

Running Head: SELF-OTHER AGREEMENT ON COMPETENCY RATINGS

Self-other Agreement on Leadership Competency Ratings:
The Moderating Effects of Rater Perspectives and Rating Ambiguity

Guangrong Dai

Paul Stiles

George Hallenbeck

Kenneth P. De Meuse

Lominger International—A Korn/Ferry Company

This paper has been presented at the
2007 Annual Meeting of the Academy of Management in Philadelphia

Abstract

This study investigated how the degree of rating ambiguity and the rater's direct perspective influenced the self-other rating agreement. Using an archival database of responses ($N=737$), this study found that when a leadership competency was ambiguous, the agreement between self and others was lower than if the competency was concrete. The opportunity for a rater to observe a competency demonstrated by the ratee also has a direct influence on self-other rating agreement on that competency. Overall rating ambiguity and rater's direct perspective explained about 25% of the variance of self-other rating agreement. The theoretical and practical implications were discussed.

Keywords:

Multirater feedback; Self-other agreement; Rating ambiguity; Rater perspective

Multirater feedback interventions have experienced rapid proliferation in many organizations around the world. Nearly all of the Fortune 500 companies have used or intend to use multirater feedback in some form (London & Smither, 1995). The need to understand multirater feedback processes has never been greater. One issue of main concern has been self-other agreement on ratings (Atwater, Waldman, Ostroff, Robie, & Johnson, 2005). As early as 1988, Harris & Schaubroeck called for direct tests of potential moderators that may influence self-other agreement. Unfortunately, our understanding of self-other disagreement has not enhanced much over the years (Smither, London, & Reilly, 2005). Only a few studies in the literature have directly tested moderators on self-other agreement and more often than not, these studies focused on overall leadership performance or a small number of performance dimensions (e.g., Atwater & Yammarino, 1992; Church, 1997; Fletcher & Baldry, 2000). In this study, we propose two factors influencing self-other agreement. More specifically, we tested to what extent self-other agreement was a function of (a) rating ambiguity and (b) the degree of rater direct perspective on leadership competencies being rated.

Review of Current Literature

Researchers have long noticed the discrepancy between self ratings and others' evaluation. Determining the degree of consensus between self ratings and other ratings is a key component of the multirater feedback system because of the potential impact on human resource management outcomes (Yammarino & Atwater, 1997). Self-ratings are inflated for reasons such as self-defensiveness (Holzbach, 1978; Steel & Ovalle, 1984), leniency (Farh & Dobbins, 1989), and attribution bias (DeVader, Bateson, & Lord, 1986; Jones & Nisbett, 1972). The correlations between self and other ratings tend to be low to moderate at best (Harris & Schaubroeck, 1988; Lawler, 1967). Past assumptions view these discrepancies as error variance that should be

reduced or eliminated. This perspective has since shifted to viewing these discrepancies as useful and meaningful information (Tornow, 1993).

Several perspectives have been proposed in the literature to interpret self-other agreement and predict its relationship with performance and other outcomes: self-consistency theory (Korman, 1970, 1976; also see Johnson & Ferstl, 1999), control theory (Carver & Scheier, 1981; also see Atwater et al., 2005), and self-awareness theory (Wegner & Vallacher, 1980; also see Atwater & Yammarino, 1992). Self-consistency theory hypothesizes that people are motivated to behave in ways that are consistent with self-perceptions. When managers receive feedback indicating that others' ratings are lower than their self-ratings, they experience cognitive discrepancy and are motivated to improve their performance in order to reduce that discrepancy. Control theory differs in the sense that individuals are continuously matching their behavior to goals or standards. For example, a manager may have a goal where they want to be seen by others as technically competent. If the manager recognizes that others do not perceive them as technically competent, they are likely to make behavioral changes in an attempt to modify others' perceptions.

Self-consistency theory and control theory explain the motivation for behavioral change in terms of perceived discrepancy. Self-awareness theory, on the other hand, is built upon self-other agreement. Self awareness is an individual difference variable and stems largely from one's ability to self-observe (Wicklund, 1979). Individuals who are self-aware are able to assess others' evaluation of the self and incorporate those assessments into their self-evaluations, and ultimately, into their behavior. A highly self-aware individual has a greater likelihood of agreement with other ratings than someone low in self-awareness (Atwater & Yammarino,

1992). The self-aware theory therefore predicts that individuals whose self ratings are in agreement with others will have superior individual and organizational outcomes.

All three perspectives may be correct to some degree. The performance outcomes following a multirater feedback process likely depend on a complex interplay of factors (Craig & Hannum, 2006). Self-other agreement may be important for some performance and outcome measures, but not for others (Atwater, Ostroff, Yammarino, & Fleenor, 1998). Bono and Colbert (2005) indicated that ratees with high core self-evaluations were more motivated to improve when their self-ratings were discrepant from others but that ratees with low core self-evaluations were more motivated to improve when others' ratings agreed with their own. London and Smither (1995) suggested that the degree of agreement within other rater group also makes differences. For example, disagreement among direct reports may cause a manager to dismiss the ratings as reflecting idiosyncrasies of the raters rather than accept them as reflections of his or her own behavior. If there is agreement within the rater group, the manager will pay more attention to self-other discrepancies and feel more pressure to be able to explain those discrepancies. The effects of self-other agreement have also been found to be moderated by culture (Atwater et al., 2005).

Although these three perspectives are different in their theoretical foundations and predictions, there is no doubt that self-other agreement has important human resource implications (Craig & Hannum, 2006; Yammarino & Atwater, 1997). Self-other agreement was related to ratee's reaction to multirater feedback (Atwater & Brett, 2005); perceived need for change (London & Smither, 1995); performance improvement after feedback (Atwater & Yammarino, 1992; Atwater et al., 2005; Johnson & Ferstl, 1999); leadership effectiveness (Atwater, Rouch, & Fischthal, 1995); mentoring functions and Protégé outcomes (John &

Veronica, 2004); work attitude (Sosik, 2001); assessment center performance (Nowack, 1997); and supervisor rating of job performance, compensation, and organizational level (Ostroff, Atwater, & Feinberg, 2004).

While there have been extensive studies on the outcomes of self-other agreement in the literature (as presented above), empirical research on the predictors of self-other agreement are rare. Currently we have only a limited understanding for these (Smither et al., 2005). A close review of related studies indicate that researchers typically focus their attention on individual characteristics such as gender, age, race, managerial experience, education, personality, and contextual factors such as industry, organizational type, and functional area (Brutus, Fleenor, & London, 1998; Fletcher & Baldry, 2000; Mabe & West, 1982; Ostroff et al., 2004; Wohlers, Hall, & London, 1993). An important factor that has been basically ignored is the content domain of the performance dimensions to be rated. Harris and Schaubroeck (1988) in their meta-analysis noticed that self-other agreement was moderated by job type. Ostroff et al., (2004) recognized that the extent of self-other agreement may be related to the competency dimensions. Existing studies in this field usually focus solely on either the overall leadership performance or a small number of dimensions. Therefore, we have a limited understanding of the relationship between self-other agreement and the characteristics of the leadership performance or competency dimensions.

One exception was Wohlers and London's (1989) study. They investigated how rating difficulty influenced self-other agreement. The current study also proposes that self-other agreement differs across leadership competency dimensions. However, our study is significantly different with Wohlers and London (1989) in three ways. First, we have a much larger sample size and believe it allows for a more reliable estimation of self-other agreement. Second, we used

the intraclass correlation absolute agreement index instead of Pearson's correlation because this index is a function of both self-other rating consistency and consensus (McGrow & Wong, 1996). Ratings by two raters may be highly correlated but can have a significant mean difference. Pearson's correlation coefficient is not sensitive to this mean difference. Third, Wohlers and London (1989) contrasted self ratings with the average of all other (raters from higher level, same level, and lower level) ratings. The current study proposes that raters from different organizational levels have different perspectives on the performance to be rated. In this context we compared self-rating with specific rater groups (bosses and direct reports). Therefore, two moderating factors were tested: rating ambiguity and rater perspective.

The Moderating Effect of Rating Ambiguity

The first moderating factor proposed in this study is rating ambiguity. According to social information-processing theory (Salancik & Pfeffer, 1978), individuals develop an attitude as a function of the information available to them at the time they express the attitude. The content of that expression is affected by the saliency and relevance of the information. When there is no salient and relevant information, egocentric bias is likely to occur.

Harris & Schaubroeck (1988) found that self-other agreement was moderated by job type. Specifically, self-supervisor and self-peer correlations were lower for managerial/professional employees than for blue-collar/service employees. The authors attributed this to the characteristics of the job: managerial/professional jobs are more ambiguous than well defined blue-collar/service jobs. Borman (1979) found that raters rated ratees significantly more accurate on certain performance dimensions than on others. Borman (1979) suggested that accuracy was highest on those dimensions for which performance was the least ambiguous and most consistent. Eichinger & Lombardo (2004) noticed that raters were more likely to say they do not

know or cannot rate clearly on competencies that are harder for many people to see directly (such as career ambition and work/life balance). It was therefore proposed in the current study that self-other agreement was moderated by rating ambiguity. People are more likely to disagree on their ratings when the competencies to be rated are more ambiguous.

We define rating ambiguity as the extent to which certain competencies can be rated upon clear and interpretable signs or standards of evaluation. Less ambiguous are those that are concrete, observable, or behaviorally defined. For example, presentation skills are considered to be less ambiguous because it can be behaviorally defined and observable. High ambiguous are those that are abstract, not observable, value related, or trait defined. For example, the competency 'ethics and values' is highly ambiguous in nature because it is abstract and hard to define behaviorally.

The current study focused on self-boss agreement and self-direct reports agreement. We hypothesized the following:

Hypothesis 1. Self-boss agreement on leadership competencies will be moderated by rating ambiguity. Self and boss would agree more on concrete competencies than on ambiguous competencies.

Hypothesis 2. Self-direct reports agreement on leadership competencies will be moderated by rating ambiguity. Self and direct reports would agree more on concrete competencies than on ambiguous competencies.

The Moderating Effect of Rater's Direct Perspective

The second factor under investigation is the extent to which a rater has a direct perspective on the competency to be rated. "Clearly, an individual's self-perception is, or at least should be, linked to self-rating. Likewise, it would be logical to assume that another person's

perception of the individual being rated should determine the rater's response (Yammarino & Atwater, 1997, p.38).” Rating distortion is likely to happen when raters do not have direct perception of the issues to be rated. For example, raters may infer specific judgments from general impressions (Murphy, Martin, & Garcia, 1982). Laing, Phillipson and Lee (1966) introduced the notion of a spiral of reciprocal perspectives. Direct perspectives are person X's views on issue Z; meta-perspectives are person X's views of person Y's views on issue Z. Comparison between one person's metaperspective and the other person's direct perspective on the same issue indicates understanding or misunderstanding. The implication of this notion on self-other agreement is that to what extent self and others agree on a certain competency will likely be moderated by how much other raters have direct perspective on the competency to be rated. When other raters do not have direct perspective, they may rate upon metaperspectives. The less the other raters have direct perspective on the issues to be rated, the higher probability that a misunderstanding will happen.

Raters from different organizational levels may have varying degrees of direct perspectives on the same competencies because of the nature of their daily interactions with the ratees. For example, boss will have a greater direct perspective for the competency 'boss relationship' than direct reports. On the contrary, direct reports will have a greater direct perspective on the competency 'caring about direct reports' than boss. Self-other agreement therefore should be influenced by other raters' direct perspectives across the leadership competencies.

Hypothesis 3. Self-boss agreement on leadership competencies will be moderated by boss's direct perspective. Self and boss would agree more on competencies

that boss has high degree of direct perspective than on competencies that boss has low degree of direct perspective.

Hypothesis 4. Self-direct reports agreement on leadership competencies will be moderated by direct reports' direct perspective. Self and direct reports would agree more on competencies that direct reports have high degree of direct perspective than on competencies that direct reports have low degree of direct perspective.

We also tested the interaction between the two moderating factors. We hypothesized that self-ratings will be in the highest level of agreement with others' ratings when other raters have high direct perspectives and can rate the competencies unambiguously.

Method

Sample

An archival database of responses to a commercial 360 leadership assessment instrument was used for this study. Criteria of qualification for the study were: the ratees must provide self-ratings and must be rated by bosses and direct reports. The original sample consisted of about 1,200 subjects with self-ratings. Cases with missing values on more than 5 per cent of the rated competencies by any of the rater groups were deleted. The frequency distribution of the missing values to the final data set showed 66% of the subjects having complete values, another 17% with only one missing value, while the remaining 17% of the sample had more than one missing value but less than 5% missing. With such a small number of missing values, we felt comfortable to replace these missing values with the middle point of the rating scale. The final data set consisted of 737 subjects.

For the concern of confidentiality, no information on age, ethnicity, or gender was provided in this archival data for the participants.

The 360 feedback instrument

The commercial 360 feedback instrument consisted of 67 items designed to assess leadership competencies. The theoretical foundation and research on the development of this instrument are summarized in Lombardo & Eichinger (2001, 2003). Items are rated on a 5-point scale indicating how much the statement described by an item is a characteristic of the person to be rated. The test-retest reliability was .75 by self rating and .82 by others rating (Lombardo & Eichinger, 2003).

Statistic Analyses

We operationalized self-other agreement with intraclass correlation (ICC) absolute agreement. As discussed earlier, we use the ICC absolute agreement index because it is sensitive to both rating consistency and consensus (McGrow & Wong, 1996). For the purpose of this study, we contrasted self ratings with boss ratings and average of direct reports ratings. To justify aggregation across raters within the same rater group, there must be enough inter-rater reliability. In the current study, all the participants have multiple direct reports ratings. As reported by Lombardo & Eichinger (2003), the mean ICC for the direct reports rating group across the 67 competencies was .75, adequate to justify aggregation (Van Velsor & Leslie, 1991). The aggregation justification was not a critical concern for the boss rater group because only 6.5% of the subjects has been rated by multiple bosses.

Coding of the Independent Variables

Coding the Competencies. The competencies were coded by three subject matter experts (SMEs) that are familiar with the competencies used in this study. Three coding numbers were

assigned to each of the 67 competencies: rating ambiguity, boss perspective, and direct reports perspective.

Again, we define rating ambiguity as the extent to which the competency can be rated upon clear and interpretable signs or standards of evaluation. The SMEs individually rated all 67 competencies on a scale of 1 – 3 (high, medium & low ambiguity, respectively). Condensed ratings revealed total agreement for 19 competencies, 40 adjacent agreements ((1,1,2); (2, 2, 1); (2,2,3); (3,3,2)), and 8 total disagreements ((3,3,1); (1,2,3); (3,1,1)). Adjacent agreements were coded using the most common response. Total disagreements were discussed until a consensus rating was agreed upon by all three SMEs. In total, 11 competencies were considered highly ambiguous (e.g., personal learning and self-development), 43 medium ambiguity (e.g., career ambition and creativity) and 13 low ambiguity (e.g., customer focus and presentation skills).

Rater's perspective is defined as a rater's ability to rate an individual clearly based on how much interaction, knowledge, or experience they have with the person being rated on a particular competency. The scale again ranged from 1 – 3 (limited perspective, moderate perspective, substantial perspective, respectively). 'Limited perspective' indicated that the person rating an individual did not have a clear perspective on the competency being rated. For example, it is hypothesized that for the competency 'Developing Direct Reports', direct reports would have a better perspective on the ratee than would the boss. 'Substantial perspective' is defined as being quite familiar with the rated individual on the competency to be rated. The degree of rater perspective was coded for two rater groups: boss and direct reports. For 'Boss Perspective' there were 17 unanimous competencies, 43 adjacent agreements, and 7 total disagreements. 'Direct Report Perspective' revealed 30 unanimous, 28 adjacent agreements, and 9 total disagreements. Again, adjacent agreements were coded using the modes and discussion

was conducted until consensus agreements were reached for total disagreements. In total, boss perspective had 18 limited perspectives (e.g., personal learning and developing direct reports), 34 moderate perspectives (e.g., organizing and time management), and 15 substantial perspectives (e.g., decision quality and intellectual horsepower). Direct report perspective yielded 15 limited perspective competencies (e.g., negotiating and self-knowledge), 31 moderate perspectives (e.g., decision quality and functional/technical skills), and 21 substantial perspectives (e.g., directing others and approachability).

Sureness Rating: The 360 degree feedback instrument publisher also provides normative all others sureness ratings for the 67 competencies. The sureness ratings were derived from a normative study using more than 2000 employed adults (Lombardo & Eichinger, 2003). The sureness rating asks raters indicating how sure they are of their ratings. The instrument publisher reported the average of all others sureness rating on each of the 67 competencies. We included this sureness rating in the current study for two purposes: to validate our coding of the independent variables, and to further examine how the sureness rating is related to the self-other agreement. We expected that the sureness rating would be significantly correlated with our coding on the rating ambiguity but not rater's perspective. To explain why, we decompose the sureness rating variance into three components: the variance due to the nature of the competency (e.g., rating ambiguity), the variance due to the rater (e.g., rater's perspective), and error variance.

$$V_{\text{sureness}} = V_{\text{competency}} + V_{\text{rater}} + V_{\text{error}}$$

The variance due to the nature of the competency is systematic. But the variance due to the rater is non-systematic across rater sources. When composite sureness rating across rater sources is calculated, the rater variance component is minimized. In contrast, the variance due to

the nature of the competency is maximized. Rating ambiguity is basically determined by the nature of the competency. But the rater's perspective is more specific to a particular rater source. For this reason, the average sureness rating is expected to be correlated with the subjective coding of rating ambiguity but not rater perspective. In fact, we believe that the average sureness rating to a large extent is the objective measure of rating ambiguity.

Because of the expected relationship between the average sureness rating and the coding of rating ambiguity, we further examined the relationship between the sureness rating and the self-other agreement. It was expected that self-other agreement was moderated by the average sureness rating in that there would be more agreement on competencies raters were surer of their ratings.

Results

Mean ratings and standard deviations by the rater sources (boss, direct reports and self) are reported in Table 1. In the current study, mean self rating across the 67 competencies is lower than the mean direct reports' rating by about one sixth standard deviation. The difference between mean self rating and mean boss' rating is trivial.

Insert Table 1 about here

Table 2 shows descriptive statistics to and the correlations among the independent and dependent variables. The average coding for each of the three independent variables is close to 2, the midpoint of the coding scale. The self-boss ICC and self-direct reports ICC are about the same (ICC=.20), indicating overall the low self-other agreement.

 Insert Table 2 about here

The correlation matrix shows that boss perspective and direct report perspective do not have a significant relationship. This finding demonstrates that our coding of direct perspective yielded variables that did in fact differ with one another. In other words, based on SME coding when considering rater type, certain competencies have varying degrees of rater perspective.

As expected, the SME coding of rating ambiguity is highly correlated with the sureness rating ($r = -.55, p < .001$). Sureness rating is not correlated with SME coding of boss perspective ($r = .19, p > .05$). But unexpectedly, the sureness rating is significantly correlated with the SME coding of direct reports perspective ($r = .32, p < .01$). Later on we will discuss why the sureness rating is not related to the SME coding of boss perspective but related to the SME coding of direct reports perspective.

The ambiguous nature of the competency being rated had a significant relationship with boss perspective ($r = -.43, p < .01$) and direct report perspective ($r = -.37, p < .01$). This suggests that with higher ambiguity of the competency being rated, the less likely there will be an opportunity for direct perspective. The ambiguous nature of the competency was also related to self-direct reports ICC ($r = -.36, p < .01$) and self-boss ICC ($r = -.41, p < .01$). These findings support hypotheses 1 & 2 such that self and boss/direct report agreement is in fact related to the ambiguous nature of the competency being assessed.

Further correlation findings show that self-direct reports ICCs have a relationship with direct report perspective ($r = .45, p < .01$) but not with boss perspective ($r = .03, p = .83$).

Similarly, self-boss ICCs were found to have a relationship with boss perspective ($r = .43, p <$

.01) but not direct reports perspective ($r=.04, p = .72$). This finding supports hypotheses 3 & 4 that suggest self-boss and self-direct report agreement is a function of rater perspective.

As expected, sureness rating is correlated with self-direct reports agreement ($r=.37, p<.01$), suggesting that self and direct reports agree more on competencies where direct reports are surer of their ratings. But sureness rating is not significantly correlated with self-boss agreement ($r=.19, p=.11$). We scatterplot sureness rating against self-agreement and identified one competency, “Managing Diversity”, as an outlier. This competency is high on sureness rating but very low on self-boss agreement ($ICC=.06$). Probably the majority of the ratees did not face the diversity issue and as a result raters could not give reliable ratings even it appears to be a much concrete competency. The zero-order correlation between sureness rating and self-boss agreement is .26 ($p<.05$) after “Managing Diversity” is removed from analysis. In sum, the impacts of the sureness rating on self-other agreement is mostly supported in this study.

The hypotheses were further tested using hierarchical regressions. Because rating ambiguity and rater perspective was significantly correlated, they were centered to mean to produce the interaction terms (Aiken & West, 1991). Rating ambiguity ($\beta = -.28, p < .05$) and boss' direct perspective ($\beta = .31, p < .05$) were significant predictors of self-boss ICC (Table 3). This shows further support for the hypotheses 1 & 3 suggesting that when item ambiguity is high, self-boss agreement will be lowered (vice versa). Similarly, when boss direct perspective is low, self-boss agreement will also be lower (vice versa). No significant interaction effect was found ($\beta = -.04, p = .92$) suggesting that the relationship between rating ambiguity and self-boss agreement is not moderated by boss perspective. Similarly, the relationship between boss perspective and self-boss agreement is not moderated by rating ambiguity.

Insert Table 3 about here

Hierarchical regression on self-direct reports ICC shows similar results for main and interaction effects. As Table 4 shows, item ambiguity ($\beta = -.23, p < .05$) and direct perspective ($\beta = .33, p < .01$) were significant predictors of self-direct reports ICC. Hypothesis 2 & 4 are supported. Finally, no significant interaction effect was found ($\beta = -.02, p = .95$) suggesting that the relationship between rating ambiguity and self-direct reports agreement is not moderated by direct reports perspective. Similarly, the relationship between direct reports perspective and self-direct reports agreement is not moderated by rating ambiguity.

Insert Table 4 about here

Discussion

Little research has been conducted on the moderating role of leadership competency dimension on self-other rating agreement. By using a 360 feedback instrument with a thorough sample of leadership competencies, we were able to examine how self-other agreement varies across the leadership competencies. The current study tested two moderators: rating ambiguity and rater perspective. In sum, the main effects of the moderators were strongly supported, but the interactive effects were not. Based on our findings when a competency is ambiguous in nature the agreement between self and other is lower than if the item was less ambiguous. In addition, the direct perspective a rater has on a ratee influences self and other agreement. In other words the opportunity for a rater to perceive a competency demonstrated by the ratee will have a direct

influence on self and other agreement on that competency. Finally, these two moderators influence self-other agreement independently. Their effects on self-other agreement are additive but not interactive.

The significant and high correlation with the sureness rating provided by the instrument publisher validated the SME coding of the rating ambiguity. As expected, the sureness rating was not significantly correlated with the SME coding of boss perspective. Unexpectedly, the sureness rating was significantly correlated with the SME coding of direct reports perspective. This may be because of the method used to calculate the composite score. As articulated previously, the variance of sureness rating can be decomposed into three components: the variance due to the nature of the competency, rater variance, and error variance. We argued that when average across raters, the rater specific variance will be minimized. But in practice, boss is usually outnumbered by other raters such as peers and direct reports. The simple average is likely to minimize boss specific variance but not the direct reports specific variance. For this reason, the sureness rating is not correlated with the SME coding of boss perspective, but correlated with the SME coding of direct reports perspective. In fact, the sureness rating has a relative higher correlation with the self-direct reports agreement than with the self-boss agreement, providing further evidence that the boss but not the direct reports specific variance is minimized when the sureness rating is averaged across raters.

In the current study the mean of direct report ratings on the 67 leadership competencies was higher than the means of self and boss ratings. The difference between self rating and boss rating was trivial. This appears to be inconsistent with the literature. For example, in Harris and Schaubroeck's (1988) meta-analytic study, self ratings were over half a standard deviation higher than supervisor rating. Using more than six thousands subjects from the U.S. and five European

countries, Atwater et al. (2005) found higher self rating than subordinates rating. It was generally believed that self-ratings tend to be inflated. But Atwater & Yammarino (1992) demonstrated that self-other rating difference resulted from both self- and other evaluations. Atwater & Yammarino (1992) assigned the self-other differences into three categories: over-estimate, in-agreement, and under-estimate. Those categorized as over-estimators rated themselves higher and were rated by others lower than those categorized in-agreement. In contrast, those categorized as under-estimators rated themselves lower and were rated by others higher than those categorized as in-agreement. Self ratings are not always higher than others ratings for all the rates.

Atwater & Yammarino's (1992) method was used to further examine the pattern of self-other disagreement in this study. The results are presented in Table 5 and 6. Atwater & Yammarino's (1992) findings are replicated here. Boss is the most accurate rater, followed by peers and direct reports (Eichinger & Lombardo, 2004). The patterns identified in this study implies that lower performers tend to overestimate themselves, but higher performers are more likely giving self ratings lower than or in agreement with others, echoing conclusions made by others (e.g., Atkins & Wood, 2002; Church, 1997; Lombardo & Eichinger, 2004).

Insert Table 5 and 6 about here

Self rating on average is lower than direct reports rating in this study. One possible reason for this difference is that the participants in general are high performers. The data is from a commercial 360 degree leadership competency feedback instrument. Organizations probably

have their high potentials or high performers but not low performers participate in a 360 degree assessment for development purpose.

In the current study, both self-boss ICC (.20) and self-DRs ICC (.20) are small, indicating in general low self-other agreement. This finding is consistent with the literature. For example, in Harris and Schaubroeck's (1988) meta-analytic study, the mean self-supervisor correlation was .22 (corrected for sampling error). In Atwater et al.'s (2005) study, the observed self-subordinate correlation was .27. ICC is a function of both consistency (as indicated by the Pearson correlation) and consensus (as indicated by the mean difference), and therefore ICC is smaller than Pearson correlation coefficient in general. Because of this, the effect sizes observed in this study are in the typical range of the self-other agreement indicated in the literature.

Theoretical Implication

It is not uncommon for different raters to provide very different ratings about a given manager (Morgeson, Mumford, & Campion, 2005). Much theory has been offered to explain why a comparison of self-ratings with those of others is important for predicting performance ratings, effectiveness, and performance outcomes. Empirical studies, however, showed inconsistent and often conflicting results (Atwater et al., 2005) regarding the impacts of self-other rating agreement. Smither et al. (2005) proposed a theoretical model for understanding performance improvement following multisource feedback. In this model, characteristics of feedback (for example, how self rating is related to others' rating) influence feedback recipients' initial reactions to the feedback, which in turn influence subsequent motivational attitudes and behaviors. One important implication of this model is that multisource feedback should not be expected to lead to substantial performance improvement for all feedback recipients. Initial reactions to feedback (such as affective reactions) can be expected to influence whether feedback

recipients use the feedback to set goals and make performance improvements. For example, reactions immediately following feedback were found to be positively related to feelings that the feedback was useful (Brett & Atwater, 2001). Several characteristics of multisource feedback are likely to influence feedback recipients' initial reactions. How self rating is related to others rating is one of the most obvious characteristics. For example, feedback recipient may construe others' negative ratings as biased or containing errors when they are not congruent with self rating.

Eichinger and Lombardo (2004) stated that it should be coupled with other data to make sense the self other rating agreement patterns. The current study indicated how self other rating agreement across the leadership competencies is moderated by the nature of the competency and rater's experience with the ratee on the issues to be rated. Rating ambiguity and rater's perspective combined together explained about 25% of the variance of self-other rating agreement. One theoretical implication of the finding is that the influence of self other agreement on feedback recipient's initial reactions to the feedback may be moderated by the performance dimension to be rated and the rater source. O'Reilly and Anderson (1980) reported that employees' perceptions of the accuracy of their performance evaluation moderated the relationship between the feedback received and their satisfaction with the appraisal. Feedback recipients react more positively when the feedback is perceived to come from a credible source (Albright & Levy, 1995). The finding of the current study implies that the credibility of a specific rater source should also depend on the performance dimensions. Boss may be perceived as the credible source for the ratings on some leadership competencies, but direct reports may be the credible source for some other leadership competencies. In other words, perceived rating accuracy or usefulness should be a function of both the rater source and the performance dimension.

One method for future research to examine the proposition introduced above is to include a sureness rating in multirater feedback. Raters will be asked to indicate how sure they are of their performance ratings. Researcher can then test the relationships among the three variables: reaction toward the feedback, perception of rating accuracy, and sureness rating. The higher a rater's sureness rating, the higher probably the feedback recipient's perceived rating accuracy of the rater. And the perceived rating accuracy will be expected to influence the feedback recipient's initial reaction toward the feedback. Sureness rating is likely to be a function of both the rating ambiguity and rater's direct perspectives. Raters will be surer of their ratings on performance dimensions that are more concrete and on which the raters have more direct experience with the ratee.

Sureness rating is a valuable piece of information. Lombardo & Eichinger (2004) found that sureness rating has a connection to job performance rating. They divided the raters into three levels of sureness—high, moderate, and low. They then correlated the leadership competency ratings with the job performance rating (obtained two years out) for each of the three group of raters and found that the surer the raters, the higher correlations between the leadership competencies and job performance rating. That is, sureness rating was found to moderate the relationship between leadership competency rating and job performance. One possible reason for the observed moderating effect of sureness rating is that the reliability of leadership competency rating was higher for the group who were surer of their ratings than the group who were less sure of their ratings. Another possible reason might be that the feedback recipients in the high level of sureness group reacted more positively to multirater feedback than those in the low level of sureness group. Including a sureness rating, together with the measures of other variables, researcher will be able to identify the antecedents (e.g., rating ambiguity and rater's perspective)

and the consequences (e.g., perceived rating accuracy and initial reaction to feedback) of sureness rating.

Practical Implications

The robust support for the main effect of rating ambiguity and perspective in this study marks an important evolution in our understanding of the rating process for multirater instruments. To the extent that the differences in agreement in this study occurred as a result of raters' inability to provide clear and direct feedback, the key question becomes "How does the experience of multirater feedback change so that the main effects of ambiguity and indirectness can be reliably reduced and result in 'true' ratings?"

The first approach that might come to mind would be to change the behavioral make-up of the competencies to make them more clearly and directly observable. This certainly represents a practical solution, but amounts to the tail wagging the dog from a construct point of view. Competencies simply are what they are. Some such as composure and listening are clearly and readily observable. Others such as strategic agility are, by nature, more difficult to scrutinize in day-to-day actions of the individual. However, because they are behaviorally distinct and have an important impact on performance in different arenas of work, they merit inclusion and should be evaluated.

Another approach would be to selectively assign competencies to different ratings groups depending on the perceived ability of those groups to clearly and directly observe the person in question. While this would likely lead to increased precision and reduced "noise" in ratings, it would also potentially leave out valid and meaningful information. By selectively eliminating input from bosses, peers, or direct reports, the resulting feedback is not truly multirater in nature.

Also, since input may be sought on just a subset of competencies, the rater's perceived role in the process may be reduced and potentially lead to less motivation to provide feedback.

A third approach that may result in positive outcomes without the compromises of the first two approaches would involve changing the rater experience to decrease the ambiguity and increase the directness of particular competencies. Rater experience can be enhanced in three ways. First, the operational definitions of the competencies could be clarified to provide more explicit mention of observable behaviors. Second, behaviorally-anchored ratings scales could be developed to provide a more specific guide for raters' evaluations. Third, video vignettes or written case studies of competencies could be provided so that raters could review them and better understand how the competency functions "in action" and at differing skill levels.

Directness can also be improved prior to rater evaluations. First, raters could be given a list of situations where each competency is most likely to come into play. This would prime their attention to when the competency may have been applied and how successfully. Second, the person being rated (i.e., the learner) could give guidance on whom to pick for his or her rater panel depending on the number and breadth of experiences that potential raters have witnessed him or her in. It may turn out that some individuals who have spent a lot of time with the learner may have only witnessed a very narrow set of behaviors. Likewise, individuals with comparatively less time may have had a broader set of situations to observe the learner in. Finally, raters could fill out a pre-rating checklist of "rater readiness" that would include a list of the most behaviorally rich settings/circumstances in which they may have had an opportunity to observe the learner. If this checklist is provided well in advance of the scheduled rating period, it may help focus raters' attention in areas that are currently sources of uncertainty.

Enhanced approaches to rater selection, training, and instruction will help make the rater better informed and prepared to provide input. Targeting interventions on decreasing ambiguity and increasing directness should not only have an impact on rating accuracy, but rater confidence and motivation may also increase because the rater may feel more secure and engaged in the process.

The findings have additional implications for interpretation and coaching. Being aware of and understanding the impacts of ambiguity and directness on raters' perceptions is valuable information for feedback coaches given the responsibility to interpret 360 reports. As the feedback coach explores the results with the learner, legitimate opportunities may emerge where the coach can offer the explanation that self and others' ratings differed on a particular competency because it was highly ambiguous or more difficult for that particular group to observe directly.

Similarly, the extent to which the learner agrees with the coaches' assessment and feels that self and others' perceptions differed simply because the raters "haven't *seen* me do that," may open up the possibility for an enhanced developmental plan. Specifically, rather than a traditional plan focused on "improvement" on a competency, the strategy would be to make something that was there all along more visible to the audience; this is referred to as the Marketing Plan path to development (Eichinger, Lombardo, & Stiber, 2005). The plan would focus on creating meaningful opportunities where the learner can clearly demonstrate to others that he or she possesses the capability in question and can apply it effectively. Of course, some potential caution needs to be exercised with these interpretation and coaching interventions. For instance, the acknowledgment of ambiguous or indirect rating circumstances can't be used as a convenient excuse for lacking skill on a particular competency. The feedback coach should be

wise to providing plausible alternatives for rating differences, especially ones that are supported by ratings on other, less difficult to rate competencies. Also, proper attention should be paid that the learner does not try to exhibit the marketed competency at the expense of engaging in more meaningful, high priority job activities.

Data gathering and reporting efforts can also be enhanced. Raters could be asked to provide a confidence rating or an “opportunity to observe” rating following each competency rating. This would allow for expanded interpretation of rating results and refined understanding of the impact of ambiguity and directness on ratings. If persistent gaps are found with certain competencies that appear attributable to ambiguity or directness, a constant may be applied to competency ratings as a form of rater error adjustment. Another option from a reporting standpoint would be to create cluster scores for competencies that are high, moderate or low in ambiguity or directness. This would allow for the opportunity to observe trends in raters’ observations across different categories.

Limitations

The results of this study should be viewed in light of its limitations. The first limitation is the subjective nature of the coding of the moderating variables. The SMEs were in disagreement by definition on their coding on about 13% of the competencies initially. This may reflect the difficulty of coding on certain competencies. Nevertheless, the correlations with the sureness rating provided by the instrument publisher to a great extent validated our subjective coding of the moderating variables. We further believe that our study is an underestimation rather than an overestimation of the moderating effects of these coded variables. Remember we used a 3-category coding: low, moderate, and high. During the coding, we felt hard to put some competencies in one category or another. A middle point between two adjacent categories may

fit these competencies better. If we use more than 3 categories of coding, we may be able to observe even stronger association between the moderators and self-other agreement.

The second limitation is about coding rating ambiguity. Rating ambiguity was coded without concern for rater source. In reality, different raters may have a different perception of rating ambiguity. The same competencies may be considered ambiguous by direct reports but not bosses. Therefore, rating ambiguity may not be independent of rater sources, as indicated by the significant correlations between rating ambiguity and rater perspectives in this study. Nonetheless, we found that rating ambiguity explained beyond rater perspectives the variance of self-other agreement, suggesting the independent effect of this coded characteristic of the leadership competency.

Third, we used the all others average sureness rating to validate our subjective coding of the moderating variables and further examined its relationship with self-other agreement. While the composite score maximized the assessment of the nature of the competencies, meaningful rater specific information was lost. We were not able to examine how self and other agreement was influenced by the sureness rating from a particular rater group. Future research may overcome this weakness by collecting sureness ratings for each of the rater sources.

Finally, we have focused only on self-boss and self-direct reports rating agreement. As organizational trends continue toward a flatter structure, employees are increasingly involved in team work and having more connections with customers. Future research should repeat this study by including other sources of ratings such as peers and customers.

In conclusion, the present research makes a significant contribution to the multirater feedback literature. At a conceptual level, it identified two moderating variables and provided understanding on how self-other agreement varies across leadership competencies. At a more

pragmatic level, it has potential implications on multisource instrument design, rater selection, feedback, coaching, and development plan.

References

- Albright, M. D., & Levy, P. E. 1995. The effects of source and performance rating discrepancy on reactions to multiple raters. *Journal of Applied Social Psychology, 25*: 577-600.
- Atkins, P. W. B., & Wood, R. E. 2002. Self versus others' ratings as predictors of assessment center ratings: Validation evidence for 360 degree feedback programs. *Personnel Psychology, 55*: 871-904.
- Atwater, L. E., & Brett, J. F. 2005. Antecedents and consequences of reactions to developmental 360 degree feedback. *Journal of Vocational Behavior, 66*: 532-548.
- Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. 1998. Self-other agreement: Does it really matter? *Personnel Psychology, 51*: 577-598.
- Atwater, L. E., Rouch, P., & Fischthal, A. 1995. The influence of upward feedback on self- and follower raters of leadership. *Personnel Psychology, 48*: 34-60.
- Atwater, L. E., Waldman, D., Ostroff, C., Robie, C., & Johnson, K. M. 2005. Self-other agreement: Comparing its relationship with performance in the U.S. and Europe. *International Journal of Selection and Assessment, 13*: 25-40.
- Atwater, L. E. & Yammarino, F. J. 1992. Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology, 45*: 141-164.
- Bono, J. E., & Colbert, A. E. 2005. Understanding responses to multi-source feedback: The role of core self-evaluations. *Personnel Psychology, 58*: 171-203.
- Borman, W. C. 1979. Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology, 64*: 410-421.

- Brett, J. F., & Atwater, L. E. 2001. 360 degree feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology*, 86: 930-942.
- Brutus, S., Fleenor, J. W., & London, M. 1998. Does 360-degree feedback work in different industries?: A between-industry comparison of the reliability and validity of multi-source performance ratings. *Journal of Management Development*. 17: 177-190.
- Carver, C. & Scheier, M. 1981. *Attention and self-regulation: A control theory approach to human behavior*. New York: Springer.
- Church, A. H. 1997. Managerial self-awareness in high performing individuals in organizations. *Journal of Applied Psychology*, 82: 281-292.
- Craig, S. B., & Hannum, K. 2006. Research update: 360-degree performance assessment. *Consulting Psychology Journal: Practice and Research*, 58: 117-124.
- DeVader, C. L., Bateson, A. G., & Lord, R. G. 1986. Attribution theory: A meta-analysis of attributional hypotheses. In E. A. Locke (Ed). *Generalizing from laboratory to field studies*. Lexington, MA: Lexington Books.
- Eichinger, R. W., & Lombardo, M. M. 2004. Patterns of rater accuracy in 360 degree feedback. *Human Resource Planning*, 27: 23-25.
- Eichinger, R. W., Lombardo, M. M., & Stiber, A. 2005. *Broadband talent managements: Paths to improvement*. Minneapolis, MN: Lominger Limited, Inc.
- Farh, J., & Dobbins, G. H. 1989. Effects of self-esteem on leniency bias in self-reports of performance: A structural equation model analysis. *Personnel Psychology*, 42: 835-850.
- Fletcher, C. & Baldry, C. 2000. A study of individual differences and self-awareness in the context of multi-source feedback. *Journal of Occupational and Organizational Psychology*, 73: 303-319.

- Harris, M. M., & Schaubroeck, J. 1988. A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, 41: 43-62.
- Holzbach, R. L. 1978. Rater bias in performance ratings: Superior, self, and peer ratings. *Journal of Applied Psychology*, 63: 579-588.
- Sosik, J. J. 2001. Self-other agreement on charismatic leadership: Relationships with work attitudes and managerial performance. *Group & Organization Management*, 26: 484-511.
- Johnson, J. W., & Ferstl, K. L. 1999. The effects of interrater and self-other agreement on performance improvement following upward feedback. *Personnel Psychology*, 52: 271-303.
- John, S. J., & Veronica, G. M. 2004. Self-other rating agreement in mentoring: Meeting protégé expectations for development and career advancement. *Group & Organization Management*, 29: 442-469.
- Jones, E. E., & Nisbett, R. E. 1972. The actor and the observer: Divergent perceptions of the causes of behavior. In Jones E.E., Kanouse D. A., Kelly H. H., Nisbett R.E., Valines S., & Weiner B. (Eds.), *Attribution: Perceiving the causes of behavior*: 79-94. Morristown, NJ: General Learning Press.
- Korman, A. K. 1970. Toward a hypothesis of work behavior. *Journal of Applied Psychology*, 54: 31-41.
- Korman, A. K. 1976. Hypothesis of work behavior revisited and an extension. *Academy of Management Review*, 1: 50-63.
- Laing, R. D., Phillipson, H., & Lee, A. R. 1966. *Interpersonal perception: A theory and a method of research*. New York: Spring Publishing Co.

- Lawler, E. E. 1967. The multitrate-multirater approach to measuring managerial job performance. *Journal of Applied Psychology*, 51: 369-380.
- Lombardo, M. M., & Eichinger, R. W. 2001. *The leadership machine*. Minneapolis MN: Lominger Limited, Inc.
- Lombardo, M. M., & Eichinger, R. W. 2003. *Leadership architect norms and validity report*. Minneapolis. MN: Lominger Limited, Inc.
- London, M., & Smither, J. W. 1995. Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory based applications and directions for research. *Personnel Psychology*, 48: 803-839.
- 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.
- McGraw, K. O., & Wong, S. P. 1996. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1: 30-46.
- Morgeson, F. P., Mumford, T. V., & Campion, M. A. 2005. Coming full circle: Using research to address 27 questions about 360-degree feedback programs. *Consulting Psychology Journal: Practice and Research*, 57: 196-209.
- Murphy, K. R., Martin, C., & Garcia, M. 1982. Do behavioral observation scales measure observation? *Journal of Applied Psychology*, 67: 562-567.
- Nowack, K. M. 1997. Self-ratings as a predictor of assessment center performance. *Journal of Social Behavior and Personality*, 12: 145-166.
- O'Reilly, C., & Anderson, J. 1980. Trust and the communication of performance appraisal information: The effects of feedback on performance and job satisfaction. *Human Communication Research*, 6: 290-298.

- Ostroff, C., Atwater, L. E., & Feinberg, B. J. 2004. Understanding self-other agreement: A look at rater and rate characteristics, context, and outcomes. *Personnel Psychology*, 57: 333-375.
- Salancik, G. R., & Pfeffer, J. 1978. A social information processing approach to job attitudes and task design. *Administrative Science Quarterly*, 23: 224-253.
- Smither, J. W., London, M., & Reilly, R. R. 2005. Does performance improve following multisource feedback? A theoretical model, meta-analysis, and a review of empirical findings. *Personnel Psychology*, 58: 33-66.
- Steel, R. P., & Ovalle, N. K. 1984. Self-appraisal based upon supervisory feedback. *Personnel Psychology*, 37: 667-685.
- Tornow, W. W. 1993. Perceptions or reality: Is multi-perspective measurement a means or an end? *Human Resource Management*, 32: 221-230.
- Van Velsor, E., & Leslie, J. B. 1991. *Feedback to managers: Vol. 1 A guide to evaluating multi-rater feedback instruments*. Greensboro, NC: Center for Creative Leadership.
- Wegner, D., & Vallacher, R. 1980. *The self in social psychology*. New York: Oxford University Press.
- Wicklund, R. 1979. The influence of self on human behavior. *American Scientist*, 67: 187-193.
- Wohlers, R. J. Hall, M., & London, M. 1993. Subordinates rating managers: Organizational and demographic correlates of self/subordinate agreement. *Journal of Occupational and Organizational Psychology*, 66: 263-275.
- Wohlers, R. J., & London, M. 1989. Ratings of managerial characteristics: Evaluation difficulty, co-worker agreement, and self-awareness. *Personnel Psychology*, 42: 235-261.

Yammarino, F., & Atwater, L. 1997. Do managers see themselves as others see them?

Organizational Dynamics, 25: 35-44.

Table 1. The mean ratings and standard deviations of the 67 competencies by rater sources

Competency	Boss		Direct Reports		Self	
	Mean	Std.	Mean	Std.	Mean	Std.
1	3.95	0.74	4.01	0.62	4.00	0.78
2	3.32	0.74	3.67	0.57	3.55	0.80
3	3.68	0.82	3.73	0.72	3.66	0.89
4	3.78	0.69	3.78	0.60	3.74	0.70
5	3.61	0.76	3.91	0.57	3.60	0.82
6	3.39	0.71	3.89	0.61	3.23	0.87
7	3.63	0.71	3.53	0.75	3.63	0.82
8	3.54	0.73	4.00	0.58	3.47	0.81
9	3.47	0.77	3.77	0.66	3.63	0.79
10	3.62	0.74	3.54	0.72	3.58	0.83
11	3.46	0.83	3.62	0.73	3.42	0.90
12	3.26	0.74	3.48	0.62	3.36	0.77
13	3.35	0.75	3.43	0.67	3.39	0.78
14	3.39	0.78	3.51	0.62	3.44	0.85
15	3.89	0.72	3.80	0.60	3.80	0.77
16	3.49	0.73	3.67	0.62	3.58	0.75
17	3.66	0.66	3.74	0.58	3.74	0.65
18	3.35	0.72	3.70	0.64	3.40	0.82
19	3.31	0.73	3.33	0.70	3.44	0.77
20	3.49	0.70	3.36	0.68	3.49	0.75
21	3.64	0.69	3.85	0.58	3.85	0.75
22	4.13	0.74	3.88	0.60	4.09	0.76
23	3.66	0.73	3.68	0.70	3.85	0.73
24	3.96	0.73	3.91	0.63	3.81	0.79
25	3.36	0.70	3.63	0.62	3.64	0.74
26	3.52	0.68	3.71	0.64	3.57	0.84
27	3.48	0.72	3.39	0.68	3.43	0.75
28	3.33	0.74	3.51	0.62	3.32	0.80
29	4.18	0.79	3.93	0.70	4.18	0.73
30	3.89	0.70	4.04	0.58	3.72	0.71
31	3.45	0.84	3.63	0.71	3.62	0.84
32	3.61	0.72	3.76	0.57	3.72	0.75
33	3.44	0.73	3.49	0.67	3.31	0.84
34	3.56	0.78	3.67	0.67	3.60	0.82
35	3.49	0.74	3.41	0.61	3.28	0.76
36	3.45	0.72	3.40	0.75	3.57	0.75
37	3.51	0.73	3.70	0.63	3.47	0.82
38	3.49	0.70	3.90	0.56	3.60	0.76
39	3.64	0.70	3.64	0.60	3.55	0.79
40	3.22	0.68	3.50	0.57	3.40	0.72
41	3.34	0.79	3.38	0.68	3.10	0.90
42	3.52	0.82	3.64	0.62	3.54	0.72

Competency	Boss		Direct Reports		Self	
	Mean	Std.	Mean	Std.	Mean	Std.
43	3.89	0.78	3.91	0.61	3.82	0.80
44	3.32	0.76	3.36	0.66	3.39	0.81
45	3.34	0.75	3.35	0.58	3.33	0.70
46	3.40	0.73	3.72	0.55	3.53	0.76
47	3.58	0.77	3.43	0.64	3.28	0.79
48	3.23	0.76	3.75	0.61	3.21	0.80
49	3.35	0.76	3.77	0.62	3.48	0.86
50	3.52	0.73	3.65	0.65	3.51	0.74
51	3.71	0.73	3.80	0.58	3.74	0.76
52	3.60	0.78	3.59	0.64	3.47	0.75
53	3.85	0.83	3.93	0.63	3.97	0.78
54	3.43	0.73	3.60	0.58	3.34	0.75
55	3.42	0.70	3.49	0.63	3.50	0.72
56	3.44	0.66	3.61	0.55	3.67	0.72
57	3.81	0.73	3.86	0.65	3.86	0.77
58	3.34	0.78	3.66	0.62	3.55	0.82
59	3.33	0.72	3.55	0.61	3.33	0.78
60	3.44	0.74	3.50	0.74	3.66	0.77
61	3.71	0.76	3.72	0.64	3.55	0.84
62	3.44	0.77	3.58	0.68	3.20	0.82
63	3.41	0.75	3.51	0.62	3.26	0.77
64	3.35	0.63	3.42	0.57	3.45	0.69
65	3.29	0.76	3.42	0.65	3.42	0.82
66	3.30	0.75	3.44	0.80	2.90	0.97
67	3.36	0.69	3.69	0.56	3.58	0.75
Mean	3.53	0.74	3.65	0.64	3.54	0.78

Table 2. Descriptive statistics and zero-order correlations

	Mean	Std.	Boss Perspective	DRs Perspective	Rating Ambiguity	Sureness	Self- Boss ICC
Boss Perspective	1.96	.71					
DRs Perspective	2.07	.73	-.17				
Rating Ambiguity	2.03	.60	-.43**	-.37**			
Sureness	3.91	.11	.19	.32**	-.55**		
Self-Boss ICC	.20	.06	.43**	.04	-.41**	.19	
Self-DRs ICC	.20	.06	.03	.45**	-.36**	.37**	.55**

** $p < .01$

Table 3. Hierarchical regression testing the influences of boss perspective and rating ambiguity

Step	IVs	ΔR^2	Std. β	Sig.
1	Boss Perspective		.31	$p < .05$
	Rating Ambiguity	.25	-.28	$p < .05$
2	Boss Perspective \times Rating Ambiguity	.00	-.04	$p = .92$

Dependent variable: Self-Boss ICC

Table 4. Hierarchical regression testing the influence of DRs perspective and rating ambiguity

Step	IVs	ΔR^2	Std. β	Sig.
1	DRs Perspective		.36	$p < .01$
	Rating Ambiguity	.24	-.23	$p < .05$
2	DRs Perspective \times Rating Ambiguity	.00	-.02	$p = .95$

Dependent variable: Self-DRs ICC

Table 5. Average self and boss ratings in different agreement groups

Self-boss agreement	Self rating	Boss rating
Under-Estimator	3.29	3.80
In-Agreement	3.52	3.50
Over-Estimator	3.83	3.27

Three mean ratings across categories differ significantly based on ANOVA, $p < .01$; all three pairs of means are significantly different using Tukey's HSD test, $p < .05$.

Table 6. Average self and direct reports ratings in different agreement groups

Self-DRs agreement	Self rating	DRs rating
Under-Estimator	3.28	3.90
In-Agreement	3.53	3.66
Over-Estimator	3.83	3.37

Three mean ratings across categories differ significantly based on ANOVA, $p < .01$; all three pairs of means are significantly different using Tukey's HSD test, $p < .05$.