.3  | -7262.4  |
| Unrestricted Log-Likelihood | -7200.4                               | -7200.4                                   | -7200.4   | -6696.2  | -6696.2  | -6696.2  |
| Number of Parameters        | 48                                    | 49  | 58  | 42   | 43   | 52   |
| AIC                         | 16812.4                               | 16313.3                                   | 15817.1   | 15342.7  | 15192.5  | 14628.8  |
| BIC                         | 16990.1                               | 16494.6                                   | 16031.7   | 15498.1  | 15351.6  | 14821.2  |
| BIC2                        | 16837.8                               | 16339.2                                   | 15847.8   | 15365.0  | 15215.3  | 14656.3  |
| RMSEA                       | .17** (.16–.17)                       | .14** (.14–.15)                           | .12** (.11–.13)   | .17** (.17–.18)                                  | .16** (.16–.17)                                | .13** (.13–.14)  |

Note.  $N = 299$ . Numbers in parentheses indicate lower and upper bounds, respectively. Degrees of freedom for the baseline model are 276 for analyses of EI and Liking, and 210 for EI and Neuroticism. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; BIC2 = Bayesian Information Criterion-2; CFI = Comparative Fit Index; RMSEA = Root mean square error of approximation; TLI = Tucker Lewis Index.  
\*  $p < .05$ . \*\*  $p < .01$ ; all values two-tailed.

for construct validity. Next, there was modest and inconsistent evidence for the correspondence between observer-rated EI and performance on a respected ability test of EI. That is, observer-rated Emotion Understanding correlated with ability-tested Emotion Understanding, whereas self-reported Emotion Understanding did not. However, observer-reported management of emotion did not correlate with ability-tested emotion management, whereas self-reported emotion management did. Third, there was a significant distinction between observer-rated EI and the related interpersonal constructs of liking, positive regard, and emotional stability/neuroticism. Confirmatory factor analyses showed better model fit when these constructs were distinct from observer-reported EI. In the case of liking, ratings of EI differed in terms of other fundamental properties. Notably, there is large dyadic reciprocity for liking, which indicates mutual feelings. Reciprocity is a key characteristic identified in the past literature for liking (Kenny, 1994), but it was only modest for observer-rated EI. Taken together, these empirical properties suggest a theoretical distinction between observer-rated EI vis-à-vis observer ratings of related constructs. Further, there was small yet significant predictive validity for observer-rated EI after controlling for liking and positive regard. This was the case for teammate effectiveness, but not individual effectiveness, which suggests that observer-rated EI is more relevant to judgments of performance in interdependent tasks.

### General Discussion

The current work provides the most comprehensive evidence to date about observer ratings of EI. We explored observer perceptions of EI as epiphenomenal—that is, theoretically interesting because they exist inside our heads and have real consequences. Individuals make judgments of other people's emotional abilities on a regular basis, and they act accordingly. In attempting to understand these important judgments, three studies established a systematic body of evidence using authoritative methods from the study of interpersonal perception (Kenny & LaVoie, 1984) and construct validation (Campbell & Fiske, 1959), and replicating patterns of findings in three studies drawing from a total of 2,521 individuals.

In doing so, we mapped out a number of key psychometric properties of these social perceptions. First, observers tended at least modestly to agree with each other, and at least modestly with individuals' own self ratings. Second, observer ratings tended to fit, also modestly, the construct validity criteria of the MTMM. Note that these MTMM analyses were a particularly conservative test because they applied the criteria to distinguishing among the factors of a four-factor model rather than distinguishing observer-rated EI from other constructs. Third, there was consistent evidence for the predictive validity of observer-rated EI, above and beyond trait affect, the Big Five personality traits, cognitive intelligence, age, gender, English as a native language, and closeness to the rater. Even when controlling for liking, observer-rated EI still predicted performance as a teammate, even if it no longer predicted performance as an individual contributor. Indeed, the better predictive power of observer-rated EI for team versus individual performance suggests that the construct is tapping into a form of effectiveness that is interpersonal—rather than intrapersonal. Although this study focused specifically on workplace-relevant cri-

teria in establishing predictive validity, this result highlights the importance of EI in the realm of interpersonal functioning more generally. Fourth, there was good evidence that observer-rated EI is distinct from related judgments that could be considered possible confounding factors, notably liking and emotional stability. In addition to this direct evidence for divergent validity in the form of confirmatory factor analysis, the MTMM findings across the four branches of EI also provided further evidence for divergent validity. That is, if the four factors were distinct from each other, then at least some branches must also be distinct from these confounding factors. If judges were merely guided by liking or even stereotypes about the person high in EI, then these nuisance factors should apply equally across the branches of EI. Further, observer-rated EI had different statistical properties from liking—notably, we replicated the longstanding finding that liking is highly mutual (Kenny, 1994), but much less so for ratings of EI. Last, somewhat surprisingly there was only poor evidence that observer ratings of EI correlate with EI ability tests. We discuss the possible interpretation and meaning of this in the following paragraphs.

### Are Observer Perceptions of EI Valid?

Although we explored observer-rated EI as an epiphenomenal construct—where these ratings are important regardless of whether they reflect an underlying reality—it is also necessary to pose the question of whether observers can actually detect another person's EI. The evidence in this article suggests caution in making this conclusion. First, many positive results that were statistically significant had modest effect sizes, and some confirmatory factor models showed that fit increased with nested models and yet fit was poor overall.

Further, it is important to consider the lack of clear evidence for any convergence between observer-rated EI and ability-tested EI. One would expect to see such associations, and their absence is concerning to the extent one believes that existing ability tests represent the EI construct validly and exhaustively. In making sense of the evidence, it is worth noting that studies using multiple ability tests of EI typically find that these tests do not correlate strongly with each other. In their *Annual Review of Psychology* chapter on EI, Mayer et al. (2008) described this observation as “both perplexing and troubling” (p. 518). As such, observer-rated EI should not necessarily be invalidated due to this lack of clear relationships with ability tests. It would have been helpful to understand the EI construct as a whole if these two forms of measurement did converge, but their nonconvergence does not necessarily implicate observer ratings in particular. The lack of clear evidence does present something of a mystery: the social perception of EI clearly matters in work teams, even if it does not correspond to the ability test measures tested.

It is worthwhile to explore whether there may be moderating conditions under which observer ratings of EI do correspond with ability testing, and potential methods to collect these judgments that could maximize this convergence—and maximize validity as a whole. Notably, observers who are more familiar with the person being rated would have a greater basis of experience from which to draw in making their judgments. Further, those who are themselves high in emotional abilities are likely to be more capable of judging other people's emotional abilities—which has been called a ‘takes one to know one’ effect (Carney & Harrigan, 2003).

In considering the validity of observer ratings, it is noteworthy that their predictive validity exceeded that of both ability-tested and self-reported EI. This is consistent with theoretical development based on Staw et al.'s (1994) three mechanisms for the association between affect and job performance, as discussed above. Individuals high in EI might have better tools to do their work, they may endear themselves to colleagues and receive tangible assistance and support, and they or may perform no better but be perceived as performing better. Given that the second and third mechanisms are mediated by colleagues' views, these colleagues' perceptions are important. As an epiphenomenal construct, the social perception of EI matters because people put it to use.

### Limitations and Future Research

A number of limitations qualify the results of this research, and suggest fertile ground for additional work.

First, the present study adapted self-report questionnaires for use in an observer-rating format, in keeping with norms from the field of personality (McCrae & Costa, 1987). However, in the domain of EI, some areas may be too intrapsychic for observers to be able to report validly. Particularly in Studies 2 and 3, we used a self-report scale that was developed to match the content domain of an EI ability test (MSCEIT, Mayer, Salovey, & Caruso, 2002). The content of this test is limited to areas that are amenable to ability testing, and other areas of EI may not have been emphasized. We suggest that future research on observer-rated EI should not start with a self-report instrument for adaptation. Rather, it should start with the theoretical question of which emotional skills should leave visible behavioral traces in the environment that observers might be able to see.

Second, these studies examined EI only in the context of workplace settings, and EI is crucial for interpersonal interaction in relationships more generally. The primary reason for the focus on the workplace is that most adults spend the majority of their waking hours at work, and for this reason it seemed like an important context to examine. Further, the participants and observers we sampled shared strong acquaintance with each other and had meaningful outcomes to their interactions together. Future work could examine social groups with other types of personal relationships, such as romantic partners, families, dormitories, fraternities and sororities, houses of worship, and military organizations.

Finally, the research presented here was observational rather than experimental. As such, we did not have the opportunity to observe the detailed psychological processes that contribute to observer judgments of EI. Having access to the particular cues that are given off by individuals higher versus lower in EI, as well as the cues utilized by perceivers, would allow a lens model analysis (Brunswik, 1955) for better understanding of the social perception process. Notably, some cues may be used successfully, whereas other cues might involve inaccurate stereotypes or red herrings.

### The Use of Observer Ratings: Promising But No Panacea

Observer ratings of EI could be seen as having a potential practical application in academic research. Measurement chal-

lenges in the field of EI have long been noted (e.g., Matthews et al., 2004), and observer ratings might be seen as a solution to that challenge. It is worthwhile to evaluate this potential choice in the context of the four criteria that Matthews, Zeidner, and Roberts (2002) outlined in their large-scale review and critique of research on EI, which any purported measure of EI should at least minimally satisfy: (a) content validity, (b) scale reliability, (c) construct validity, including convergent and divergent validity, and (d) predictive validity for relevant criteria. Observer reports were found here to satisfy these four requirements to varying extents. As such, using observer reports to measure EI must be done with caution. We note that these criteria are satisfied with statistical significant results, yet often with modest effect sizes, modest model fit, and modest variance explained in predictive models. Along these lines, although there was significant consensus across judges, a relatively large number of judges would be required to create a composite score that approaches conventional levels of total reliability. Based on the effect sizes reported above, it would require at least 12 raters to achieve total reliability in the .90 range for each components of EI, and at least 4 raters for a .70 level of reliability. That said, one of the benefits of observer ratings is that one can aggregate information from as many informants as needed—whereas, by contrast, there is only one self who is available for self-ratings (Paulhus et al., 1998). In any use of observer reports, each rater should ideally rate multiple targets, so that any individual-level rater bias in how people use the response scales can be measured and corrected. To the extent that the results above provide only modest evidence for the validity of observer-reported EI, based on what is known at this point, we suggest that they be used in developmental versus evaluative domains. The data in this study were collected in classroom contexts that emphasized the value of direct and honest feedback for the sake of colleagues' learning, and we do not know how results would generalize to settings that have instrumental stakes for the raters and targets. Future research can enhance our knowledge of when and how observer ratings of EI might be useful as a measurement device.

In this article we argue that people judge each other's social skills and use these judgments, for better or for worse. In today's workplace, such judgments are often explicit hiring criteria, and they are typically made by interviewers on an idiosyncratic rather than systematic basis. In social settings, people use these judgments to steer toward and away from potential relationships. Social perceptions are prevalent, important, and potentially useful. In the case of coaching for professional development and leadership training, it is valuable to know what other people think of you—given that other people are using these impressions. The strongest findings in this article were the consistent results for predictive validity. We found in each study that observer ratings of EI predict performance ratings from both supervisors and peers. Thus individuals can likely benefit from feedback about others' perceptions of them. We conclude that observer judgments of EI are consequential. Taken together, this study offers a substantial body of evidence that better helps us to understand the nature of observer judgments of EI.

### References

- Averill, J. R. (2004). A tale of two snakes: Emotional intelligence and emotional creativity compared. *Psychological Inquiry*, 3, 228–233.

- Bar-On, R., & Handley, R. (2003). *The BarOn EQ-360: Technical manual*. Toronto, Canada: Multi-Health Systems.
- Boyatzis, R. E., & Sala, F. (2004). The Emotional Competence Inventory (ECI). In G. Geher (Ed.), *Measuring Emotional Intelligence* (pp. 147–180). Hauppauge, New York: Nova Science.
- Brackett, M., Mayer, J. D., & Warner, R. M. (2004). Emotional intelligence and its relation to everyday behaviour. *Personality and Individual Differences, 36*, 1387–1402. [http://dx.doi.org/10.1016/S0191-8869\(03\)00236-8](http://dx.doi.org/10.1016/S0191-8869(03)00236-8)
- Brackett, M. A., Rivers, S. E., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: A comparison of self-report and performance measures of emotional intelligence. *Journal of Personality and Social Psychology, 91*, 780–795. <http://dx.doi.org/10.1037/0022-3514.91.4.780>
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*, 193–217. <http://dx.doi.org/10.1037/h0047470>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105. <http://dx.doi.org/10.1037/h0046016>
- Carmeli, A., & Josman, Z. E. (2006). The relationship among emotional intelligence, task performance, and organizational citizenship behaviors. *Human Performance, 19*, 403–419. [http://dx.doi.org/10.1207/s15327043hup1904\\_5](http://dx.doi.org/10.1207/s15327043hup1904_5)
- Carney, D. R., & Harrigan, J. A. (2003). It takes one to know one: Interpersonal sensitivity is related to accurate assessments of others' interpersonal sensitivity. *Emotion, 3*, 194–200. <http://dx.doi.org/10.1037/1528-3542.3.2.194>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511571312>
- Charbonneau, D., & Nicol, A. A. M. (2002). Emotional intelligence and leadership in adolescents. *Personality and Individual Differences, 33*, 1101–1113. [http://dx.doi.org/10.1016/S0191-8869\(01\)00216-1](http://dx.doi.org/10.1016/S0191-8869(01)00216-1)
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255. [http://dx.doi.org/10.1207/S15328007SEM0902\\_5](http://dx.doi.org/10.1207/S15328007SEM0902_5)
- Costa, P. T., Jr., & McCrae, R. R. (1992). Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory. *Psychological Assessment, 4*, 5–13. <http://dx.doi.org/10.1037/1040-3590.4.1.5>
- Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior, 13*, 225–245. <http://dx.doi.org/10.1007/BF00990295>
- Côté, S. (1999). Affect and performance in organizational settings. *Current Directions in Psychological Science, 8*, 65–68. <http://dx.doi.org/10.1111/1467-8721.00016>
- Côté, S., & Miners, C. T. H. (2006). Emotional intelligence, cognitive intelligence, and job performance. *Administrative Science Quarterly, 51*, 1–28.
- D'Augelli, A. R. (1973). The assessment of interpersonal skills: A comparison of observer, peer, and self ratings. *Journal of Community Psychology, 1*, 177–179. [http://dx.doi.org/10.1002/1520-6629\(197304\)1:2<177::AID-JCOP2290010206>3.0.CO;2-5](http://dx.doi.org/10.1002/1520-6629(197304)1:2<177::AID-JCOP2290010206>3.0.CO;2-5)
- Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology, 75*, 989–1015. <http://dx.doi.org/10.1037/0022-3514.75.4.989>
- Day, A. L., & Carroll, S. A. (2004). Using an ability-based measure of emotional intelligence to predict individual performance, group performance, and group citizenship behaviours. *Personality and Individual Differences, 36*, 1443–1458. [http://dx.doi.org/10.1016/S0191-8869\(03\)00240-X](http://dx.doi.org/10.1016/S0191-8869(03)00240-X)
- Downs, A., Strand, P., & Cerna, S. (2007). Emotion understanding in English- and Spanish-speaking preschoolers enrolled in HeadStart. *Social Development, 16*, 410–439. <http://dx.doi.org/10.1111/j.1467-9507.2007.00391.x>
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7*, 1–26. <http://dx.doi.org/10.1214/aos/1176344552>
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall. <http://dx.doi.org/10.1007/978-1-4899-4541-9>
- Elfenbein, H. A. (2007). Emotion in organizations: A review and theoretical integration. *Academy of Management Annals, 1*, 371–457.
- Elfenbein, H. A., Foo, M. D., White, J. B., Tan, H. H., & Aik, V. C. (2007). Reading your counterpart: The benefit of emotion recognition ability for effectiveness in negotiation. *Journal of Nonverbal Behavior, 31*, 205–223.
- Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, H. Goldsmith, & K. R. Scherer (Eds.), *Handbook of the Affective Sciences* (pp. 572–595). New York, Oxford: New York: Oxford University Press.
- Funder, D. C. (1991). Global traits: A neo-Allportian approach to personality. *Psychological Science, 2*, 31–39. <http://dx.doi.org/10.1111/j.1467-9280.1991.tb00093.x>
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55*, 149–158. <http://dx.doi.org/10.1037/0022-3514.55.1.149>
- Funder, D. C., & West, S. G. (1993). Consensus, self-other agreement, and accuracy in personality judgment: An introduction. *Journal of Personality, 61*, 457–476. <http://dx.doi.org/10.1111/j.1467-6494.1993.tb00778.x>
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A Very Brief Measure of the Big Five Personality Domains. *Journal of Research in Personality, 37*, 504–528. [http://dx.doi.org/10.1016/S0092-6566\(03\)00046-1](http://dx.doi.org/10.1016/S0092-6566(03)00046-1)
- Guttman, L., & Levy, S. (1991). Two structural laws for intelligence. *Intelligence, 15*, 79–103. [http://dx.doi.org/10.1016/0160-2896\(91\)90023-7](http://dx.doi.org/10.1016/0160-2896(91)90023-7)
- Harms, P. D., & Credé, M. (2010). Emotional Intelligence and Transformational and Transactional Leadership: A meta-analysis. *Journal of Leadership & Organizational Studies, 17*, 5–17. <http://dx.doi.org/10.1177/1548051809350894>
- Harris, M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43–62. <http://dx.doi.org/10.1111/j.1744-6570.1988.tb00631.x>
- Herold, D. M., Fedor, D. B., Caldwell, S., & Liu, Y. (2008). The effects of transformational and change leadership on employees' commitment to a change: A multilevel study. *Journal of Applied Psychology, 93*, 346–357. <http://dx.doi.org/10.1037/0021-9010.93.2.346>
- James, L. R., Demaree, R. G., & Wolf, G. (1993).  $r_{wg}$ : An assessment of within-group interrater agreement. *Journal of Applied Psychology, 78*, 306–309. <http://dx.doi.org/10.1037/0021-9010.78.2.306>
- John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: The big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61*, 521–551. <http://dx.doi.org/10.1111/j.1467-6494.1993.tb00781.x>
- John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol. 2, pp. 102–138). New York: Guilford Press.
- Jordan, P. J., & Ashkanasy, N. M. (2005). Emotional intelligence, emotional self-awareness, and team effectiveness. In V. U. Druskat, F. Sala, & G. J. Mount (Eds.), *The impact of emotional intelligence on individual and group performance* (pp. 145–163). Mahwah, NJ: Erlbaum.
- Joseph, D. L., Jin, J., Newman, D. A., & O'Boyle, E. H. (2014). Why does self-reported emotional intelligence predict job performance? A meta-analytic investigation of mixed EI. *Journal of Applied Psychology*.

- Joseph, D. L., & Newman, D. A. (2010). Discriminant validity of self-reported emotional intelligence: A multitrait-multisource study. *Educational and Psychological Measurement, 70*, 672–694. <http://dx.doi.org/10.1177/0013164409355700>
- Kellett, J. B., Humphrey, R. H., & Sleeth, R. G. (2006). Empathy and the emergence of task and relations leaders. *The Leadership Quarterly, 17*, 146–162. <http://dx.doi.org/10.1016/j.leaqua.2005.12.003>
- Kenny, D. A. (1994). *Interpersonal Perception: A Social Relations Analysis*. New York: Guilford Press.
- Kenny, D. A. (1998). SOREMO [computer program]. <http://nw3.nai.net/~dakenny/kenny.htm>
- Kenny, D. A., & LaVoie, L. (1984). The Social Relations Model. *Advances in Experimental Social Psychology, 18*, 141–182. [http://dx.doi.org/10.1016/S0065-2601\(08\)60144-6](http://dx.doi.org/10.1016/S0065-2601(08)60144-6)
- Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgements by the self and knowledgeable others. *Journal of Personality, 64*, 311–337. <http://dx.doi.org/10.1111/j.1467-6494.1996.tb00513.x>
- Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*, 63–82. <http://dx.doi.org/10.3102/00346543075001063>
- Law, K. S., Wong, C. S., & Song, L. J. (2004). The construct and criterion validity of emotional intelligence and its potential utility for management studies. *Journal of Applied Psychology, 89*, 483–496. <http://dx.doi.org/10.1037/0021-9010.89.3.483>
- Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.
- Lee, K., & Allen, N. J. (2002). Organizational citizenship behavior and workplace deviance: The role of affect and cognitions. *Journal of Applied Psychology, 87*, 131–142. <http://dx.doi.org/10.1037/0021-9010.87.1.131>
- Libbrecht, N., Lievens, F., & Schollaert, E. (2010). Measurement equivalence of the Wong and Law Emotional Intelligence Scale across self and other ratings. *Educational and Psychological Measurement, 70*, 1007–1020. <http://dx.doi.org/10.1177/0013164410378090>
- Lopes, P. N., Grewal, D., Kadis, J., Gall, M., & Salovey, P. (2006). Evidence that emotional intelligence is related to job performance and affect and attitudes at work. *Psicothema, 18*(Suppl), 132–138.
- Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*, 280–296. <http://dx.doi.org/10.1037/0021-9010.67.3.280>
- MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion, 8*, 540–551. <http://dx.doi.org/10.1037/a0012746>
- MacCann, C., Wang, L., Matthews, G., & Roberts, R. D. (2010). Emotional intelligence and the eye of the beholder: Comparing self-and parent-rated situational judgments in adolescents. *Journal of Research in Personality, 44*, 673–676. <http://dx.doi.org/10.1016/j.jrp.2010.08.009>
- Matthews, G., Roberts, R. D., & Zeidner, M. (2004). Seven myths about emotional intelligence. *Psychological Inquiry, 15*, 179–196. [http://dx.doi.org/10.1207/s15327965pli1503\\_01](http://dx.doi.org/10.1207/s15327965pli1503_01)
- Matthews, G., Zeidner, M., & Roberts, R. D. (2002). *Emotional intelligence: Science and Myth*. Cambridge, MA: MIT Press.
- Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence, 27*, 267–298. [http://dx.doi.org/10.1016/S0160-2896\(99\)00016-1](http://dx.doi.org/10.1016/S0160-2896(99)00016-1)
- Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology, 59*, 507–536. <http://dx.doi.org/10.1146/annurev.psych.59.103006.093646>
- Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications* (pp. 3–31). New York: Basic Books.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2000). Models of emotional intelligence. In R. J. Sternberg (Ed.), *Handbook of Intelligence* (pp. 396–420). Cambridge, UK: New York: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511807947.019>
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) User's Manual*. Toronto, Canada: MHS Publishers.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2008). Emotional intelligence: New ability or eclectic traits? *American Psychologist, 63*, 503–517. <http://dx.doi.org/10.1037/0003-066X.63.6.503>
- Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*, 97–105. <http://dx.doi.org/10.1037/1528-3542.3.1.97>
- McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81–90. <http://dx.doi.org/10.1037/0022-3514.52.1.81>
- 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. <http://dx.doi.org/10.1037/0033-295X.102.2.246>
- Modassar, A., & Singh, T. (2008). Relationship of emotional intelligence with transformational leadership and organizational citizenship behavior. *International Journal of Leadership Studies, 4*, 3–21.
- Newman, D. A., & Lyon, J. S. (2009). Recruitment efforts to reduce adverse impact: Targeted recruiting for personality, cognitive ability, and diversity. *Journal of Applied Psychology, 94*, 298–317. <http://dx.doi.org/10.1037/a0013472>
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259. <http://dx.doi.org/10.1037/0033-295X.84.3.231>
- Oatley, K. (2004). Emotional intelligence and the intelligence of emotions. *Psychological Inquiry, 15*, 216–221.
- O'Boyle, E. H., Jr., Humphrey, R., Pollack, J., Hawver, T. H., & Story, P. A. (2011). Emotional intelligence and job performance: A meta-analysis. *Journal of Organizational Behavior, 32*, 788–818. <http://dx.doi.org/10.1002/job.714>
- Paulhus, D. L., Lysy, D. C., & Yik, M. S. M. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality, 66*, 525–554. <http://dx.doi.org/10.1111/1467-6494.00023>
- Petrides, K. V., Furnham, A., & Mavroveli, S. (2007). Trait emotional intelligence: Moving forward in the field of EI. In G. Matthews, M. Zeidner, & R. Roberts (Eds.), *The science of emotional intelligence: Knowns and unknowns* (pp. 151–166). New York: Oxford University Press.
- Podsakoff, P. M., MacKenzie, S. B., & Bommer, W. H. (1996). Meta-analysis of the relationships between Kerr and Jermier's substitutes for leadership and employee job attitudes, role perceptions, and performance. *Journal of Applied Psychology, 81*, 380–399. <http://dx.doi.org/10.1037/0021-9010.81.4.380>
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879–903. <http://dx.doi.org/10.1037/0021-9010.88.5.879>
- Rivers, S. E., Brackett, M. A., Salovey, P., & Mayer, J. D. (2007). Measuring emotional intelligence as a set of mental abilities. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *The science of emotional intelligence: Knowns and unknowns* (pp. 230–257). New York: Oxford University Press.
- Roberts, R. D., Zeidner, M., & Matthews, G. (2001). Does emotional intelligence meet traditional standards for an intelligence? Some new data and conclusions. *Emotion, 1*, 196–231. <http://dx.doi.org/10.1037/1528-3542.1.3.196>
- Roseman, I. J. (2001). A model of appraisal in the emotion system:

- Integrating theory, research, and applications. In K. R. Scherer & A. Schorr (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 68–91). New York: Oxford University Press.
- Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS Test*. Baltimore: Johns Hopkins University Press.
- Rubin, R. S., Munz, D. C., & Bommer, W. H. (2005). Leading from within: The effects of emotion recognition and personality on transformational leadership behavior. *Academy of Management Journal*, *48*, 845–858. <http://dx.doi.org/10.5465/AMJ.2005.18803926>
- Sala, F. (2005). *Emotional Competence Inventory: Technical Manual*. Hay Group. Retrieved February 26, 2009, from <http://www.eiconsortium.org/>
- Song, L. J., Huang, G. H., Peng, K. Z., Law, K. S., Wong, C. S., & Chen, Z. J. (2010). The differential effects of general mental ability and emotional intelligence on academic performance and social interactions. *Intelligence*, *38*, 137–143. <http://dx.doi.org/10.1016/j.intell.2009.09.003>
- Staw, B. M., & Barsade, S. G. (1993). Affect and managerial performance: A test of the sadder-but-wiser vs. happier-and-smarter hypotheses. *Administrative Science Quarterly*, *38*, 304–331. <http://dx.doi.org/10.2307/2393415>
- Staw, B. M., Sutton, R. I., & Pelled, L. H. (1994). Employee positive emotion and favorable outcomes at the workplace. *Organization Science*, *5*, 51–71. <http://dx.doi.org/10.1287/orsc.5.1.51>
- Swann, W. B., Jr., Chang-Schneider, C., & Larsen McClarty, K. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist*, *62*, 84–94. <http://dx.doi.org/10.1037/0003-066X.62.2.84>
- Swann, W. B., Jr., Stein-Seroussi, A., & McNulty, S. E. (1992). Outcasts in a white-lie society: The enigmatic worlds of people with negative self-conceptions. *Journal of Personality and Social Psychology*, *62*, 618–624. <http://dx.doi.org/10.1037/0022-3514.62.4.618>
- Tellegen, A. (1982). *Brief manual of the differential personality questionnaire*. Minneapolis: University of Minnesota Press.
- Tett, R. P., Fox, K. E., & Wang, A. (2005). Development and validation of a self-report measure of emotional intelligence as a multidimensional trait domain. *Personality and Social Psychology Bulletin*, *31*, 859–888. <http://dx.doi.org/10.1177/0146167204272860>
- Van der Zee, K., Thijs, M., & Schakel, L. (2002). The relationship of emotional intelligence with academic intelligence and the Big Five. *European Journal of Personality*, *16*, 103–125. <http://dx.doi.org/10.1002/per.434>
- Watson, D., & Clark, L. A. (1991). Self- versus peer ratings of specific emotional traits: Evidence of convergent and discriminant validity. *Journal of Personality and Social Psychology*, *60*, 927–940. <http://dx.doi.org/10.1037/0022-3514.60.6.927>
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. <http://dx.doi.org/10.1037/0022-3514.54.6.1063>
- Wechsler, D. (1944). *The measurement of adult intelligence* (3rd ed.). Baltimore: Williams and Wilkins. <http://dx.doi.org/10.1037/11329-000>
- Wong, C. S., & Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: An exploratory study. *The Leadership Quarterly*, *13*, 243–274. [http://dx.doi.org/10.1016/S1048-9843\(02\)00099-1](http://dx.doi.org/10.1016/S1048-9843(02)00099-1)

Received November 23, 2013

Revision received September 26, 2014

Accepted October 15, 2014 ■

### E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!