Multirater assessment of creative contributions to team projects in organizations

Giovanni B. Moneta a; Teresa M. Amabile b; Elizabeth A. Schatzel b; Steven J. Kramer c

a London Metropolitan University, London, UK  
b Harvard Business School, Cambridge, MA, USA  
c Independent consultant,

First published on: 08 May 2009

To cite this Article  Moneta, Giovanni B., Amabile, Teresa M., Schatzel, Elizabeth A. and Kramer, Steven J. (2010) 'Multirater assessment of creative contributions to team projects in organizations', European Journal of Work and Organizational Psychology, 19: 2, 150 — 176, First published on: 08 May 2009 (iFirst)

To link to this Article: DOI: 10.1080/13594320902815312
URL: http://dx.doi.org/10.1080/13594320902815312

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Multirater assessment of creative contributions to team projects in organizations

Giovanni B. Moneta, 
London Metropolitan University, London, UK

Teresa M. Amabile, and Elizabeth A. Schatzel 
Harvard Business School, Cambridge, MA, USA

Steven J. Kramer 
Independent consultant

This study examined the convergent and construct validity of ratings of individual creative contributions in a team context. A sample of 201 employees and supervisors, working on 26 team projects, completed the NEO-Five Factor Inventory and rated themselves and their teammates monthly on a single item measuring creative contributions to the project. The convergent validity of the ratings was supported because there was consistency among other ratings of the same targets and among different types of ratings (peer, supervisor, and self ratings) of the same targets. The construct validity of the ratings was partly supported because there were positive associations between individuals’ peer-rated creativity and their Extraversion, and between individuals’ self-rated and supervisor-rated creativity and their Openness to Experience. From peers and the self, women had lower creativity ratings than men, but other ratings were not influenced by the gender of the judge or the difference in gender of the target–judge dyad. The implications of these findings are discussed.

Keywords: Creativity; Five-factor model; Gender; Multilevel modelling; Organizational performance; Social relations model.

Creativity is an increasingly important concern for managers and organizational scholars (e.g., Bharadwaj & Menon, 2000). Organizational creativity refers to the generation of novel and useful ideas or products within an
organization, in which the term “product” is broadly defined to include processes, procedures, and services (Amabile, 1988; Woodman, Sawyer, & Griffin, 1993). Most research in the field focuses on individual creativity, and the personal characteristics and work environment factors that influence it (e.g., George & Zhou, 2001; Oldham & Cummings, 1996). The validity of research findings on organizational creativity depends on having valid measures of the construct. However, the validity of creativity measures in organizations has received relatively little attention from organizational researchers.

The methodology for assessing creativity has instead received considerable attention from nonorganizational psychologists over the past 50 years, with methods ranging from paper-and-pencil creativity tests to ranking of the creativity of artists and scientists by domain experts. Psychological research on creativity over the past 20 years has predominantly adopted the consensual assessment technique (Amabile, 1982, 1996), which relies on the consensual judgement of multiple experts. This article applies the consensual assessment technique in organizations and assesses some psychometric properties of individual creativity scores. Because most creative work being done in contemporary organizations occurs in team projects, this article investigates individual creative contributions in a team context.

Organizations are interested in maximizing the creativity of team project outcomes. Recent research suggests that individual and team creativity are dynamically related. For example, Taggar (2002) found that team creativity as rated by an external judge was predicted by both average team member creativity and group-level processes such as team citizenship and communication. Moreover, Pirola-Merlo and Mann (2004) found preliminary evidence in support of a multilevel model in which average team member creativity predicts team creativity at particular time-points, and average team creativity across time-points predict creativity of the team project outcome. As such, the development of valid methods to assess individual creative contributions to the team project can help to understand the dynamic relationships between individual creativity (i.e., individual creative contributions to a team project), team creativity (i.e., creative contributions to a team project achieved by team members acting together), and creativity of team project outcomes.

THE CONSENSUAL ASSESSMENT OF CREATIVITY

The consensual definition of creativity (Amabile, 1982, 1996) states that the most appropriate assessment of the creativity of an idea or product is the consensus of independent experts—that is, people familiar with the domain in question—who are blind to individual characteristics of the author of the idea or product. Therefore, the measurement of a person’s creativity...
depends on experts’ judgements of the extent to which creativity (a property) exists in the person’s products or ideas (objects).

The key assumption underlying the consensual definition and assessment technique of creativity is that although certain thinking processes and personality characteristics might be associated with creativity they are not, themselves, creativity. Ultimately, it is the fruit of those thinking processes and personality dynamics, the actual work produced by the individual, that is, the manifestation of creativity. A second assumption is that although it may be difficult for experts to articulate objective features of products or ideas that serve as markers of creativity, creativity is a quality that experts can recognize when they see it, and something that they can generally agree upon in their independent judgements.

Based on these assumptions, most studies applied the consensual assessment technique employing a single global item to measure creativity (e.g., Dollinger, 2007; Shalley & Perry-Smith, 2001; Zhou & Oldham, 2001). The experts typically rate the individual’s idea or product on a Likert scale indicating the extent to which the idea or product is “creative” relative to other persons’ ideas or products. As such, the experts assess the creativity of a given idea or product within the relevant historical and social context.

APPLICATION OF THE CONSENSUAL ASSESSMENT TECHNIQUE TO ORGANIZATIONAL TEAMWORK

We propose two modifications in order to make the consensual assessment technique applicable to organizational team projects. First, because the final product is a collective entity, raters should be required to disentangle the individual’s work from that of the whole team and evaluate his or her creative performance in contributing to the team product as a whole. Second, because no blind and independent experts are available, multiple quasi-experts with different perspectives—the peers who are members of the same team, the team supervisor, and the self—should be utilized as judges of individual creativity. Because all three sources of information are selective and potentially biased, the ratings provided by the three types of quasi-expert are regarded as imperfect indicators of a common latent construct, that is, the actual individual creative contributions to the team project.

THREE PERSPECTIVES ON INDIVIDUAL CREATIVITY IN A TEAM CONTEXT

Organizational researchers have investigated whether using observers from different perspectives might produce more valid job performance ratings. Borman (1997) argued that observers from different perspectives may
attribute different weights to target behaviour, observe different samples of target behaviour, and focus on different areas of target behaviour. As such, the variance of ratings between perspective groups would not necessarily be idiosyncratic, and might in part capture complementary components of target performance. However, a number of studies have found that the variance attributable to idiosyncrasy is much greater than the variance attributable to perspective (e.g., Mount, Judge, Scullen, Sytsma, & Hezlett, 1998; Scullen, Mount, & Goff, 2000; Viswesvaran, Schmidt, & Ones, 2002). In particular, these studies suggest that the rating variance within peer, supervisor, and self groups is larger than the rating variance between these groups. As such, there would seem to be little benefit in using observers from different perspectives.

Research in the field of creativity has traditionally adopted the product-oriented definition of creativity, and has hence privileged the judgement of experts over the judgement of peers, and has discounted the self perspective. However, some creativity scholars have recently argued in favour of a process-oriented concept of creativity as a learning process that energizes individual behaviour even when it falls short in producing an output that would be appraised as creative by experts (e.g., Beghetto & Plucker, 2006; Moran & John-Steiner, 2003; Runco, 2005). In particular, Beghetto and Kauffman (2007) advance that this form of creativity, that they label “mini-c”, is necessary in the genesis of any idea that is eventually judged to be creative by experts. Mini-c creativity essentially is an intrapsychic experience of novelty, sense of ongoing progress, and meaningfulness that might not always be detected and adequately appraised by others. If the product-oriented and process-oriented views of creativity are complementary, then there should be benefit in using multiperspective ratings of individual creative contributions to the team project.

Based on the lens model of perception and judgement (Brunswik, 1956; Funder & Sneed, 1993), we propose that the peers, supervisor, and self have different perspectives and, thus, access partially overlapping sets of cues of creative performance. Relative to the self perspective, the peers and supervisor are more likely to capture only actual instances of creative work, as opposed to unexpressed creative cognition. However, they are also likely to draw from a narrower pool of instances of creative work done by the individual because they are unaware of much of it. Relative to the peers and supervisor perspectives, the self perspective is likely to draw from a larger pool of instances of creative work done, but it is more likely to depend on intrapsychic events that are not actual instances of creative contributions. Relative to peers, supervisors may be in a more privileged position to fairly compare the work of different individuals under their supervision. However, the peers as a whole may have better first-hand knowledge of the individual’s work, particularly in the development phase of a creative idea, when the individual has not yet proposed it to the supervisor.
Overall, we propose that all three perspectives are partially valid because they may miss certain valid cues of creative performance such as subtle indicators of creative work, and may capture invalid cues such as irrelevant personal characteristics and psychological processes. Moreover, we propose that there should be comparatively more consistency between peer ratings and supervisor ratings of the same targets because both types of judges focus primarily on observable target behaviour, whereas there should be comparatively less consistency between self ratings and other ratings (peer ratings and supervisor ratings) of the same targets because targets are likely to focus on cognitive processes and outcomes that others cannot directly observe.

GOALS OF THIS STUDY

We studied a sample of employees and supervisors working on team projects who completed a personality inventory and rated themselves and their teammates monthly on a single item measuring creative contributions to the project. The goal of the study was to assess the convergent validity of the creativity ratings, their construct validity in respect to personality traits, and method effects linked to gender.

Convergent validity of other ratings

We first examined the convergent validity of the creativity ratings as consistency among other ratings of the same targets. The consensual assessment technique in psychological studies typically generates moderate average interrater correlation coefficients among the experts, on the order of $r = .3$ (Amabile, 1982, 1996). This figure is in line with estimates of consistency among raters for a wide range of judgements, including interpersonal liking (Park & Flink, 1989) and behaviours manifested during social interactions (see review by Kenny, Mohor, & Levesque, 2001).

The other ratings gathered in this study conform to a “round-robin” pattern, in which each team member is both a target and a rater of all other team members. The Social Relations Model (SRM; Kenny & La Voie, 1984) allows partitioning the variance of round-robin ratings into five components: target variance, judge variance, general reciprocity covariance, dyad variance, and error variance. The target variance measures the extent to which the ratings capture true individual differences of the targets, with the caveat that the term “true” should be given a relativistic interpretation, as it is possible that there is perfect consistency among the judges in rating a target and they are all equally wrong. The judge variance measures the extent to which there are individual scaling differences between the judges; in the case of evaluative target characteristics such as “creativity” this
variance component can be mostly attributed to differences in judge leniency. The general reciprocity covariance measures the tendency for highly rated targets to give high ratings to others, and vice versa. The dyad variance measures the extent to which there is positive reciprocity in creativity ratings between specific pairs of individuals. The error variance measures random error. This partitioning allows a composite assessment of convergent validity.

Campbell and Fiske (1959) argued that the evaluation of construct validity should control for method effects, as method variance may inflate the observed relationships among variables measured using a common method. Target variance is a measure of convergent validity, whereas judge, general reciprocity, and dyad variance are measures of distinct method effects due to judges. As such, the convergent validity of other ratings is supported if the target variance is large compared to all the sources of method variance. The application of the Social Relations Model to trait ratings (Kenny, 1994) typically generates moderate target variance (15%), slightly greater judge (20%) and dyad variance (20%), and negligible general reciprocity variance, which is instead pronounced for judgements of interpersonal liking. Paulhus and Reynolds (1995) designed a series of studies of personality perception in order to maximize target variance at the expense of judge variance and other method effects, and achieved estimates of almost 30% for target variance and of less than 11% for judge variance; these studies provide an upper standard point for the application of the consensual assessment technique in the context of organizational teamwork. Therefore, we hypothesize:

**Hypothesis 1:** There will be consistency among observer ratings of the same target of the order of a target variance of 30%.

Convergent validity of different types of ratings

We then examined the convergent validity of the creativity ratings as consistency among different types of ratings (peer, supervisor, and self ratings) of the same targets. In a meta-analysis of performance ratings, Harris and Schaubroeck (1988) found fair correlations between peer ratings and supervisor ratings of employees \((r = .62)\) and moderate correlations between self ratings and supervisor ratings \((r = .35)\), and between self ratings and peer ratings \((r = .36)\). These figures were moderated neither by type of rating (trait versus behavioural) nor by scale format (global vs. dimensional). A subsequent meta-analysis (Conway & Huffcutt, 1997) confirmed that supervisor and peer ratings converge more strongly than either does with self ratings. These studies provide a standard for the application of the
consensual assessment technique in the context of organizational teamwork. Therefore, we hypothesize:

**Hypothesis 2:** Peer, supervisor, and self ratings of the same targets will be positively associated with one another; in particular, (a) the correlation between peer ratings and supervisor ratings will be good ($0.5 < p < 0.7$), (b) the correlation between peer ratings and self ratings will be moderate ($0.2 < p < 0.4$), and (c) the correlation between supervisor ratings and self ratings will be moderate ($0.2 < p < 0.4$).

**Construct validity**

We examined the construct validity (Cronbach & Meehl, 1955) of each of the three types of ratings as the extent to which they tie into a nomological network of related constructs and measures. If peer, supervisor, and self ratings are valid indicators of individual creative accomplishments in project teamwork, they should be related to personal characteristics that are associated with other real-life creative accomplishments, and should be unrelated to personal characteristics that are not associated with other real-life creative accomplishments. This rationale rests on the assumption that the factors influencing individual creative accomplishment in project teamwork do not differ from the factors influencing individual creative accomplishment in other contexts. This assumption is shared within the social-psychological approach to organizational creativity (e.g., Amabile & Tighe, 1993; Woodman et al., 1993).

Because real-life creative accomplishment is associated with personality traits, individual differences in certain traits of the Five-Factor Model (FFM; e.g., Goldberg, 1993) can be used to assess construct validity. In a meta-analysis of scientific and artistic creativity, Feist (1998) found that more creative individuals were higher on Openness to Experience and Extraversion and, in some studies only, lower on Conscientiousness and Agreeableness. Neuroticism tended to be associated with creative accomplishment in artists but not scientists. Openness to Experience was the trait most systematically related to creativity across studies, followed by Extraversion. The effect of Extraversion was entirely attributable to the confidence component, and there was no effect involving sociability.

McCrae (1987) argued that Openness to Experience is conducive to creative accomplishment because it represents a broad tendency towards unconventionality, intellectual curiosity, and imaginativeness. As such, openness may enhance the engagement with open-ended tasks and, thus, increase the likelihood of creative accomplishment. Furthermore, openness may motivate a person to acquire intellectual and divergent thinking skills.
through practice, and the acquired skills may increase the likelihood of creative accomplishment. King, McKee Walker, and Broyles (1996) argued that Extraversion is conducive to creative accomplishment because it is a disposition to experience positive affectivity and self-confidence, and these tendencies facilitate engagement and effectiveness in creative work.

Because findings on Openness to Experience and Extraversion are those that appear most consistently in the literature, we use them as the key indicators of construct validity in our study. The remaining FFM traits are not expected to correlate with creative accomplishment and hence are used to detect method effects linked to target personality. Therefore, we hypothesize:

**Hypothesis 3:** Peer, supervisor, and self ratings (a) will be positively associated with target Openness to Experience and Extraversion, and (b) will not be associated with target Conscientiousness, Agreeableness, and Neuroticism.

**Method effects linked to gender**

The current study investigated a sample of mostly engineers or scientists with less than 25% of women. Most projects involved science- or engineering-based product development. In this context, although there is no evidence of gender differences in creative skills as measured by standardized tests (e.g., Saeki, Fan, & van Dusen, 2001; Shi, Xu, Zhou, & Zha, 1999), if women rated themselves as less creative than men and were rated as less creative than men by others we could not infer that the ratings are biased. As such, no hypothesis is posited on differences in rated creativity between male and female participants.

Nevertheless, the stereotypically masculine nature of the performance tasks may induce subtle forms of gender bias. Two types of gender differences in creativity ratings would reveal the presence of method effects linked to gender: (a) Other ratings may be influenced by the gender of the judge, and (b) other ratings may be influenced by the difference in gender of the target–judge dyad. We could assess these gender differences on peer ratings because each team had peers of both genders, but we could not assess them on supervisor ratings because there was only one supervisor per team and the number of teams was too small to perform comparisons between teams led by leaders of different gender. We hypothesize that no such differences will be observed:

**Hypothesis 4:** Peer ratings of the same targets will not be associated with (a) judge gender and (b) target–judge gender difference.
METHODS

Participants

Twenty-six project teams from seven companies within three industries (consumer products, chemicals, and high-tech) were selected specifically because, according to top managers in the companies, creativity was both possible and desirable in the teams’ projects. Most of the members of each team had the team’s project as their main work assignment. All teams had a designated supervisor who was also an active member of the team. Data were collected from the onset to the end of each project, with a mean of 123.7 days ($SD = 49.1$ days) and a range of 57–256 days. The total number of team members was 267.

A total of 238 team members participated in the study. Ten were members of two different teams that participated in the study at two different points in time; for each of these individuals, we randomly retained the data relative to one team only. A total of 201 participants (177 employees and 24 supervisors) completed both the creativity ratings and the personality inventory, and were retained in the analyses of this study. The average team size for the analysis was 7.7 participants, with a range of 3–16.

Employees were 133 men and 44 women. The mean participant age was 37.6 years ($SD = 10.3$ years) with a range of 22–67 years; the mean tenure in the company was 7.5 years ($SD = 9.1$ years) with a range of 0.4–36 years. The sample was highly educated; 148 employees (83.6%) had received at least a college degree, and 66 (37.2%) held advanced degrees.

Supervisors were 20 men and 4 women. The mean participant age was 41.6 years ($SD = 8.5$ years) with a range of 29–64 years; the mean tenure in the company was 10.7 years ($SD = 10.1$ years) with a range of 0.2–29 years. The sample was highly educated; all supervisors had received at least a college degree, and 17 (70.8%) held advanced degrees.

Assessments

Creativity ratings. At the end of every month, all participants (both employees and team supervisors) rated themselves and their teammates on the item, “Creative contribution to the project during the past month”, with a response scale ranging from 1 (“Very low/very poor”) to 7 (“Very high/very good”), with a mid-point of 4 (“Average”). This round-robin design generated four types of ratings: peer ratings (judges = employees, targets = employees), supervisor ratings (judges = supervisors, targets = employees), self ratings (judges = targets), and employee ratings (judges = employees, targets = supervisors). We also examined the other ratings (judges = employees and supervisors, targets = employees and supervisors). The
employee ratings were included in the analysis of the other ratings but were not analysed separately because of the insufficient number of targets (supervisors).

The data include a total of 4044 other ratings from 201 participants, and 569 self ratings from 195 participants (6 participants did not provide self ratings). The mean number of judges per target was 8.0 \( (SD = 3.4 \text{ judges}) \) with a range of 2–15 judges. The mean number of repeated other ratings per target–judge dyad was 2.5 \( (SD = 1.2 \text{ other ratings}) \) with a range of 1–8 other ratings. The mean number of repeated self ratings per participant was 2.8 \( (SD = 1.4 \text{ self ratings}) \) with a range of 1–8 self ratings.

Aggregate scores of creative contributions to the project were computed for each participant separately for peer, supervisor, and self ratings. The 4044 other ratings were averaged within subject for each judge separately, so that each target had one average rating from each judge. The total number of individual mean other ratings was 1618. Of these, 1284 were peer ratings, 168 were supervisor ratings, and the remaining 166 were employee ratings.

**Personality traits.** At the beginning of the study, all participants completed the NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992) containing 60 items measuring the FFM traits. Items were scored on the standard 5-point NEO scale, ranging from 1 ("Strongly disagree") to 5 ("Strongly agree"). Trait scores were computed by summing up the scores of the constituent items.

**Statistical analysis**

**Overview.** The round-robin ratings were analysed using the Multilevel Social Relations Model (MSRM; Snijders & Kenny, 1999), which extends the original Social Relations Model (SRM; Kenny & La Voie, 1984) and provides maximum likelihood, maximally efficient estimates. In particular, the MSRM allows for missing dyadic observations, testing specialized models such as the one with no general reciprocity, analysing simultaneously multiple groups of different sizes while controlling for the between-groups variance, and including target and judge covariates in the model. All three types of ratings (peer, supervisor, and self ratings) could be simultaneously included in an "MSRM with roles" as discussed by Snijders and Kenny (1999). We tried such a model but failed to achieve convergence of its parameter estimates, probably because of the relatively small number of participants in the supervisor role. Therefore, we have resorted to a separate modelling of the three types of ratings. Other ratings and peer ratings were analysed using the MSRM, whereas supervisor and self ratings were analysed using ordinary multilevel modelling (e.g., Goldstein, 1995).
Models of other ratings and peer ratings. Other ratings and peer ratings were analysed using the three-level MSRM model with a group effect presented by Snijders and Kenny (1999, pp. 748–749) as their Equation 7, in which level 1 is the creativity judgement, level 2 is the target–judge dyad, and level 3 is the team.

The model without covariates of an individual’s creative contribution to the project was:

$$\text{CREATIVITY}_{ijk} = \mu + \text{TEAM}_k + \text{TARGET}_{ik} + \text{JUDGE}_{jk} + \text{DYAD}_{(ij)k} + \text{ERROR}_{ijk}$$

in which, $i$ identifies the target, $j$ the judge, and $k$ the team. CREATIVITY$_{ijk}$ is the average creativity rating received by target $i$ from judge $j$, in which both $i$ and $j$ belong to team $k$. The constant (grand mean) $\mu$ represents the average rating in the population. The team effect TEAM$_k$ represents the tendency of team $k$ to be consistently higher or lower in all judgements compared to the average team. The target effect TARGET$_{ik}$ represents the tendency of target $i$ to be rated consistently higher or lower compared to the average target and, thus, can be regarded as the estimate of the target’s true creativity in the project. The judge effect JUDGE$_{jk}$ represents the tendency of judge $j$ to consistently rate targets higher or lower compared to the average judge and, thus, is the estimate of the judge’s level preference. The dyad effect DYAD$_{(ij)k}$ represents the tendency of judge $j$ to provide a higher or lower rating of target $i$ compared to the average rating that judge $j$ provides to the average target and the average rating that target $i$ receives from the average judge. Finally, ERROR$_{ijk}$ represents random error. The constant $\mu$ is a fixed effect, whereas all the other terms are random effects assumed to vary in the population according to a multivariate normal distribution with null mean vector and variances and covariances to be estimated. Five variance components can be estimated: target variance, judge variance, general reciprocity (i.e., target–judge covariance), dyad variance, and error variance. The between-teams variance is not estimated as a separate component, but it is incorporated and controlled for when calculating the target–judge covariance matrix.

We utilized the model without covariates to test if there was general reciprocity. This was achieved by comparing the goodness of fit of the model in which the target–judge covariance was a free parameter with that of the model in which the target–judge covariance was constrained to be null, and retained the best model. Furthermore, we used the selected model without covariates to estimate the average within-team consistency among the ratings of the same target as the percentage of variance accounted for by target [(target variance)/(target variance + judge variance + dyad variance + error variance)]. If consistency is perfect, target will explain 100% of the
variance, whereas if consistency is null, target will explain 0% of the variance. This estimate of consistency among other ratings of the same target provides a test of Hypothesis 1.

The model with covariates (equation not shown) included the FFM traits of target and judge as well as their gender indicator and the absolute target–judge gender difference as fixed effects. The estimates of target traits provide a test of Hypothesis 3. The estimates of judge gender and target–judge gender difference provide a test of Hypotheses 4a and 4b, respectively. Finally, the inclusion of judge traits in the model allows an exploration of method effects arising from the judges’ personality.

Models of supervisor and self ratings. Supervisor and self ratings were analysed using an ordinary two-level model (e.g., Goldstein, 1995) in which individuals constitute level 1 and teams constitute level 2. The model without covariates of an individual’s creative contribution to the project was:

\[ \text{CREATIVITY}_{ij} = \mu + \text{TEAM}_j + \text{TARGET}_{ij} + \text{ERROR}_{ij} \]

in which, \( i \) identifies the target and \( j \) the team. In the case of supervisor ratings, \( j \) identifies both the team and the judge. In the case of self ratings, \( i \) identifies both the target and the judge. CREATIVITY\(_{ij}\) is the average creativity rating of target \( i \) in team \( j \). The constant \( \mu \) is a fixed effect, whereas the team effect \( \text{TEAM}_j \) and the target effect \( \text{TARGET}_{ij} \) are random effects. Three variance components can be estimated: team variance, target variance, and error variance.

The model with covariates (equation not shown) for self ratings included the target FFM traits and gender as fixed effects. The estimates of target traits provide a test of Hypothesis 3.

The model with covariates (equation not shown) for supervisor ratings included the FFM traits of target and judge as well as their gender indicator and the absolute target–judge gender difference as fixed effects. The estimates of target traits provide a test of Hypothesis 3. The estimates of judge gender and target–judge gender difference provide a test of Hypotheses 4a and 4b, respectively. Finally, the inclusion of judge traits in the model allows an exploration of method effects arising from the judges’ personality.

Assessment of consistency among different types of ratings. The multilevel models without covariates are measurement models that produce empirical Bayes individual estimates of the target effect that are more reliable and valid measures of individual creativity than the observed creativity ratings (see treatment of empirical Bayes estimates in Raudenbush & Bryk, 2002). The individual target effects of the MSRM models of other ratings and peer ratings are measures of individual creative contributions to the team project.
adjusted for (a) judge variance, (b) target–judge covariance, (c) dyad variance, (e) team variance, and (f) error variance. The individual target effects of the ordinary multilevel models of supervisor ratings and self ratings are measures of individual creative contributions to the team project adjusted for (a) team variance and (b) error variance. We assessed the consistency among different types of ratings (peer, supervisor, and self ratings) of the same targets by computing the correlations between the target effects estimated for each individual from the models of the different types of ratings. The estimated correlations provide a test of Hypothesis 2.

Estimation of models. We estimated the multilevel models using the program MLwiN 1.0 (Goldstein et al., 1998) and the MSRM models using Snijders’s (1999) macro SRM.OBE. The significance of the fixed effects was evaluated by the ratio of the point estimate divided by its standard error, which is asymptotically distributed according to a standardized normal distribution (e.g., Goldstein, 1995). The significance of a variance component was evaluated by the chi square of change between the model containing the variance component and the subset model excluding the variance component (e.g., Goldstein, 1995).

RESULTS

Data description

Table 1 shows the means, standard deviations, reliabilities, and correlations of the target individual mean ratings of creative contributions to the project, target trait scores, and target gender. The reliabilities of the creativity ratings are test–retest correlations between the mean first rating and the mean last rating received by each target. The reliabilities of the traits are Cronbach’s alphas. The temporal stability was high for peer ratings and fair for supervisor and self ratings. The three types of ratings had moderate-to-good correlations with one another and modest positive correlations with target Openness to Experience and target Extraversion. From peers and the self, women had lower creativity ratings than men.

Models of other ratings and peer ratings

Table 2 shows the estimates of the fixed and random effects of the MSRM models of other ratings. Model 1 is the model without covariates, which allows for general reciprocity. Model 2 is the simplification of Model 1 with the target–judge covariance for the same individual constrained to be null, which does not allow for general reciprocity. The chi square of change was nonsignificant but close to being significant, \( \chi^2 = 3.38, \ df = 1, \ p < .067, \)
TABLE 1
Means, standard deviations, and intercorrelations of the study variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>X</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Peer ratings of creative contributions to the project&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4.72</td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Supervisor ratings of creative contributions to the project&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4.73</td>
<td>1.11</td>
<td>.46**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Self ratings of creative contributions to the project&lt;sup&gt;a&lt;/sup&gt;</td>
<td>5.08</td>
<td>0.92</td>
<td></td>
<td>.28**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Openness to Experience&lt;sup&gt;b&lt;/sup&gt;</td>
<td>30.26</td>
<td>6.06</td>
<td></td>
<td></td>
<td>.17*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Extraversion&lt;sup&gt;b&lt;/sup&gt;</td>
<td>31.18</td>
<td>6.91</td>
<td></td>
<td></td>
<td></td>
<td>.18*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Conscientiousness&lt;sup&gt;b&lt;/sup&gt;</td>
<td>35.10</td>
<td>6.16</td>
<td>-.06</td>
<td></td>
<td></td>
<td></td>
<td>.13</td>
<td>-.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Agreeableness&lt;sup&gt;b&lt;/sup&gt;</td>
<td>32.70</td>
<td>5.60</td>
<td></td>
<td>-.16*</td>
<td>-.02</td>
<td>-.12</td>
<td>-.00</td>
<td>.17*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Neuroticism&lt;sup&gt;b&lt;/sup&gt;</td>
<td>15.56</td>
<td>7.43</td>
<td></td>
<td>-.03</td>
<td>-.09</td>
<td>-.09</td>
<td>-.18*</td>
<td>-.44**</td>
<td>-.26**</td>
<td>-.25**</td>
<td></td>
</tr>
<tr>
<td>9. Gender (men = 0, women = 1)</td>
<td>N/A</td>
<td>N/A</td>
<td>-.30**</td>
<td>-.05</td>
<td>-.23**</td>
<td>.19*</td>
<td>.16*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Correlation coefficients are point-biserial for Gender and product-moment for all other variables. Reliability estimates (test–retest for creativity ratings and Cronbach’s alpha for traits) are shown in parentheses. n = 168 for supervisor ratings of creative contribution to project, n = 195 for self-ratings of creative contribution to project, and n = 201 for all other variables.

<sup>a</sup>Range of the scale = 1 ("Very low/very poor")–7 ("Very high/very good").

<sup>b</sup>Range of the scale = 12–60.

*<sup>p</sup> < .05, **<sup>p</sup> < .01.
indicating that the more parsimonious Model 2, which assumes that there is no general reciprocity, is slightly preferable. Therefore, the other-rated creativity of an individual does not seem to correlate with that individual’s ratings of all members of the team.

Model 2 estimated the components of creativity rating variance to be 29.8% (.504/1.691) for target, 22.5% (.381/1.691) for judge, 5.7% (.097/1.691) for dyad, and 41.9% (.709/1.691) for error. Model 1, which allows for general reciprocity, generated virtually identical estimates of the variance components. The target variance, which is an estimate of consistency among other ratings of the same targets controlled for method effects, approached

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated fixed and random effects, with standard errors in parentheses, of the Multilevel Social Relations Models (MSRM) of other-rated creative contributions to the project</td>
</tr>
<tr>
<td><strong>Effects</strong></td>
</tr>
<tr>
<td>Random effects</td>
</tr>
<tr>
<td>Target variance</td>
</tr>
<tr>
<td>Judge variance</td>
</tr>
<tr>
<td>Target-judge covariance</td>
</tr>
<tr>
<td>Dyad variance</td>
</tr>
<tr>
<td>Error variance</td>
</tr>
<tr>
<td>Fixed effects</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Target covariates</td>
</tr>
<tr>
<td>Openness to Experience</td>
</tr>
<tr>
<td>Extraversion</td>
</tr>
<tr>
<td>Conscientiousness</td>
</tr>
<tr>
<td>Agreeableness</td>
</tr>
<tr>
<td>Neuroticism</td>
</tr>
<tr>
<td>Gender (men = 0, women = 1)</td>
</tr>
<tr>
<td>Judge covariates</td>
</tr>
<tr>
<td>Openness to Experience</td>
</tr>
<tr>
<td>Extraversion</td>
</tr>
<tr>
<td>Conscientiousness</td>
</tr>
<tr>
<td>Agreeableness</td>
</tr>
<tr>
<td>Neuroticism</td>
</tr>
<tr>
<td>Gender (men = 0, women = 1)</td>
</tr>
<tr>
<td>Target-judge covariates</td>
</tr>
<tr>
<td>Gender difference (no = 0, yes = 1)</td>
</tr>
</tbody>
</table>

Models were estimated on the round-robin ratings including 177 employees and 24 supervisors as judges and targets. “–” means that the corresponding effect was excluded.

*p < .05, **p < .01.
the upper standard point of 30%. Therefore, Hypothesis 1 is supported for other ratings.

Model 3 is the extension of Model 2 with covariates. The estimates of the effects of target Openness to Experience and Extraversion were both positive, but only the second achieved significance. Therefore, for other ratings, Hypothesis 3a is supported only in respect to Extraversion. Target Conscientiousness, Agreeableness, and Neuroticism were not significant predictors of other ratings. Therefore, Hypothesis 3b is supported for other ratings. The estimate of the effect of target gender was significantly negative, indicating that women are rated as less creative than men. The estimate of judge gender was not significant. Therefore, Hypothesis 4a is supported for other ratings. The estimate of the target–judge gender difference was nonsignificant. Therefore, Hypothesis 4b is supported for other ratings. Finally, the estimate of the effect of judge Agreeableness was significantly positive, revealing an unexpected source of bias in judgements.

Models 1 through 3 yielded virtually identical estimates and significance levels when fitted on peer ratings only, excluding the supervisors as both judges and targets (results not shown).

Models of supervisor ratings

Table 3 shows the estimates of the fixed and random effects of the ordinary multilevel models of supervisor ratings. Model 4 is the model without covariates. The team variance was nearly null and nonsignificant, indicating that contextual effects are weak for supervisor ratings.

Model 5 is the extension of Model 4 with target and judge covariates. The estimates of the effects of target Openness to Experience and Extraversion were both positive, but only the first achieved significance. Therefore, for supervisor ratings, Hypothesis 3a is supported in respect to target Openness to Experience but not in respect to target Extraversion. Target Conscientiousness, Agreeableness, and Neuroticism were not significant predictors of supervisor ratings. Therefore, Hypothesis 3b is supported for supervisor ratings. Moreover, none of the judge traits was a significant predictor of supervisor ratings. Finally, the estimate of the effect of target gender was nonsignificant, indicating that supervisors rate women as creative as men.

Models of self ratings

Table 3 also shows the estimates of the fixed and random effects of the ordinary multilevel models of self ratings. Model 6 is the model without covariates estimated on all participants (employees and supervisors). The team variance was not significant, indicating that contextual effects are weak for self ratings.
Model 7 is the extension of Model 6 with covariates. The estimates of the effects of target Openness to Experience and Extraversion were both positive, but only the first achieved significance. Therefore, for self ratings, Hypothesis 3a is supported only for Openness to Experience. Moreover, target Agreeableness and Neuroticism were not significant predictors of self ratings, whereas the estimate of the effect of target Conscientiousness was significantly positive. Therefore, for self ratings, hypothesis 3b is supported only in respect to Agreeableness and Neuroticism. Finally, the estimate of
the effect of gender was significantly negative, indicating that women rate themselves as less creative than men.

Models 6 and 7 yielded virtually identical estimates when fitted on employees only, eliminating supervisors’ self ratings (results not shown).

**Consistency among different types of ratings**

The multilevel models without covariates provided individual estimates of the target effects controlled for various sources of method variance. These estimated target effects were then used to estimate the consistency among different ratings of the same targets. The target effect of peer ratings had a correlation of .56 with the target effect of supervisor ratings and of .32 with the target effect of self ratings. The target effect of supervisor ratings had a correlation of .27 with the target effect of self ratings. All the correlations were significant at least at the $p < .01$ level. Therefore, Hypothesis 2 is supported.

**Post hoc analyses on gender differences in rated creativity**

In an effort to determine if the found gender differences in rated creativity are due to confounders, we examined other demographic variables that might differentiate the performance of men from the performance of women. We found no gender differences in educational attainment, perhaps the demographic variable most likely to be related to creative performance. However, we found that women were younger ($X = 35.5$ years) than men ($X = 38.9$ years) and had a shorter tenure in the company ($X = 5.7$ years) than men ($X = 8.6$ years). However, when age and tenure were added as target covariates to the models of other, peer, and self ratings they were nonsignificant, whereas the gender indicator remained significantly negative (results not shown).

**DISCUSSION**

This study applied the consensual assessment technique of individual creativity (Amabile, 1982, 1996) to the assessment of individual creative contributions to the project in organizational teamwork. The evidence on the convergent and construct validity of the creativity ratings indicates the need for improvements and suggests directions for future research.

**Study findings**

*Convergent validity of other ratings.* The analyses of other ratings (including both employees and supervisors as targets and judges) and peer
ratings (including only employees as targets and judges) estimated the average within-team consistency among ratings of the same targets to be nearly one-third of the total rating variance. This estimate was obtained controlling for method effects attributable to judge, general reciprocity, dyadic reciprocity, and team. As such, the consistency of dyadic creativity ratings exceeds the standards of a wide range of dyadic ratings (e.g., Kenny et al., 2001) and matches the upper standard point of dyadic trait ratings that Paulhus and Reynolds (1995) achieved through study designs that maximize target variance at the expense of judge variance and other method effects. The estimated consistency supports Hypothesis 1 and implies that other ratings and peer ratings of the same targets have satisfactory convergent validity.

Convergent validity of different types of ratings. The analysis of different types of ratings (peer, supervisor, and self ratings) of the same targets estimated the consistency between peer ratings and supervisor ratings to be good, and the consistency between self ratings and other ratings (peer and supervisor ratings) to be moderate. These estimates were obtained after correcting supervisor and self ratings for method effects attributable to team, and after correcting peer ratings for method effects attributable to team, judge, and dyadic reciprocity. As such, peer, supervisor, and self ratings match the standards of the original consensual assessment technique (Amabile, 1982, 1996) and a wide range of job performance ratings with different contents and scale-formats (Harris & Schaubroeck, 1988). The estimated consistency supports Hypothesis 2 and implies that the three types of ratings of the same targets have satisfactory convergent validity.

Construct validity. The analyses of target personality traits indicate that none of the three types of ratings fully captures the construct of individual creative contributions to the team project, and that the three types of ratings measure partially overlapping parts of the same construct. Peer ratings are more positively associated with target Extraversion than with target Openness to Experience, and only the association with Extraversion is significant. Both supervisor ratings and self ratings are more positively associated with target Openness to Experience than with target Extraversion, and only the association with Openness to Experience is significant. Therefore, the associations of target creativity with target Openness to Experience and Extraversion proposed by Hypothesis 3a are not fully supported by any one type of rating.

The analyses of target personality traits indicate that all three types of ratings are unrelated to target Conscientiousness, Agreeableness, and Neuroticism with one exception: Self-rated creativity was positively predicted by Conscientiousness. This unexpected finding suggests the
presence of a bias in the self perspective, in that individuals may misinterpret their hard work, orderliness, and achievement as cues of creative performance. The remaining findings are consistent with Feist’s (1998) meta-analysis. In particular, Feist indicated that Neuroticism is associated with real-life creative accomplishment in artists but not scientists; the lack of association between rated creativity and target Neuroticism found in the present study is consistent in that the majority of study participants were scientists. Therefore, the lack of association of target creativity with target Conscientiousness, Agreeableness, and Neuroticism proposed by Hypothesis 3b is supported by other ratings (peer and supervisor ratings), and partially supported by self ratings.

**Method effects linked to gender.** The analyses of other ratings (peer and supervisor ratings) indicated that rated creativity is not associated with the gender of the judge or with the difference in gender of the target–judge dyad. These findings support Hypothesis 4 and suggest that method effects linked to gender are either absent or small for other ratings.

Nevertheless, the analysis indicated that all three types of ratings (peer, supervisor, and self ratings) are negatively associated with target female gender, although the association was nonsignificant for supervisor ratings. Women are consistently rated as less creative than men by judges of both genders, and they consistently rate themselves as less creative than men. These gender differences held even after controlling for age and organizational tenure, which were lower for women in this sample. There is no evidence of gender differences in creative skills as measured by standardized tests (e.g., Saeki et al., 2001; Shi et al., 1999). Moreover, evidence from studies of self-evaluation of performance suggests that the observed gender difference might be caused by a judgemental bias. Beyer (1990, 1998, 1999) found that in masculine tasks (such as tests of knowledge of sports figures or mathematics) women tend to underestimate their objective performance, whereas men tend to overestimate theirs. Women’s lower self ratings of creative performance observed in this study might be due to the stereotypically masculine nature of the performance tasks, since most projects involved science- or engineering-based product development. Beyer did not investigate gender differences in peer ratings of performance in masculine tasks. Although we cannot rule out the possibility that there are real differences in creativity between genders in this particular domain, it is possible that there is a gender bias in both self and peer ratings. This possibility should be examined in future studies.

**Post hoc method effect linked to judge personality.** In the model of other ratings (peer and supervisor ratings) we found that judges higher on Agreeableness tended to rate all targets as more creative, although the
association was nonsignificant for supervisor ratings. Therefore, judge Agreeableness seems to induce a leniency bias in peer-rated creativity.

Overall assessment. Overall, the evidence on the convergent and construct validity of the creativity ratings is mixed. On the positive side, the convergent validity of other ratings (peer and supervisor ratings) of the same targets is good. Moreover, the convergent validity of different types of ratings (peer, supervisor, and self ratings) of the same targets is satisfactory. Finally, other ratings do not seem to be affected by method effects arising from the gender of the judges or from the difference in gender of the target-judge dyad. On the negative side, none of the three types of ratings correlates positively with both target Openness to Experience and Extraversion, which are the personality traits most robustly associated with real-life creative accomplishment. Moreover, the presence of a gender bias in both self and peer ratings cannot be ruled out at this stage. Finally, a target Conscientiousness bias seems to affect self ratings, and a judge Agreeableness bias seems to affect peer ratings.

Study limitations

This study has four main limitations that should be overcome in future research. First, the proposed extension of the consensual assessment technique used a single global item to measure creativity. In order to improve the assessment, we propose to decompose the construct of individual creative performance into facets, to devise multiple items to tap the facets, and to create a multi-item scale to measure the construct. Formally, this is the direction undertaken by researchers who have developed multi-item behavioural measures of individual creativity (e.g., Oldham & Cummings, 1996; Scott & Bruce, 1994). However, our recommendation is that, in order to keep the distinction between creativity of the work done and personality, the items should not tap behavioural patterns that are descriptive of the creative personality. They should instead tap facets of an employee’s actual creative contributions to the project such as the nature, frequency, level, and impact of his or her creative ideas.

Second, due to the relatively small number of participants in the team supervisor role, this study fitted two different types of models (a three-level MSRM model and an ordinary two-level model) to the three different types of ratings (peer, supervisor, and self ratings). This has led to loss of statistical efficiency and notably the impossibility to estimate the variance accounted for by judge perspective. As such, it is impossible to determine whether the different patterns of peer, supervisor, and self ratings with target Extraversion and Openness to Experience are due to judge perspective over and beyond judge idiosyncrasy. With a larger number of teams and hence
team supervisors, the three types of ratings could be analysed in a single MSRM model and provide an estimate of the judge perspective variance that could be compared with estimates from studies of other job performance constructs (e.g., Scullen et al., 2000; Viswesvaran et al., 2002).

Third, this study measured the focal construct of creativity across multiple rating sources, but has not also measured other work related performance constructs in order to provide an assessment of discriminant validity with respect to the constructs being assessed. As such, it is impossible to determine whether the proposed measures of creativity are empirically distinguishable from other facets of job performance such as quantity, quality or effort. In order to improve the assessment, we propose a design in which raters evaluate multiple performance constructs. With a minimum of two performance constructs and three methods (peer, supervisor, and self perspective), a confirmatory factor model based on the Multitrait–Multimethod Matrix (Campbell & Fiske, 1959) could be estimated, and the predicting role of nomological variables such as the personality measures could be examined with respect to the different performance latent variables. This would allow for a dual assessment of discriminant validity in that there would be evidence for the relationships among the performance factors as well as potentially different patterns of the performance factors with the predictor variables.

Finally, this study measured creative contributions to the project using an absolute scale, whereas the original consensual assessment technique requires judges to rate the creativity of an individual’s idea relative to other persons’ ideas. The implication of using an absolute scale is that judges are free to adopt different standards of comparison. For example, a judge might rate targets relative to their differing professional or personal profiles, whereas another judge might rate targets irrespectively of their profiles. In turn, the variability of standard of comparisons will increase error variance and might introduce nonrandom error. Empirical research is needed to compare the properties of absolute and relative measures of creative contributions to the project. Until then, the findings of this study do not generalize to relative measures of creativity.

Does judge perspective matter?

Prior studies of multirater assessment of various facets of individual job performance have found that the largest chunk of the total rating variance is accounted for by idiosyncratic judge effects, the second-largest chunk is accounted for by target performance, and the third-largest chunk is accounted for by judge perspective; moreover, the variance attributable to judge perspective is practically negligible (e.g., Scullen et al., 2000; Viswesvaran et al., 2002). Nevertheless, no prior study has assessed the
judge perspective in relation to ratings of individual creative contributions to the team project.

The present study did not estimate the variance in creativity ratings that is attributable to differences in peer, supervisor, and self perspective, and hence is not directly comparable to prior studies. Nevertheless, our findings indicate that the nomological associations of target creativity with target Openness to Experience and target Extraversion are not fully captured by any one type of perspective. In terms of the lens model of perception and judgement (Brunswick, 1956; Funder & Sneed, 1993) and the process-oriented approach to creativity (e.g., Beghetto & Kauffman, 2007; Runco, 2005), these findings suggest that peers, supervisors, and the self have somewhat different lenses through which to perceive individual creativity in a team context. As a result, each captures somewhat different valid cues of creativity, and misses somewhat different valid cues of creativity.

Valid cues that are more likely to be produced by extraverted targets include high levels of externally observable energy, engagement, risk taking, confidence in pursuing an endeavour, and confidence in promoting one's own ideas and solutions. Because teammates are frequently exposed to these cues in the process of collaborating with each other on the team project, such cues are likely best captured by these peers. Supervisors who are team leaders, although they are also exposed to these cues, may be somewhat farther removed from them as a result of their leadership positions. And individuals themselves are likely less influenced by their own extraversion because of the other-directed nature of such behaviour.

Valid cues that are more likely to be produced by targets who are open to experience include engaging in divergent thinking, exploring new possibilities, independently acquiring new information, connecting the specific issue at hand with a broader context, being interested in new ideas, seeking out and engaging new experience, and expressing variety in actions and procedures. Because of their intrapsychic nature, these processes are most readily available to the self, and are less accessible to external observers such as peers. Nevertheless, perhaps because the monitoring of subordinates is a key leadership task, supervisors capture valid cues of creative performance that stem from their subordinates’ openness to experience.

Disentangling individual contributions out of a team product

Recent studies suggest that individual job performance in team projects, overall team performance, and team project outcomes are intertwined (e.g., Bliese, 2000; Kozlowski & Klein, 2000; Pirola-Merlo & Mann, 2004; Taggar, 2002). In turn, understanding the relationships between individual- and team-level variables is important in order to either maximize a specific variable at a specific level (e.g., creativity of the team project outcome) or
optimize a pattern of variables located at different levels. However, the attempt to gauge individual-level creative contributions to the team-level processes and outcomes raises issues concerning feasibility and application.

A key issue is to what extent it is technically possible to disentangle individual contributions out of a group product if something beyond individual creativity happens to turn individual contributions into a creative group product. Our application of the MSRM used only one parameter, the team variance, to account for all team-level contextual influences on target creative performance and judge ratings. This limitation can be overcome in future applications. For example, the technique would allow using team-level measures of creativity, assuming the same value for each member of the same team, as predictors of target and judge variance. Moreover, the technique can take as input both individual- and team-level measures that are gathered from the onset to the end of the team project, and hence allow modelling of dynamic relationships between variables located at different levels. As such, there is no obvious upper limit of the extent to which individual creative contributions can be singled out.

Another key issue is whether the consensual assessment of individual creative contributions to the project in organizational teamwork can be integrated in a reward system. There are four main reasons for not undertaking such an application at this stage. First, the evidence on the validity of the individual creativity scores is insufficient in respect to discriminant validity. Second, we have assessed consistency among raters (i.e., the extent to which group member ratings on a common target tend to correlate with one another) but not consensus (i.e., the extent to which group member ratings on a common target are virtually the same and hence interchangeable), which would be desirable in a reward system. Third, the measure of creative contributions to the project utilized in this study is not relative to other persons’ contributions, and hence leaves respondents free to adopt different standards of comparison. As such, the findings of this study do not generalize to relative measures of creativity, which would be desirable in a reward system. Finally, studies that evaluated the effects of various reward systems on group project outcomes found that groups tend to perform better when they are given group rewards relative to when they are given individual or hybrid rewards (e.g., Rosenbaum et al., 1980; Wageman, 1995). Therefore, the consensual assessment of individual creative contributions to the project should be used for research purposes and in order to foster organizational learning.

CONCLUSIONS

The assessment of individual creative contributions to team projects in organizations is important in order to be able to investigate the complex
relationships between individual-level contributions, team-level performance, and team-level outcomes. This study begins to probe the perceptual processes by which observers with different perspectives—the peer, the supervisor, and the self—for impressions of an individual’s creative performance in the context of organizational team projects, and suggests that all three perspectives capture valid and complementary cues of individual creative performance, and each misses somewhat different valid cues of creative performance. Moreover, it provides organizational researchers with an extended consensual assessment technique that produces convergent measures and identifies sources of bias that can be controlled for in hypothesis testing. Finally, it outlines a conceptual-methodological framework that, if appropriately developed, can eventually lead to an assessment technique that has both convergent and discriminant validity. In all, this preliminary study suggests that individual creative performance in team projects may differ from other facets of job performance and hence may require specific forms of assessment and management.

REFERENCES


Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.


