Journal of Applied Psychology
2011, Vol. 96, No. 5, 1033–1044

0021-9010/11/$12.00 DOI: 10.1037/a0023368

Rating Leniency and Halo in Multisource Feedback Ratings: Testing Cultural Assumptions of Power Distance and Individualism-Collectivism

Kok-Yee Ng, Christine Koh, Soon Ang, Jeffrey C. Kennedy, and Kim-Yin Chan
Nanyang Technological University

This study extends multisource feedback research by assessing the effects of rater source and raters' cultural value orientations on rating bias (leniency and halo). Using a motivational perspective of performance appraisal, the authors posit that subordinate raters followed by peers will exhibit more rating bias than superiors. More important, given that multisource feedback systems were premised on low power distance and individualistic cultural assumptions, the authors expect raters' power distance and individualism-collectivism orientations to moderate the effects of rater source on rating bias. Hierarchical linear modeling on data collected from 1,447 superiors, peers, and subordinates who provided developmental feedback to 172 military officers show that (a) subordinates exhibit the most rating leniency, followed by peers and superiors; (b) subordinates demonstrate more halo than superiors and peers, whereas superiors and peers do not differ; (c) the effects of power distance on leniency and halo are strongest for subordinates than for peers and superiors; (d) the effects of collectivism on leniency were stronger for subordinates and peers than for superiors; effects on halo were stronger for subordinates than superiors, but these effects did not differ for subordinates and peers. The present findings highlight the role of raters' cultural values in multisource feedback ratings.

Keywords: multisource feedback, culture, power distance, individualism-collectivism, rating bias

Organizations routinely use subjective performance ratings for administrative as well as developmental purposes. However, these ratings are often challenged for their validity because "bias pervades the typical rating" (Wherry & Bartlett, 1982, p. 550). Instead of measuring ratees' performance, "ratings were stronger reflections of raters' overall biases" (Lance, 1994, p. 768). Two biases that are ubiquitous in performance ratings are (a) leniency, defined as the tendency for raters to assign higher ratings and (b) halo—raters' failure to differentiate among different dimensions of the ratee's behaviors (Murphy & Balzer, 1989; Saal, Downey, & Lahey, 1980). As a result of these biases, Murphy and Cleveland (1995) noted that "it is not unusual to find that 80% to 90% of all employees are rated as 'above average'" (p. 275).

Rating biases are likely to be even more prevalent in multisource feedback (MSF) systems, which have become very popular in organizations in and outside the United States. This is because unlike conventional feedback systems that rely on superiors' ratings alone, MSF involves ratings from peers and subordinates. Ratings from these nontraditional sources of feedback are relatively less understood (Mount & Scullen, 2001), and MSF scholars have warned that these raters may not be willing to provide honest feedback because of the risk of potential repercussions from ratees (Kudisch, Fortunato, & Smith, 2006; Murphy & Cleveland, 1995; Westerman & Rosse, 1997). This suggests that peers and subordinates could be more likely to inflate their ratings (leniency) across multiple behavioral dimensions (halo) in order to avoid consequences of negative feedback.

Furthermore, rating biases could be exacerbated when raters possess cultural beliefs that are inconsistent with the practice of giving upward or lateral feedback (Leslie, Gryskiewicz, & Dalton, 1998). This concern stems from the widespread recognition that MSF, as a practice that originated in the United States, is premised on underlying assumptions of individualistic and low-power distance values (Fletcher & Perry, 2001; Leslie et al., 1998; Shipper, Hoffman, & Rotondo, 2007; Stone-Romero & Stone, 2002; Varela & Premeaux, 2008). For instance, seeking (or providing) "objective" feedback on an individual's behaviors is based on individualistic values that emphasize personal striving and self-assertiveness (De Luque & Sommer, 2000; Morrison, Chen, & Salgado, 2004; Stone-Romero & Stone, 2002). Involving peers and subordinates also signals a redistribution of evaluative powers away from the superior and, hence, is more compatible with low-power distance values that are less sensitive to status and hierarchy (Leslie et al., 1998; Shipper et al., 2007). These arguments suggest that raters, particularly peers and subordinates, may be even more prone to rating biases when their power distance and individualism-collectivism value orientations are inconsistent with the premise of MSF.

Surprisingly, even though concerns relating to nontraditional sources of feedback and culture have been raised in the MSF literature, no empirical study has examined both of these issues simultaneously. As rating bias can undermine the quality of MSF ratings (Antonioni & Woehr, 2001), understanding how different

This article was published Online First April 11, 2011.

Kok-Yee Ng, Soon Ang, Jeffrey C. Kennedy, and Kim-Yin Chan, Division of Strategy, Management and Organization, Nanyang Technological University, Singapore; Christine Koh, Division of Information Technology and Operations Management, Nanyang Technological University.

Correspondence concerning this article should be addressed to Kok-Yee Ng, Center for Innovation Research in Cultural Intellegence, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798. E-mail: akyng@ntu.edu.sg

raters (especially peers and subordinates) and their cultural value orientations affect rating behaviors will offer important insights to organizations adopting MSF in the global workplace. In addressing this research question, we aim to close two major gaps in the existing research.

First, research has shown that rater effects account for substantially more variance in MSF ratings than ratee effects, particularly for peer and subordinate ratings (Greguras & Robie, 1998; Scullen, Mount, & Goff, 2000). According to Scullen et al. (2000), rater effects were 2.08 times larger than ratee effects for peer ratings, 1.86 times larger for subordinate ratings, and only 1.21 times larger for boss ratings. These findings highlight the importance of rater idiosyncratic effects but do not specify what rater attributes are likely to affect MSF ratings. We address this gap by focusing on raters' power distance and individualism-collectivism value orientations because they are relevant to the cultural assumptions of MSF, and to respond to the increasing use of MSF in today's diverse and global workplace (Atwater, Wang, Smither, & Fleenor, 2009).

Although power distance and individualism-collectivism originated as constructs that distinguish people across cultures (e.g., Hofstede, 2001), more recent research has argued that there is substantial within-culture variation in value orientations arising from regional, ethnic, religious, and generational differences (Hofstede, 2001; Kirkman, Lowe, & Gibson, 2006). This has prompted a growing number of researchers to examine power distance and individualism-collectivism as individual-differences constructs that vary within a single culture (e.g., Chan & Drasgow, 2001; Farh, Hackett, & Liang, 2007; Ilies, Wagner, & Morgeson, 2007; Kirkman, Chen, Farh, & Chen, 2009; Ng & Van Dyne, 2001). Consistent with these studies, we examine power distance and individualism-collectivism as individuals' beliefs about desirable end states that will affect their rating behaviors in the context of MSF.

Second, existing research comparing rating biases across different rater sources have produced inconsistent findings. For instance, several studies found peer ratings to be more lenient than supervisor ratings (Schneier, 1977; Springer, 1953; Zedeck, Imparto, Krausz, & Oleno, 1974), but others did not find significant differences across the different groups of raters (e.g., Harris & Schaubroeck, 1988; Holzbach, 1978; Klimoski & London, 1974; Tsui, 1984; Wohlers, Hall, & London, 1993). Findings on halo are likewise mixed. For example, whereas Lawler (1967) found peer ratings to exhibit comparable halo effects as superior ratings, Klimoski and London (1974) found that halo effect was stronger for peer than for superior ratings. In all these studies, comparisons of ratings across rater source were based on aggregate ratings pooled within each rater source (i.e., comparing aggregated superior ratings against aggregated peer and aggregated subordinate ratings). This aggregate approach offers less precise findings for understanding raters' rating bias because it does not take into account (a) individual raters' idiosyncracies, which have been found to account for a major portion of MSF ratings (Mount, Judge, Scullen, Sytsma, & Hezlett, 1998; Scullen et al., 2000) and b) the nonindependence of MSF ratings, because multiple observations are provided to a common ratee (i.e., ratee effects).

As advocated by Murphy and Cleveland (1995), we examine rater bias at the individual rater level, rather than at the aggregate level. We use a nested rating design in which raters were nested within ratee (i.e., each ratee received ratings from at least one superior, three peers, and three subordinates). This allows us to more rigorously assess how, for the same target of observation (i.e., ratee), individual raters' biases differ as a function of their hierarchical level and cultural value orientations. We adopt hierarchical linear modeling to focus on individual raters' biases (Level 1) while controlling for common ratee effects (Level 2). By addressing the methodological weaknesses in prior studies, our study aims to offer more conclusive insights on the effects of rater source on rating biases.

Below, we develop our research hypotheses and report our study comprising 1,447 superiors, peers, and subordinates who provided ratings on 172 ratees in an MSF program conducted for leadership development purposes.

Rater Source on Rating Bias

Research and anecdotal evidence have shown that a major cause of rating biases is raters' motivation (Harris, 1994; Murphy & Cleveland, 1995; Spence & Keeping, 2010). This perspective recognizes that raters are not "passive measurement instruments" whose sole interest is to provide accurate measures of ratees' performance (Murphy & Cleveland, 1995, p. 215). Rather, rating behaviors are often consciously or unconsciously guided by raters' goals. Two rater goals that are widely documented in the literature are (a) helping ratees improve their performance (Murphy & Cleveland, 1995; Wong & Kwong, 2007) and (b) averting negative consequences for oneself, and/or for the relationship with the ratee (Harris, 1994; Murphy & Cleveland, 1995; Spence & Keeping, 2010). These two goals work in conflict to affect rating behaviors. Wong and Kwong (2007) showed that raters who are motivated to help ratees improve their performance by identifying their strengths and weaknesses are more likely to provide ratings with less leniency and less halo. However, raters who are motivated to avoid negative outcomes are more likely to give ratings with greater leniency and halo.

In the context of a developmental MSF, the anonymity of response and the absence of administrative consequences for the ratees should ideally encourage raters to help their ratees improve their performance. In reality, however, raters are still likely to be concerned with potential negative consequences of providing accurate feedback (Spence & Keeping, 2010). This is because unfavorable feedback, even when provided for developmental purposes, has been found to arouse anger and discouragement in ratees (Brett & Atwater, 2001), which could adversely affect raters. Mount and Scullen (2001) also observed that subordinates, in particular, are likely to "feel intimidated or uncomfortable about rating the boss, even though assurances have been made that the ratings are confidential and anonymous" (p. 157). Hence, raters asked to provide developmental MSF generally have conflicting goals.

We contend that the degree of conflict between helping ratees improve and avoiding negative outcomes differs across raters with different hierarchical relationships with the ratee. A key organizational feature that distinguishes superior, peer, and subordinate raters is their level of *position power*, defined as the amount of influence one has over the other arising from one's formal position in the organization (Bass, 1960). Position power can be further described as comprising reward, coercive, and legitimate power

(French & Raven, 1959; Yukl & Falbe, 1991), all of which can affect raters' motivation and rating behaviors. Specifically, raters who possess greater reward and coercive power should have less concern for providing candid feedback because they control more organizational resources than the ratees (e.g., Bettenhausen & Fedor, 1997; Kudisch et al., 2006; Murphy & Cleveland, 1995). This should help mitigate raters' leniency arising from fear of negative consequences (Spence & Keeping, 2010). Raters with greater legitimate power to develop their ratees should be more motivated to help their ratees improve, because doing so is aligned with the expectations of members of their role set (Pfeffer & Salancik, 1975). This should encourage raters to provide feedback that will help ratees discern their strengths and weaknesses, thus resulting in ratings with less halo.

Taken together, we propose that superiors, compared with peers and subordinates, are least concerned with negative outcomes and most motivated to help their ratees improve. This is because superiors, by virtue of their reward and coercive power, should be less concerned with negative consequences compared with peer and subordinate raters (Kudisch et al., 2006). Moreover, compared with peers and subordinates, superiors have greater legitimate power to help their employees improve their performance through useful feedback (Kudisch et al., 2006; London, 2003; Westerman & Rosse, 1997). Therefore, we expect superiors to exhibit the least leniency and halo in their ratings.

In contrast, subordinates have the weakest reward and coercive power because they are dependent on their superior ratees for important resources at work that include tangible resources such as rewards and job opportunities, as well as intangible resources such as information and advice (Yukl & Falbe, 1991). As a result, subordinates are hesitant to provide feedback that may invoke negative emotions from their ratees (cf. Brett & Atwater, 2001), which could lead to negative consequences such as administrative reprisals or the withholding of valued resources (Bettenhausen & Fedor, 1997; Kudisch et al., 2006; Murphy & Cleveland, 1995). In addition, subordinates are likely to feel the least legitimate to provide feedback to their superiors for improvement, because upward ratings seriously violate the status hierarchy inherent in organizational structures (Westerman & Rosse, 1997). Hence, we expect subordinates' ratings to display greatest leniency and halo.

Peers should be more concerned than superiors with potential negative consequences of unfavorable feedback because of their mutual dependence on each other for resources such as information and cooperation (Fedor, Bettenhausen, & Davis, 1999). Compared with superiors, peers are also less likely to feel responsible for their peers' development. Fedor et al. (1999) noted that peer appraisals can be perceived as additional responsibilities that go beyond one's psychological contract of their job and employment relationships. However, compared with subordinates, peers possess greater reward and coercive power to the extent that peers can choose to provide or withhold intangible resources such as information, cooperation, and harmony due to their equal position power in the organization. Peers are also likely to feel more legitimate than subordinates to help their ratees identify their strengths and weaknesses because of their level of proximity and interdependence with the ratees (Fedor et al., 1999). Hence, we expect peers to exhibit more bias than superiors, but less bias than subordinates, in their MSF ratings.

Hypothesis 1 (H1): Compared with superiors, subordinates are more likely to show rating bias in terms of (a) leniency and (b) halo.

Hypothesis 2 (H2): Compared with superiors, peers are more likely to show rating bias in terms of (a) leniency and (b) halo.

Hypothesis 3 (H3): Compared with peers, subordinates are more likely to show rating bias in terms of (a) leniency and (b) halo.

Next, we argue that raters' cultural value orientations will further accentuate the effects of rater source on rating leniency and halo.

Moderating Effects of Cultural Value Orientations

Value orientations refer to an individual's beliefs about desirable end states that guide selection or evaluation of behavior and events (Schwartz & Bilsky, 1987). By specifying what is right and wrong, cultural value orientations guide individuals to select certain behaviors over others according to their internalized criteria and goals (Erez & Earley, 1993). Below, we consider how individual differences in power distance and individualism-collectivism, two cultural assumptions underlying MSF, moderate effects of rater source on rating leniency and halo.

Power Distance Value Orientation

Power distance describes the extent to which individuals accept social stratification and unequal distribution of power in the society. Individuals with high power distance believe that they should respect and obey those with authority and power, and adhere more to organizational hierarchy than individuals with low power distance (Hofstede, 2001). Consequently, individuals with high power distance are more conscious of status differences and how their behaviors should properly reflect these differences during interactions. However, those with low power distance are less attuned to distinctions arising from status positions and value equal participation in decisions (Atwater et al., 2009).

Power distance has direct implications for MSF. Several scholars have suggested that because of the acceptance of unequal power distribution by employees in high-power distance cultures, soliciting feedback from individuals with lower position and less power is seen as a severe violation of the status hierarchy (Fletcher & Perry, 2001; Shipper et al., 2007). Shipper et al. (2007), for instance, in comparing participants' reactions to their MSF ratings, found that those from countries with high power distance, such as Malaysia, reported lower commitment and higher tension compared with participants from low-power distance countries such as Ireland.

Interestingly, no research has focused on the effects of power distance from the rater's perspective. Building on our earlier arguments that superiors, peers, and subordinates will display differing degrees of leniency and halo because of different motivation, we argue that raters' power distance value orientation will further moderate the relationship between rater source and rating bias. Specifically, we expect that the influence of power distance values on rating leniency and halo will be stronger for subordinates than for superiors. Subordinates with high power distance, being

more sensitized to status markers than their low-power distance counterparts (Earley, 1999; Ng & Van Dyne, 2001), are likely to perceive a greater psychological distance between themselves and their superiors in terms of reward, coercive, and legitimate power (Antonakis & Atwater, 2002). Furthermore, high-power distance subordinates are concerned with managing "face" issues to maintain and cultivate their social networks with people from higher social positions (Kim & Nam, 1998). To these individuals, providing negative feedback to one's superior causes the superior to lose "face" and threatens the social order (Bond, Wan, Leung, & Giacalone, 1985; Kim & Nam, 1998). Taken together, we argue that high power distance will accentuate subordinates' motivation to avoid negative consequence, and attenuate subordinates' motivation to help ratees improve, so that high-power distance subordinates will display greater leniency and halo than their low-power distance counterparts.

In comparison, superiors' motivation to provide feedback should be less affected by raters' power distance values, because providing feedback to help subordinates improve is a typical and legitimate role for superiors. This argument is supported by Scullen et al.'s (2000) finding that across all rater sources, superiors' ratings demonstrated the least idiosyncratic effects.

Likewise, compared with subordinates, we propose that peers' power distance will have less effect on their motivation to give feedback, and hence on their rating bias. This is because there is no substantial power difference between peers and their ratees to trigger the effect of power distance orientation. Given that concern for "face protection" is less salient for individuals with equal or more power than the target (Kim & Nam, 1998), we do not expect the effect of power distance on rating leniency and halo to differ for peers and superiors.

Hypothesis 4 (H4): Power distance moderates the relationship between rater source and rating bias such that power distance will exert a greater positive impact on subordinates' rating bias in terms of (a) leniency and (b) halo, compared with superiors.

Hypothesis 5 (H5): Power distance moderates the relationship between rater source and rating bias such that power distance will exert a greater positive impact on subordinates' rating bias in terms of (a) leniency and (b) halo, compared with peers.

Individualism-Collectivism Value Orientation

Individualism-collectivism describes how an individual sees him- or herself in relation with the collective (Hofstede, 2001). Individualists view the self as independent of others, focus on personal goals, act on personal beliefs and values, and emphasize task outcomes, whereas collectivists construe the self as an interdependent entity, adopt group goals, act according to social norms, and stress good interpersonal relationships (Markus & Kitayama, 1991; Triandis, 1995). Related to the emphasis on harmony and relationship, collectivists are concerned with maintaining the face of others in their group, and hence strive to avoid or prevent the embarrassment of others. However, individualists are more concerned with preserving their own face. Taken together, these attributes of individualism-collectivism suggest that collectivists, being more concerned with how their behaviors affect relationships and group harmony, may be more motivated to avoid the negative social consequences of giving negative feedback than to provide feedback to help ratees improve.

Of greater interest in our study is how ratings by superiors, peers, and subordinates are influenced by their individualism-collectivism value orientation. We argue that individualism-collectivism orientation will affect ratings of subordinates and peers more than that of superiors. As with our earlier arguments, superiors' formal power should provide a stronger motivation for giving accurate feedback to help ratees improve, which will minimize the effects of individual differences in individualism-collectivism on rating leniency and halo (cf. Scullen et al., 2000).

Compared with superiors, we expect subordinates' collectivistic value orientation to exert a stronger and positive effect on their rating leniency and halo. Because collectivistic subordinates tend to prefer personalized relationships with their superior more than individualists (Hogg et al., 2005), they are more motivated to avoid damaging their relationship with their superior than to provide accurate feedback to help their superior improve. Furthermore, given that collectivists are generally less driven than individualists to act in ways that are consistent with their private thoughts (Markus & Kitayama, 1991), we expect collectivistic subordinates to have less motivation to provide accurate feedback, compared with their individualistic counterparts.

Likewise, we expect peers' collectivistic value orientation to exert a stronger and positive effect on their ratings compared with superiors. Compared with superiors, peer raters are likely to experience a greater dilemma when providing accurate feedback (Antonioni & Park, 2001). Although accurate feedback will help peer members develop and perform better, which in turn will help the team, accurate and unfavorable feedback can damage working relationships and weaken the team's social climate (de Nisi & Mitchell, 1978; Drexler, Beehr, & Stetz, 2001). Therefore, we argue that collectivistic peers are more motivated to avoid negative social consequences of candid feedback, whereas individualistic peers are more motivated to help their team members improve in order to enhance self and team performance. Thus, we expect collectivistic peers to display more leniency and halo than individualistic peers (Atwater et al., 2009; Hofstede, 2001).

Hypothesis 6 (H6): Individualism-collectivism moderates the relationship between rater source and rating bias such that individualism-collectivism will exert a greater positive impact on subordinates' rating bias in terms of (a) leniency and (b) halo, compared with superiors.

Hypothesis 7 (H7): Individualism-collectivism moderates the relationship between rater source and rating bias such that individualism-collectivism will exert a greater positive impact on peers' rating bias in terms of (a) leniency and (b) halo, compared with superiors.

Method

Participants and procedure. The present hypotheses were tested using MSF ratings collected for developmental purposes in the Singapore Armed Forces. According to the GLOBE's study of 62 societies (House, Hanges, Javidan, Dorfman, & Gupta, 2004), Singapore ranks 17th on ingroup collectivism and 42nd on power

distance. Despite the relatively high collectivism score, significant within-country variation in individuals' value orientations can be expected. This is because Singapore is a multiracial, multicultural, multireligious, and multilingual society. Most people speak at least two languages—English and their native tongue (Mandarin, Malay, or Tamil)—and are generally bicultural, endorsing a mix of traditional Asian and Western values (Chen, Ng, & Rao, 2005). Moreover, Singapore has a long history of significant foreign investment by businesses from the United States and European Union countries, resulting in a cultural heritage that "reflects values of both the East and the West" (Li, Ngin, & Teo, 2007, p. 950).

Participants were military officers attending a 9-month leadership program. As part of the program, participants obtained MSF from their direct superior(s), peers, and subordinates in their units. In addition to providing feedback on ratees' leadership behaviors, observers also provided information on their own cultural value orientations (power distance and individualism-collectivism) and demographics. The final sample comprises 172 ratees who had fulfilled the organization's specified rating requirements (at least one superior, three peers, and three subordinates). In total, 1,447 observer ratings were obtained. On average, each ratee received 8.7 ratings (ranging from 7 to 14), with an average of 1.3 superior (ranging from 1 to 2), 3.6 peer (ranging from 3 to 6), and 4.1 subordinate (ranging 3 to 8) ratings. Ratees were predominantly male (94%), with an average age of 33.4 years ($SD = 1.5$) and have spent an average of 2.6 years ($SD = 2.0$) in the current unit within the organization. Raters were also mostly male (90%), with an average age of 33.8 years ($SD = 7.2$) and have spent an average of 3.7 years ($SD = 4.7$) in the unit within the organization.

Measures.

Rating leniency. In field research where true scores of performance are not available, rating leniency is operationalized by the level of ratings, with higher ratings demonstrating greater leniency (e.g., Bernardin, Cooke, & Villanova, 2000; Heidemeier & Moser, 2009; Mount, 1984; Tsui & Barry, 1986). Consistent with these studies, leniency is operationalized with the average rating that raters provide on ratees' leadership skills. These leadership skills were assessed using an instrument developed by the organization for their leadership development program. The instrument measures six critical skills covering conceptual, task, and relational aspects of leadership that are important to the organization. Raters were presented with a total of 119 behavioral statements and asked to rate how accurately each statement described the ratee (1 = *not at all*, 7 =*to a very great extent*). Cronbach's alpha for the six scales all exceeded 0.80 (ranging from 0.84 to 0.90). Consistent with existing research (e.g., Brett & Atwater, 2001), the six scales were aggregated to obtain an overall leadership rating (Cronbach's $\alpha = 0.97$).

Rating halo. Consistent with previous studies (e.g., Vance, Winne, & Wright, 1983; Wong & Kwong, 2007), rating halo is operationalized as the standard deviation of the six scales, with lower variability representing greater halo.

Rater source. Rater source was coded using two dummy variables, subordinates (1 = subordinate, 0 = not subordinate) and peers (1 = peer, 0 = not peer), with superiors being the omitted reference category. The superior is chosen as the reference source, as research shows that superior ratings tend to be the most valid (Scullen et al., 2000).

Power distance. Power distance was measured with three items adapted from Dorfman and Howell (1988) (1 = *strongly disagree*, 7 = *strongly agree*; Cronbach's $\alpha = 0.81$). A sample item was "I believe it is important to respect the decisions made by those who have more power." The higher the score, the greater the power distance.

Individualism-collectivism. Individualism-collectivism was measured with three items adapted from Earley (1993) (1 = *strongly disagree*, 7 = *strongly agree*; Cronbach's $\alpha = 0.94$). A sample item was "I prefer to work in a group rather than by myself." The higher the score, the greater the collectivism.

Controls. Research on relational demography suggests that ratees who share similar demographic attributes (e.g., age, gender, job and company tenure) as raters tend to be rated more positively (Tsui & O'Reilly, 1989). Thus, both raters' and ratees' age (measured in years), gender (coded as 1 = male, 2 = female), and tenure in the organizational unit (in years) were controlled for in the analyses. In addition, raters' liking for the ratee was controlled for, given that research has found raters' affect for ratees to be a significant predictor of rating bias (e.g., Antonioni & Park, 2001; Tsui & Barry, 1986). Liking was measured with three items adapted from Tsui and Barry (1986) (1 = *not at all well*, 5 = *to a very great extent*; Cronbach's $\alpha = 0.80$). A sample item was "How well do you like this person?"

Analyses. Before the hypotheses were tested, the discriminant validity of the measures was examined using confirmatory factor analysis (CFA) with LISREL 8 (Jreskog & Sörbom, 2006). Results of the proposed four-factor structure (leadership rating, power distance, individualism-collectivism, and liking) demonstrated good fit with the data, $\chi^2(84, N = 1{,}477) = 526.37$, $p = .00$; root-mean-square error of approximation (RMSEA = .060); standardized root-mean-square residual (SRMR = .030); nonnormed fit index (NNFI) = .98; comparative fit index (CFI) = .98. All items loaded significantly on their predicted constructs, with standardized factor loadings ranging from .68 to .98. Composite reliabilities all exceeded 0.70 (rating = .97; power distance = .81; collectivism = .95; and liking = .80).

Relative fit of the hypothesized four-factor model was further compared with several alternate models. Results showed that the four-factor model demonstrated better fit compared with (a) a three-factor model that combined leadership rating and liking, $\Delta\chi^2(87 - 84 = 3, N = 1{,}447) = 1325.99$, $p < .001$; (b) a two-factor model that combined leadership rating and liking as one factor, and power distance and collectivism as another factor, $\Delta\chi^2(89 - 84 = 5, N = 1{,}447) = 2640.07$, $p < .001$; (c) a two-factor model that combined power distance, collectivism, and liking as one factor, $\Delta\chi^2(89 - 84 = 5, N = 1{,}447) = 3225.98$, $p < .001$; and (d) a one-factor model in which all items loaded on a single factor, $\Delta\chi^2(90 - 84 = 6, N = 1{,}447) = 6746.25$, $p < .001$. Taken together, the fit indices of the nested models show that leadership rating, individualism-collectivism, power distance, and liking were distinct constructs.

The hypotheses were tested using hierarchical linear modeling (HLM) (Bryk & Raudenbush, 1992). The data are inherently nested in nature, with each ratee being rated by multiple raters. This gives rise to the possibility of nonindependence of data, which violates an underlying assumption of ordinary least squares estimation. HLM addresses this by maintaining the appropriate level of analysis (Hofmann, 1997) and allows the examination of

the effect of Level 1 (rater) predictors while controlling for Level 2 (ratee) effect. All Level 1 variables were group mean centered because the primary interest involves only Level 1 predictors, as this "removes all between-cluster variation from the predictor and yields a 'pure' estimate of the pooled within-cluster (i.e., level 1) regression coefficient" (Enders & Tofighi, 2007, p. 128). Sensitivity analysis with grand mean centering showed a similar pattern of results.

Following recommended practice (Kreft & de Leeuw, 1998), a null model with no predictors was first specified to test whether there is significant variation between ratees in rating leniency and halo, a necessary precondition for supporting the present hypotheses. Next, controls were added for both Level 1 (rater age, gender, organizational unit tenure, liking) and Level 2 (ratee age, gender, organizational unit tenure) equations in Step 1, rater source (peer and subordinate vs. superior) in Step 2, and cultural values and the interaction terms between cultural values and rater source in Step 3. Results are reported in the final step of the present analyses. The *t* tests associated with the estimated parameters provide a direct test of the present hypotheses. The effect size of each step was computed by comparing each step's new value of σ^2 (within-group variance) with the σ^2 of the null model:

$$(R^2_{\text{Level 1 model}}) = (\sigma^2_{\text{null model}} - \sigma^2_{\text{current model}})/\sigma^2_{\text{null model}}.$$

This ratio represents the percentage of the Level 1 (rater) variance in the dependent variable accounted for by the added predictors (Bliese, 2002; Bryk & Raudenbush, 1992; Hofmann, 1997).

Results

Table 1 presents descriptives, correlations, and internal consistencies of all variables. Table 2 presents the results of our HLM analyses.

Results of our null models for leniency and halo showed that there was significant between-ratee variance at $p < .01$ (ICC for leniency = .175; ICC for halo = .050). Hence, the prerequisite condition of systematic between-ratee variance in the dependent variables was satisfied.

Not surprisingly, results of Models 3 and 6 showed that raters who had greater liking for their ratees showed more leniency ($\gamma = 0.44$, $p < .01$) and halo ($\gamma = 0.03$, $p < .01$). All the other controls were not significantly related to either leniency or halo.

H1–H3 predicted that rater source will exert a main effect on rating leniency and halo. Specifically, H1 and H2, respectively, predicted that subordinates and peers will show more rating bias than superiors in terms of leniency (H1a, H2a) and halo (H1b, H2b). Results of Model 3 showed that, compared with superiors, subordinates ($\gamma = 0.26$, $p < .01$) and peers ($\gamma = 0.11$, $p < .01$) were more lenient, thus supporting H1a and H2a. Results of Model 6 showed that, compared with superiors, subordinates exhibited greater halo ($\gamma = 0.02$, $p < .01$). However, the effect for peers was not significant ($\gamma = 0.01$, *ns*). Hence, H1b was supported but not H2b.

H3 predicted that subordinates will show more leniency (H3a) and halo (H3b) than peers. Given that both peers and subordinates were included categories in our equation, we compared the two regression coefficients (γ_{sub} and γ_{peer}) using the contrast test in HLM (Hox, 2002; Raudenbush & Bryk, 2002). A contrast test is essentially a composite hypothesis that tests whether the two regression coefficients are equal (i.e., $\gamma_{\text{sub}} = \gamma_{\text{peer}}$) and is based on an asymptotic chi-square test. A benefit of using the contrast test is the "protection against heightened probability of Type I errors that arises from performing many univariate tests" (Raudenbush & Bryk, 2002, p. 60).

Results of the contrast test showed that the difference in the two coefficients for leniency was significant, $\chi^2(1, N = 1{,}447) = 58.25$, $p < .01$, indicating that subordinates gave more lenient ratings than peers. Likewise, comparison of peers' and subordinates' coefficients for halo was significant, $\chi^2(1, N = 1{,}447) = 9.61$, $p < .01$, indicating that subordinates showed greater rating halo compared with peers. Hence, H3a and H3b were supported.

Table 1
Means, Standard Deviations, Scale Reliabilities, and Intercorrelations Among Variables[a]

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Rating leniency	5.57	0.77	(.97)												
2. Rating halo	−0.32	0.19	.43**	—											
3. Collectivism (IC)	5.32	1.48	.10**	.04	(.94)										
4. Power distance (PD)	3.95	1.34	−.05*	−.03	−.28**	(.81)									
5. Peer[b]	0.41	0.49	−.14**	.07**	.02	−.01	—								
6. Subordinate[b]	0.45	0.50	.18**	.08**	−.09**	.05*	−.75**	—							
7. Rater age	33.78	7.21	.00	.06*	.04	−.04	−.03	.21**	—						
8. Rater gender[c]	1.10	0.30	−.06*	.03	−.05	.04	.02	.05	.09**	—					
9. Rater org. tenure	3.66	4.73	.05*	.07*	−.02	.00	−.12**	.15**	.39**	.13**	—				
10. Liking	3.68	0.65	.38**	.12**	.16**	−.07**	−.05*	−.12**	.18**	−.12**	.07**	(.80)			
11. Ratee age	33.44	1.46	−.05	−.05	−.05	.05	.01	.00	.10**	−.07*	.04	.07*	—		
12. Ratee gender[c]	1.06	0.23	−.06*	.01	−.03	−.03	.00	−.01	.02	.02	−.02	.00	−.04	—	
13. Ratee org. tenure	2.62	2.04	−.05	−.03	−.04	.01	.00	.01	.08**	.01	.28**	.08**	.16**	−.01	—

Note. Reliability coefficients are in parentheses along the diagonal. Correlations between ratee-level variables (11–13) and rater-level variables are calculated by assigning the same ratee score to all raters rating the ratee ($N = 1{,}447$). Intercorrelations between the ratee-level variables are calculated at the ratee level ($N = 172$). org. = organizational.

[a] $N = 1{,}447$. [b] Rater type was coded using two dummy variables—subordinate (1 = yes, 0 = no) and peer (1 = yes, 0 = no), with superior being the omitted reference category. [c] Gender was coded as 1 = male, 2 = female.

* $p < .05$. ** $p < .01$.

Table 2
HLM Results Predicting Rating Leniency and Halo

Variable	Leniency			Halo		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Level 1 (Rater)						
Intercept	5.58**	5.58**	5.58**	−0.32**	−0.32**	−0.32**
Age	−0.01**	0.00	0.00	0.00	0.00	0.00
Gender	0.02	−0.03	−0.03	0.02	0.02	0.02
Org. tenure	0.02**	0.00	0.00	0.00	0.00	0.00
Liking	0.40**	0.44**	0.44**	0.03**	0.04**	0.03**
Peer		0.11**	0.11**		0.01	0.01
Subordinate (Sub)		0.25**	0.26**		0.02**	0.02**
Power distance (PD)			−0.01			0.00
Collectivism (IC)			0.04*			0.01
PD × Peer			0.03			0.00
PD × Sub			0.07**			0.02**
IC × Peer			0.05*			0.02
IC × Sub			0.08**			0.03**
Level 2 (Ratee)						
Age	−0.02	−0.02	−0.02	0.00	0.00	0.00
Gender	−0.19	−0.19	−0.21	0.01	0.01	0.01
Org. tenure	−0.02	−0.02	−0.02	0.00	0.00	0.00
Level 1 variance explained ($R^2_{\text{Level 1}}$)[a]	0.13	0.19	0.27	0.04	0.06	0.23

Note. N (Level 1) = 1,447; N (Level 2) = 172. Entries corresponding to the predicting variables are estimations of the fixed effects (standard errors not shown). HLM = hierarchical linear modeling; Org. = Organizational.
[a] All Level 1 variables were group mean centered. All Level 2 variables were grand mean centered.
* $p < .05$. ** $p < .01$.

The next two hypotheses, respectively, proposed that subordinates' power distance will exert a stronger positive impact on their ratings compared with superiors (H4a: leniency; H4b: halo) and peers (H5a: leniency; H5b: halo). Results demonstrate that the interaction terms between subordinate and power distance in predicting leniency (Model 3: $\gamma = 0.07$, $p < .01$) and halo (Model 6: $\gamma = 0.02$, $p < .01$) were significant. Specifically, the effects of power distance on both leniency and halo were significantly stronger for subordinates than for superiors. Thus, H4a and H4b were supported.

To test whether subordinates' power distance exerted a stronger influence on ratings compared with peers (H5), we similarly defined a contrast between the two regression coefficients (i.e., $\gamma_{PD \times sub} = \gamma_{PD \times peer}$). Results of the contrast test for both leniency and halo showed that the Subordinate × Power Distance Effect was significantly stronger than the Peer × Power Distance Effect in predicting leniency, $\chi^2(1, N = 172) = 4.61$, $p < .05$, and halo, $\chi^2(1, N = 172) = 6.28$, $p < .05$. Thus, H5a and H5b were supported.

As we expected, the interaction term between peers and power distance was not significant for both leniency (Model 3: $\gamma = 0.03$, *ns*) and halo (Model 6: $\gamma = 0.00$, *ns*), suggesting that the effect of power distance on ratings did not differ between peers and superiors.

The last two hypotheses, respectively, proposed that subordinates' (H6) and peers' (H7) individualism-collectivism will exert a stronger positive impact on their leniency (H6a, H7a) and halo (H6b, H7b) compared with superiors. Results demonstrate that the interaction terms between individualism-collectivism and subordinates were significant for both leniency (Model 3: $\gamma = 0.08$, $p < .01$) and halo (Model 6: $\gamma = 0.03$, $p < .01$), thus suggesting that the effect of individualism-collectivism was stronger for subordinates' ratings than for superiors' ratings. Hence, H6a and H6b were supported. For peers, the interaction term between individualism-collectivism and peers was significant for leniency (Model 3: $\gamma = 0.05$, $p < .05$) but not for halo (Model 6: $\gamma = 0.02$, *ns*). Hence, H7a was supported but not H7b.

Although not hypothesized, we tested whether the effect of individualism-collectivism on rating bias was stronger for subordinates than for peers. Results of the contrast tests for both leniency and halo showed that the Subordinate × Individualism-Collectivism terms were not significantly different from the Peer × Individualism-Collectivism terms for both leniency, $\chi^2(1, N = 172) = 1.80$, $p = ns$, and halo, $\chi^2(1, N = 172) = 1.33$, $p = ns$, suggesting that the effect of individualism-collectivism on ratings did not differ between subordinates and peers.

The main and interaction effects of rater cultural value orientations explained an additional 8% of variance for leniency and 17% for halo. Overall, our hypothesized model accounted for 27% of the within-ratee (i.e., Level 1) variance in MSF ratings for leniency and 23% for halo.

Figure 1 illustrates the general form of interaction between rater source and rater value orientations on leniency and halo.

Discussion

Our study examines a long-standing concern that nontraditional sources of feedback may be more susceptible to rating bias such as leniency and halo (Kudisch et al., 2006; Murphy & Cleveland, 1995; Westerman & Rosse, 1997). More important, we provide timely insights to the cultural assumptions of MSF by examining the moderating impact of raters' power distance and

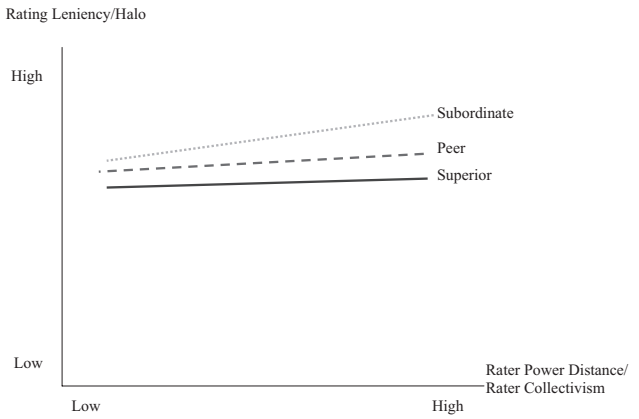

Figure 1. General form of interaction between rater source and rater's power distance/collectivism on rating leniency and halo.

individualism-collectivism orientations on rating leniency and halo. Results of our HLM analyses provide important verifications as well as novel insights that can contribute to the science and practice of MSF, in the following ways.

First, our study generally supports our first three hypotheses that rater source exerts a systematic effect on rating leniency and halo. As expected, our research shows that subordinates consistently display more leniency and halo in their ratings compared with peers and superiors. This finding corroborates with anecdotes of subordinates' hesitation to violate status hierarchy with upward feedback (Kudisch et al., 2006; Westerman & Rosse, 1997). Results for peers' ratings are more mixed and potentially interesting. Our study supports the ordered effects for peers' leniency (subordinates > peers > superiors) but not for peers' halo (subordinates > peers = superiors). This pattern of findings seems to suggest that of the three rater sources, peers, who are in between superior and subordinate raters, face the greatest conflict between avoiding negative consequences versus helping ratees improve their performance. This dilemma appears to result in a creative rating pattern that seeks to avoid negative social consequences (through giving lenient feedback) and, at the same time, helps ratees identify their strengths and weaknesses to improve (through giving more discriminant feedback, i.e., less halo).

Second, we found clear support for our proposed moderating effects of power distance on rater source and rating behaviors. Consistent with H4, power distance exerts the strongest impact on subordinates' rating leniency and halo. These findings suggest that subordinates with high power distance are most susceptible to rating bias, because they are more likely than others to perceive that upward feedback violates the subordinate-superior status hierarchy. Compared with subordinates, superiors' and peers' ratings were less affected by raters' power distance. This could be due to the fact that status difference—an important condition to activate power distance—is not as salient for peer and superior raters.

Third, and consistent with our final hypothesis, the effects of individualism-collectivism on rating leniency and halo were stronger for subordinates than for superiors. We expect individualism-collectivism to have the least impact on rating leniency and halo for superiors because of their formal power, and our results support this argument. Results for peers are mixed. Compared with superiors, peers' individualism-collectivism value exerts a stronger effect on rating leniency, but not on halo. This finding could imply that whereas collectivistic peers are more motivated to avoid damaging relationships with ratees than individualistic peers (and hence, provided more lenient ratings), both collectivistic and individualistic peers are equally motivated to provide feedback that helps ratees identify their strengths and weaknesses (and hence, did not differ on halo).

In summary, our results show that rater characteristics play an important role in affecting MSF ratings. In our study, 83% of the total variance in leniency, and 95% of the total variance in halo, reside at Level 1 (i.e., rater level), rather than at Level 2 (i.e., ratee level). This is consistent with Scullen et al.'s (2000) conclusion that rater effects account for more variance in MSF ratings than ratee effects. More important, our study shows that raters' hierarchical level, collectivism, and power distance values together explain 14% of Level 1 variance in leniency and 19% of Level 1 variance in halo. Specifically, our results largely support our prediction that peers' and subordinates' ratings are more likely to be influenced by their power distance and individualism-collectivism value orientations, compared with superiors. Subordinates, in particular, are the most susceptible to rating biases when they possess value orientations that are inconsistent with the premise of MSF. Peers' ratings are less different from superiors', especially when it concerns halo. Our findings also corroborate Scullen et al.'s conclusion that rater biases accounted for more variance in subordinate ratings than in either superior or peer ratings. Our study contributes to Scullen et al.'s study by showing that power distance and individualism-collectivism are specific idiosyncratic attributes that can explain meaningful variance in subordinates' upward ratings in MSF. They also provide important empirical evidence that sheds some light on the cultural assumptions, and hence, cultural boundaries of MSF.

Theoretical implications and future research directions. Findings of this study have several implications for future research on MSF ratings in a global work environment. First, our present findings are based on MSF ratings gathered for developmental purposes. Future research should examine how the purpose of MSF would affect the results reported in this study. For instance, existing studies have found that ratings collected for administrative purposes are more prone to rating bias than those collected for developmental purposes (Farh, Cannella, & Bedeian, 1991; Harris, Smith, & Champagne, 1995; Jawahar & Williams, 1997). This is because, compared with developmental ratings, administrative ratings carry more important stakes for the ratees (e.g., no pay increase) that are likely to accentuate potential repercussions on the raters (e.g., retaliatory behaviors or decreased motivation to perform). This argument implies that MSF conducted for administrative purposes could present a "strong" situation that diminishes interindividual differences in promoting rating bias, thus suggesting that the effects of rater source and rater cultural value orientations could be dampened.

Second, our present findings on interaction effects involving rater's cultural values highlight the possible cultural boundaries of the usefulness of MSF. Echoing the recent recommendation by Gelfand, Erez, and Aycan (2007) for future research on cross-cultural organizational behavior to move beyond individualism-collectivism, we urge future studies to examine other cultural values that are relevant to predicting rating leniency and halo. For

instance, future research could examine raters' social axioms, defined as individuals' general beliefs about themselves and their environment (Leung et al., 2002), to augment our existing understanding of value-based cultural dimensions. In particular, the social axiom of societal cynicism, which describes individuals' mistrust of people and social institutions, could affect raters' views toward MSF, and hence, influence their willingness to provide accurate ratings.

Third, future research could assess raters' motivation (e.g., Murphy, Cleveland, Skattebo, & Kinney, 2004; Wong & Kwong, 2007) as a set of mediating mechanisms that explain the effects of rater source and cultural values on rating behaviors. This will empirically ascertain, for instance, whether subordinates with high collectivistic and power distance orientations give more lenient and less differentiated ratings because they are more motivated to avert negative consequences and less motivated to provide feedback that will help ratees improve. Having a more in-depth understanding of the underlying mechanisms can in turn offer more precise insights on the interventions that are needed to encourage accurate feedback, thus ensuring that the purpose of MSF is realized.

Strengths and limitations. Our study has several methodological strengths. First, direct assessment of raters' cultural value orientations is rare in MSF studies because of concerns of rater survey fatigue and nonparticipation. This could be why even though scholars have highlighted the critical role of culture in MSF more than a decade ago (London & Smither, 1995), systematic empirical research on the topic is scarce. The few empirical studies on culture and MSF (Atwater et al., 2009; Entrekin & Chung, 2001; Heidemeier & Moser, 2009; Shipper et al., 2007) did not directly assess raters' or ratees' value orientations but instead examined group differences based on Hofstede's (2001) or GLOBE's (House et al., 2004) societal level scores for power distance and individualism-collectivism. Our study, by assessing raters' cultural values at the individual level of analysis, acknowledges important intracultural variation in value orientations (Hofstede, 2001; Kirkman et al., 2006), and also offers a unique opportunity to examine the psychological impact of culture on MSF participants.

Second, our nested design and our HLM analyses enable us to test our hypotheses at the level of the individual rater, and at the same time account for ratee effects. Our study also presents a novel application of HLM in the study of rater bias. As such, this article advances a methodological improvement over previous research that has examined mean rating level differences across the three rater sources using *t* tests or analysis of variance.

As with most MSF research, restriction in variance in performance is likely to have attenuated relationships in our model (Lebreton, Burgess, Kaiser, Atchley, & James, 2003). This is particularly true for our study conducted with a sample of military leaders undergoing a leadership development program. In spite of such restriction, and the conventional belief that ratings should be determined largely by ratee's performance, our ability to explain a significant amount of variance in MSF ratings using raters' attributes such as hierarchical level, collectivism, and power distance suggests that these effects are pervasive and nontrivial (Prentice & Miller, 1992). Furthermore, the practical significance of these effects is accentuated when MSF is used for administrative purposes with high-stakes consequences.

Our sampling from one organization in Singapore raises the issue of external generalizability. However, we note that an advantage of sampling from one organization is that systematic organizational factors that could affect rating behaviors, such as feedback environment, are kept constant. Sampling from one culture also mitigates the problem of measurement equivalence, as research has found that performance ratings across cultures are not completely invariant (e.g., Ployhart, Wiechmann, Schmitt, Sacco, & Rogg, 2003). Nonetheless, future research should replicate our design in different organizational settings and different cultures to ascertain the generalizability of results obtained in our present study. Future studies that plan to sample more widely from multiple organizations will need to take into account differences in the social context of the appraisal that could influence results (Levy & Williams, 2004).

Practical implications. Our study on rating leniency and halo in MSF ratings has important implications for organizations. MSF research has shown that managers who receive unfavorable feedback are more likely to improve than those who receive favorable feedback (Reilly, Smither, & Vasilopoulos, 1996; Walker & Smither, 1999). This suggests that when raters provide inflated feedback, they are less likely to "jolt" ratees to improve themselves. Moreover, when raters provide ratings that vary little across behavioral dimensions, they offer fewer insights to ratees on their strengths and weaknesses. Our study shows specifically that cultural profiles characterized by collectivistic and high-power distance values could pose potential challenges to the value of MSF. We suggest two complementary sets of recommendations to address some of these challenges.

Our first set of recommendations focuses on the raters. As the influence of culture on rating behaviors is often implicit, raters can first be made aware of how their cultural values, in conjunction with their organizational role vis-à-vis the ratee, may shape their rating tendencies. Such cultural self-awareness, an important aspect of cultural intelligence (Ang et al., 2007), can be coupled with rater training to develop a shared understanding of the nature of feedback that is useful for ratees to improve themselves. For instance, Murphy and Cleveland (1995) suggest that "frame-of-reference" training that is traditionally used to help raters differentiate performance levels in their ratees can be applied to help raters identify what rating behaviors are appropriate versus inappropriate when providing feedback. This helps to reduce the influence of raters' idiosyncratic effects on rating behaviors. In addition, raters providing consistently high ratings across items could be prompted by the computerized MSF system for narrative justification of these ratings. Holding raters accountable for their performance ratings in this way has been shown to increase accuracy (Mero & Motowidlo, 1995).

Our second set of recommendations focuses on the recipients of MSF. During debriefing sessions, MSF recipients should be encouraged to consider the possibility of leniency and halo when acting on the feedback they receive. This is especially important in situations in which cultural factors are likely to be salient (e.g., in high-power distance and collectivistic cultures). For instance, recipients should be advised to pay attention to the relatively lower scores, as well as small differences between scale scores. In addition, including instrument norms in MSF reports can help recipients compare their scores with organizational benchmarks to identify areas for improvement.

Conclusion. Involving peers and subordinates in the feedback process is a distinctive hallmark of MSF. Underlying the involvement of these nontraditional sources of feedback is the implicit assumption of low power distance and individualistic values. This prompted us to examine whether peers and subordinates with high power distance and collectivistic orientations are more susceptible to rating bias than superiors. Our results show that the joint effects of rater source and cultural value orientations in predicting rating leniency and halo are substantial, thus highlighting the importance of considering cultural assumptions when implementing MSF. Besides contributing new theoretical insights to the MSF literature, our study also highlights the methodological importance of using HLM to control for ratee effects in order to yield more robust results. Practically, our findings underscore the importance of ensuring that organizational members, particularly subordinates and peers with high power distance and collectivistic values, are ready for MSF.

References

Ang, S., Van Dyne, L., Koh, S. K., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural intelligence: An individual difference with effects on cultural judgment and decision making, cultural adaptation, and task performance. *Management and Organization Review, 3,* 335–371. doi:10.1111/j.1740-8784.2007.00082.x

Antonakis, J., & Atwater, L. (2002). Leader distance: A review and a proposed theory. *Leadership Quarterly, 13,* 673–704. doi:10.1016/S1048-9843(02)00155-8

Antonioni, D., & Park, H. (2001). The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management, 27,* 479–495. doi:10.1177/014920630102700405

Antonioni, D., & Woehr, D. J. (2001). Improving the quality of multisource rater performance. In D. Bracken, C. Timmreck, & A. Church (Eds.), *Handbook of multisource feedback* (pp. 114–129). San Francisco, CA: Jossey-Bass.

Atwater, L. E., Wang, M., Smither, J. W., & Fleenor, J. W. (2009). Are cultural characteristics associated with a relationship between self and others' ratings of leadership? *Journal of Applied Psychology, 94,* 876–886. doi:10.1037/a0014561

Bass, B. M. (1960). *Leadership, psychology, and organizational behavior.* New York, NY: Harper.

Bernardin, H. J., Cooke, D. K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology, 85,* 232–236. doi:10.1037/0021-9010.85.2.232

Bettenhausen, K. L., & Fedor, D. B. (1997). Peer and upward appraisals: A comparison of their benefits and problems. *Group and Organization Management, 22,* 236–263. doi:10.1177/1059601197222006

Bliese, P. D. (2002). Multilevel random coefficient modeling in organization research: Examples using SAS and S-PLUS. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 401–445). San Francisco, CA: Jossey-Bass.

Bond, M. H., Wan, K., Leung, K., & Giacalone, R. A. (1985). How are responses to verbal insult related to cultural collectivism and power distance? *Journal of Cross-Cultural Psychology, 16,* 111–127. doi:10.1177/0022002185016001009

Brett, J. F., & Atwater, L. E. (2001). 360º feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology, 86,* 930–942. doi:10.1037/0021-9010.86.5.930

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury, CA: Sage Publications.

Chan, K. Y., & Drasgow, F. (2001). Toward a theory of individual differences and leadership: Understanding the motivation to lead. *Journal of Applied Psychology, 86,* 481–498. doi:10.1037/0021-9010.86.3.481

Chen, H., Ng, S., & Rao, A. R. (2005). Cultural differences in consumer impatience. *Journal of Marketing Research, 42,* 291–301. doi:10.1509/jmkr.2005.42.3.291

De Luque, M. F. S., & Sommer, S. M. (2000). The impact of culture on feedback-seeking behavior: An integrated model and propositions. *Academy of Management Review, 25,* 829–849. doi:10.2307/259209

de Nisi, A. S., & Mitchell, J. L. (1978). An analysis of peer ratings as predictors and criterion measures and a proposed new application. *Academy of Management Review, 3,* 369–374. doi:10.2307/257678

Dorfman, P. W., & Howell, J. P. (1988). Dimensions of national culture and effective leadership patterns: Hofstede revisited. In J. L. C. Cheng & R. B. Peterson (Eds.), *Advances in international comparative management* (Vol. 3, pp. 127–150). Stamford, CT: JAI Press.

Drexler, J. A., Beehr, T. A., & Stetz, T. A. (2001). Peer appraisals: Differentiation of individual performance on group tasks. *Human Resource Management, 40,* 333–345. doi:10.1002/hrm.1023

Earley, P. C. (1993). East meets West meets Mideast: Further explorations of collectivistic and individualistic work groups. *Academy of Management Journal, 36,* 319–348. doi:10.2307/256525

Earley, P. C. (1999). Playing follow the leader: Status-determining traits in relation to collective efficacy across cultures. *Organizational Behavior and Human Decision Processes, 80,* 192–212. doi:10.1006/obhd.1999.2863

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121–138. doi:10.1037/1082-989X.12.2.121

Entrekin, L., & Chung, Y. W. (2001). Attitudes towards different sources of executive appraisal: A comparison of Hong Kong Chinese and American managers in Hong Kong. *International Journal of Human Resource Management, 12,* 965–987.

Erez, M., & Earley, P. C. (1993). *Culture, self-identity, and work.* New York, NY: Oxford University Press.

Farh, J.-L., Cannella, A. A., Jr., & Bedeian, A. G. (1991). The impact of purpose on rating quality and user acceptance. *Group and Organization Management, 16,* 367–386. doi:10.1177/105960119101600403

Farh, J. L., Hackett, R. D., & Liang, J. (2007). Individual-level cultural values as moderators of perceived organizational support-employee outcome relationships in China: Comparing the effects of power distance and traditionality. *Academy of Management Journal, 50,* 715–729.

Fedor, D. B., Bettenhausen, K. L., & Davis, W. (1999). Peer reviews: Employees' dual roles as raters and recipients. *Group and Organization Management, 24,* 92–120. doi:10.1177/1059601199241006

Fletcher, C., & Perry, E. L. (2001). Performance appraisal and feedback: A consideration of national culture and a review of contemporary research and future trends. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (Vol. 1, pp. 125–144). Thousand Oaks, CA: Sage.

French, J. R. P., Jr., & Raven, B. H. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in social power* (pp. 150–167). Ann Arbor, MI: Institute for Social Research.

Gelfand, M. J., Erez, M., & Aycan, Z. (2007). Cross-cultural organizational behavior. *Annual Review of Psychology, 58,* 479–514. doi:10.1146/annurev.psych.58.110405.085559

Greguras, G. J., & Robie, C. (1998). A new look at within-source interrater reliability of 360-degree feedback ratings. *Journal of Applied Psychology, 83,* 960–968. doi:10.1037/0021-9010.83.6.960

Harris, M. M. (1994). Rater motivation in the performance appraisal context: A theoretical framework. *Journal of Management, 20,* 737–756. doi:10.1016/0149-2063(94)90028-0

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-boss,

self-peer, and peer-boss ratings. *Personnel Psychology, 41,* 43–62. doi:10.1111/j.1744-6570.1988.tb00631.x

Harris, M. M., Smith, D. E., & Champagne, D. (1995). A field study of performance appraisal purpose: Research versus administrative-based ratings. *Personnel Psychology, 48,* 151–160. doi:10.1111/j.1744-6570.1995.tb01751.x

Heidemeier, H., & Moser, K. (2009). Self–other agreement in job performance ratings: A meta-analytic test of a process model. *Journal of Applied Psychology, 94,* 353–370. doi:10.1037/0021-9010.94.2.353

Hofmann, D. A. (1997). An overview of the logic and rationale of hierarchical linear models. *Journal of Management, 23,* 723–744. doi:10.1177/014920639702300602

Hofstede, G. (2001). *Culture's consequences.* Beverly Hills, CA: Sage.

Hogg, M. A., Martin, R., Epitropaki, O., Mankad, A., Svensson, A., & Weeden, K. (2005). Effective leadership in salient groups: Revisiting leader-member exchange theory from the perspective of the social identity theory of leadership. *Personality and Social Psychology Bulletin, 31,* 991–1004. doi:10.1177/0146167204273098

Holzbach, R. L. (1978). Rater bias in performance ratings: Superior, self and peer ratings. *Journal of Applied Psychology, 63,* 579–588. doi:10.1037/0021-9010.63.5.579

House, R., Hanges, P., Javidan, M., Dorfman, P., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies.* Thousand Oaks, CA: Sage.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum.

Ilies, R., Wagner, D. T., & Morgeson, F. (2007). Explaining affective linkages in teams: Individual differences in susceptibility to contagion and individualism-collectivism. *Journal of Applied Psychology, 92,* 1140–1148. doi:10.1037/0021-9010.92.4.1140

Jawahar, I. M., & Williams, C. R. (1997). When all the children are above average: The performance appraisal purpose effect. *Personnel Psychology, 50,* 905–925. doi:10.1111/j.1744-6570.1997.tb01487.x

Jöreskog, K. G., & Sörbom, D. (2006). *Windows LISREL 8.8.* Chicago, IL: Scientific Software.

Kim, J. Y., & Nam, S. H. (1998). The concept and dynamics of face: Implications for organizational behavior in Asia. *Organization Science, 9,* 522–534. doi:10.1287/orsc.9.4.522

Kirkman, B. L., Chen, G., Farh, J., & Chen, Z. X. (2009). Individual power distance orientation and follower reactions to transformational leaders: A cross-level, cross-cultural examination. *Academy of Management Journal, 52,* 744–764.

Kirkman, B. L., Lowe, K. B., & Gibson, C. B. (2006). A quarter century of*Culture's Consequences*: A review of empirical research incorporating Hofstede's cultural values framework. *Journal of International Business Studies, 37,* 285–320. doi:10.1057/palgrave.jibs.8400202

Klimoski, R. J., & London, M. (1974). Role of the rater in performance appraisal. *Journal of Applied Psychology, 59,* 445–451. doi:10.1037/h0037332

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling.* Thousand Oaks, CA: Sage.

Kudisch, J. D., Fortunato, V. J., & Smith, A. F. R. (2006). Contextual and individual difference factors predicting individuals' desire to provide upward feedback. *Group and Organization Management, 31,* 503–529. doi:10.1177/1059601106286888

Lance, C. E. (1994). Test of a latent structure of performance ratings derived from Wherry's (1952) theory of ratings. *Journal of Management, 20,* 751–771.

Lawler, E. E., III. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology, 51,* 369–381. doi:10.1037/h0025095

Lebreton, J. M., Burgess, J. R. D., Kaiser, R. B., Atchley, E. K., & James, L. R. (2003). The restriction of variance hypothesis and interrater reliability and agreement: Are ratings from multiple sources really dissimilar? *Organizational Research Methods, 6,* 80–128. doi:10.1177/1094428102239427

Leslie, J. B., Gryskiewicz, N. D., & Dalton, M. A. (1998). Understanding cultural influences on 360-degree feedback process. In W. W. Tornow & M. London (Eds.), *Maximizing the value of 360-degree feedback: A process for successful individual and organizational development* (pp. 196–216). San Francisco, CA: Jossey-Bass.

Leung, K., Bond, M. H., Reimel de Carrasquel, S., Muñoz, C., Hernández, M., Murakami, F., . . . Singelis, T. M. (2002). Social axioms: The search for universal dimensions of general beliefs about how the world functions. *Journal of Cross-Cultural Psychology, 33,* 286–302. doi:10.1177/0022022102033003005

Levy, P. E., & Williams, J. R. (2004). The social context of performance appraisal: A review and framework for the future. *Journal of Management, 30,* 881–905. doi:10.1016/j.jm.2004.06.005

Li, J., Ngin, P. M., & Teo, A. C. Y. (2007). Culture and leadership in Singapore: Combination of the East and the West. In J. Chhokar, F. C. Brodbeck, & R. J. House (Eds.), *Culture and leadership across the world: The GLOBE book of in-depth dtudies of 25 societies* (pp. 947–968). Mahwah, NJ: Lawrence Erlbaum Associates.

London, M. (2003). *Job feedback: Giving, seeking, and using feedback for performance improvement.* Mahwah, NJ: Lawrence Erlbaum.

London, M., & Smither, J. W. (1995). Can multisource feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and direction for research. *Personnel Psychology, 48,* 803–839. doi:10.1111/j.1744-6570.1995.tb01782.x

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98,* 224–253. doi:10.1037/0033-295X.98.2.224

Mero, N. P., & Motowidlo, S. J. (1995). Effects of rater accountability on the accuracy and the favorability of performance ratings. *Journal of Applied Psychology, 80,* 517–524. doi:10.1037/0021-9010.80.4.517

Morrison, E. W., Chen, Y.-R., & Salgado, S. R. (2004). Cultural differences in newcomer feedback seeking: A comparison of the United States and Hong Kong. *Applied Psychology: An International Review, 53,* 1–22. doi:10.1111/j.1464-0597.2004.00158.x

Mount, M. K. (1984). Psychometric properties of subordinate ratings of managerial performance. *Personnel Psychology, 37,* 687–702. doi:10.1111/j.1744-6570.1984.tb00533.x

Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology, 51,* 557–576. doi:10.1111/j.1744-6570.1998.tb00251.x

Mount, M. K., & Scullen, S. E. (2001). Multisource feedback ratings: What do they really measure? In M. London (Ed.), *How people evaluate others in organizations* (pp. 155–176). Mahwah, NJ: Lawrence Erlbaum.

Murphy, K. R., & Balzer, W. K. (1989). Rater errors and rating accuracy. *Journal of Applied Psychology, 74,* 619–624. doi:10.1037/0021-9010.74.4.619

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives.* Thousand Oaks, CA: Sage.

Murphy, K. R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters who pursue different goals give different ratings. *Journal of Applied Psychology, 89,* 158–164. doi:10.1037/0021-9010.89.1.158

Ng, K. Y., & Van Dyne, L. (2001). Individualism-collectivism as a boundary condition for effectiveness of minority influence in decision making. *Organizational Behavior and Human Decision Processes, 84,* 198–225. doi:10.1006/obhd.2000.2927

Pfeffer, J., & Salancik, G. R. (1975). Determinants of supervisory behaviors: A role set analysis. *Human Relations, 28,* 139–154. doi:10.1177/001872677502800203

Ployhart, R. E., Wiechmann, D., Schmitt, N., Sacco, J. M., & Rogg, K. (2003). The cross-cultural equivalence of job performance ratings. *Human Performance, 16,* 49–79. doi:10.1207/S15327043HUP1601_3

Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin, 112,* 160–164. doi:10.1037/0033-2909.112.1.160

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Reilly, R. R., Smither, J. W., & Vasilopoulos, N. L. (1996). A longitudinal study of upward feedback. *Personnel Psychology, 49,* 599–612. doi:10.1111/j.1744-6570.1996.tb01586.x

Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin, 88,* 413–428. doi:10.1037/0033-2909.88.2.413

Schneier, E. C. (1977). Operational utility and psychometric characteristics of behavioral expectations scales: A cognitive reinterpretation. *Journal of Applied Psychology, 62,* 541–548. doi:10.1037/0021-9010.62.5.541

Schwartz, S. M., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology, 53,* 550–562. doi:10.1037/0022-3514.53.3.550

Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85,* 956–970. doi:10.1037/0021-9010.85.6.956

Shipper, F., Hoffman, R. C., & Rotondo, D. M. (2007). Does the 360 feedback process create actionable knowledge equally across cultures? *Academy of Management Learning and Education, 6,* 33–50.

Spence, J. R., & Keeping, L. M. (2010). The impact of non-performance information on ratings of job performance: A policy-capturing approach. *Journal of Organizational Behavior, 31,* 587–608. doi:10.1002/job.648

Springer, D. (1953). Ratings of candidates for promotion by co-workers and supervisors. *Journal of Applied Psychology, 64,* 124–131.

Stone-Romero, E. F., & Stone, D. L. (2002). Cross-cultural differences in responses to feedback: Implications for individual, group, and organizational effectiveness. *Research in Personnel and Human Resources Management, 21,* 275–331. doi:10.1016/S0742-7301(02)21007-5

Triandis, H. C. (1995). *Individualism and collectivism.* Boulder, CO: Westview Press.

Tsui, A. S. (1984). A role set analysis of managerial reputation. *Organizational Behavior and Human Performance, 36,* 64–96. doi:10.1016/0030-5073(84)90037-0

Tsui, A. S., & Barry, B. (1986). Interpersonal affect and rating errors. *Academy of Management Journal, 29,* 586–599. doi:10.2307/256225

Tsui, A., & O'Reilly, C. A., III. (1989). Beyond simple demographic effects: The importance of relational demography in superior-subordinate dyads. *Academy of Management Journal, 32,* 402–423. doi:10.2307/256368

Vance, R. J., Winne, P. S., & Wright, E. S. (1983). A longitudinal examination of rater and ratee effects in performance ratings. *Personnel Psychology, 36,* 609–620. doi:10.1111/j.1744-6570.1983.tb02238.x

Varela, O. E., & Premeaux, S. F. (2008). Do cross-cultural values affect multisource feedback dynamics? The case of high power distance and collectivism in two Latin American countries. *International Journal of Selection and Assessment, 16,* 134–142. doi:10.1111/j.1468-2389.2008.00418.x

Walker, A. G., & Smither, J. W. (1999). A five-year study of upward feedback: What managers do with their results matter. *Personnel Psychology, 52,* 393–423. doi:10.1111/j.1744-6570.1999.tb00166.x

Westerman, J. W., & Rosse, J. G. (1997). Reducing the threat of rater nonparticipation in 360-degree feedback systems: An exploratory examination of antecedents to participation in upward ratings. *Group and Organization Management, 22,* 288–309. doi:10.1177/1059601197222008

Wherry, R. J., Sr., & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology, 35,* 521–551. doi:10.1111/j.1744-6570.1982.tb02208.x

Wohlers, A. J., Hall, M. J., & London, M. (1993). Subordinates rating managers: Organizational and demographic correlates of self-subordinate agreement. *Journal of Occupational and Organizational Psychology, 66,* 263–275.

Wong, K. F. E., & Kwong, J. Y. Y. (2007). Effects of rater goals on rating patterns: Evidence from an experimental field study. *Journal of Applied Psychology, 92,* 577–585. doi:10.1037/0021-9010.92.2.577

Yukl, G., & Falbe, C. M. (1991). Importance of different power sources in downward and lateral relations. *Journal of Applied Psychology, 76,* 416–423. doi:10.1037/0021-9010.76.3.416

Zedeck, S., Imparto, N., Krausz, M., & Oleno, T. (1974). Development of behaviorally anchored rating scales as a function of organizational level. *Journal of Applied Psychology, 59,* 249–252. doi:10.1037/h0036521

Received December 14, 2009
Revision received January 12, 2011
Accepted January 24, 2011 ■